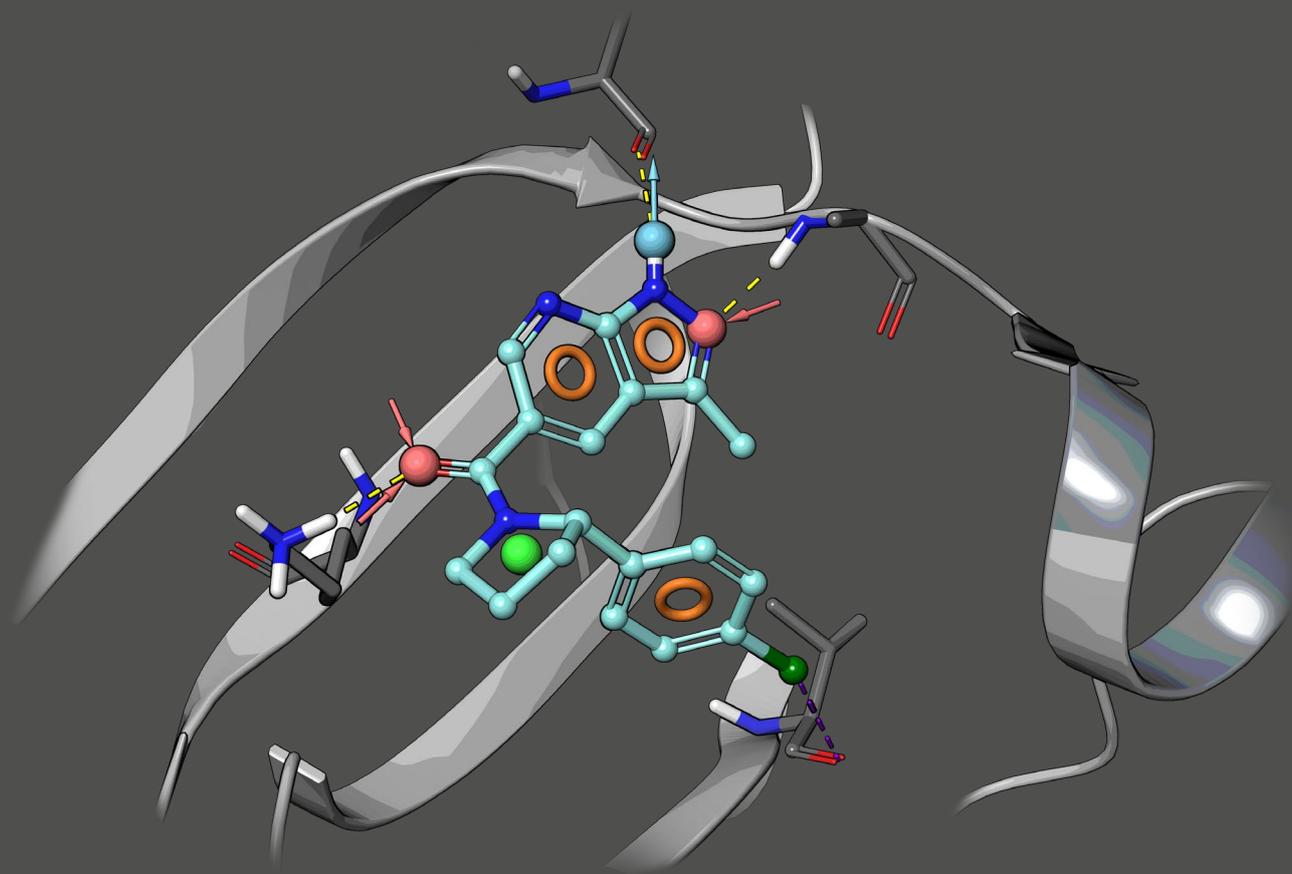


А.Ю. Кудинов, А.С. Бунев

ОСНОВЫ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ БИОЛОГИЧЕСКИ АКТИВНЫХ ВЕЩЕСТВ



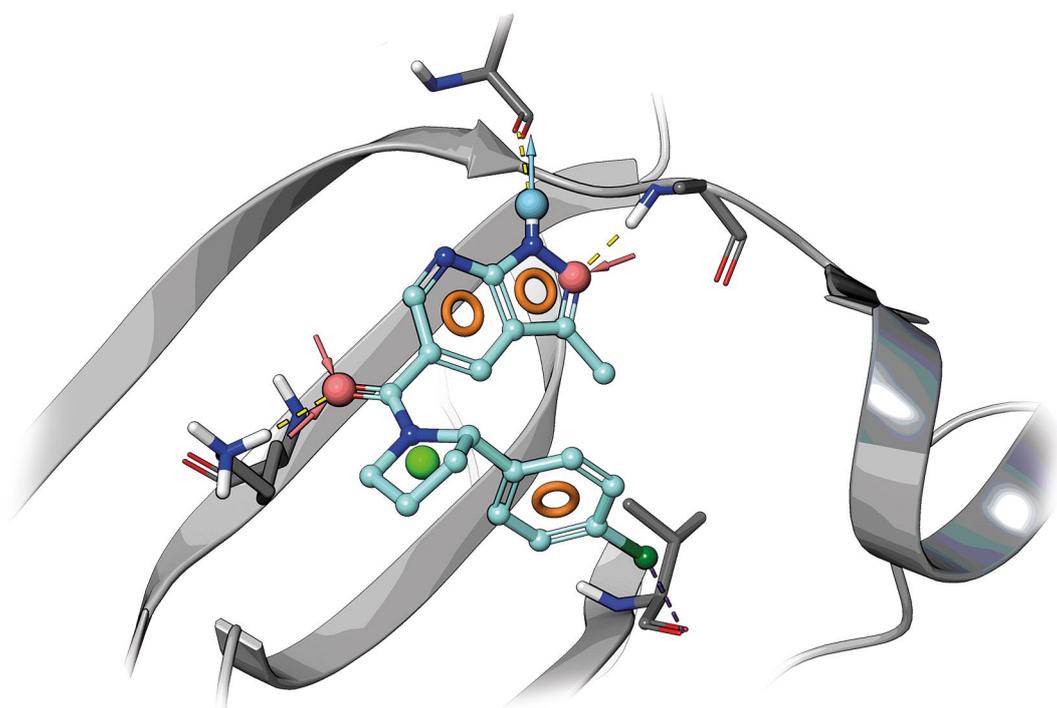
Тольятти
2026

Министерство науки и высшего образования Российской Федерации
Тольяттинский государственный университет

А.Ю. Кудинов, А.С. Бунев

ОСНОВЫ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ БИОЛОГИЧЕСКИ АКТИВНЫХ ВЕЩЕСТВ

Лабораторный практикум



Тольятти
Издательство ТГУ
2026

УДК 577.3(075.8)+615.1(075.8)

ББК 28.072я73+52.8я73

К887

Рецензенты:

д-р хим. наук, эксперт лаборатории волновых процессов и систем управления,
профессор кафедры инновационной фармацевтики, медицинской техники
и биотехнологий Московского физико-технического института *К.В. Балакин*;
канд. хим. наук, профессор центра медицинской химии
Тольяттинского государственного университета *О.Б. Григорьева*.

Кудинов, Андрей Юрьевич.

К887 Основы молекулярного моделирования биологически активных веществ : лабораторный практикум / А.Ю. Кудинов, А.С. Бунев. – Тольятти : Издательство ТГУ, 2026. – 168 с.

ISBN 978-5-8259-1778-8.

Лабораторный практикум представляет собой руководство к лабораторным занятиям по дисциплине «Основы молекулярного моделирования биологически активных веществ». В нем приведены теоретические материалы, лабораторные работы и порядок их выполнения.

Лабораторный практикум предназначен для студентов, обучающихся по направлению подготовки 04.03.01 «Химия» очной формы обучения.

УДК 577.3(075.8)+615.1(075.8)

ББК 28.072я73+52.8я73

Рекомендовано к изданию научно-методическим советом Тольяттинского государственного университета.

ISBN 978-5-8259-1778-8

© Кудинов А.Ю., Бунев А.С., 2026

© ФГБОУ ВО «Тольяттинский
государственный университет», 2026

Введение

Сегодня невозможно представить процесс разработки лекарства без использования компьютерной химии. Молекулярный докинг, QSAR (*англ.* Quantitative Structure-Activity Relationship), структурно-ориентированный виртуальный скрининг, виртуальные библиотеки – понятия, прочно закрепившиеся в современной медицинской химии.

Данный лабораторный практикум направлен на ознакомление с основными инструментами молекулярного моделирования, а также приобретение некоторых навыков работы в этой области, которые могут пригодиться вам в собственных исследованиях.

В ходе выполнения лабораторных работ вы сможете ознакомиться с протоколами подготовки структур белков и лигандов, различными методами конформационного поиска, фармакофорного моделирования и скрининга. В практикуме подробно разобраны процедуры выполнения молекулярного докинга (а также ковалентной и индуцированной стыковки).

Практические указания к каждой лабораторной работе предваряются небольшим теоретическим введением со ссылками на оригинальные источники. Для успешной защиты работы настоятельно рекомендуется ознакомиться с источниками, приведенными в списке литературы.



Лабораторная работа 1

ОСНОВЫ РАБОТЫ В MAESTRO. ПОДГОТОВКА СТРУКТУР ЛИГАНДОВ И БЕЛКА

Цель работы: освоение основных подходов к подготовке структур биологической мишени и лиганда для задач молекулярного моделирования.

Вам предстоит:

- 1) получить начальные навыки работы с графическим интерфейсом Maestro;
- 2) познакомиться со стандартным протоколом подготовки моделей белков, полученных методом рентгеноструктурного анализа (РСА), с помощью инструмента Protein Preparation Workflow [1, 2];
- 3) познакомиться со стандартным протоколом подготовки структур лигандов с помощью инструмента LigPrep [1, 3];
- 4) получить начальные навыки поиска структур белков в Банке данных белковых структур (*англ.* PDB, Protein Data Bank) [4].

Теоретическое введение

Для выполнения большинства задач CADD (*англ.* Computer-Aided Drug Design), будь то виртуальный скрининг, моделирование молекулярной динамики или фармакофорный поиск, требуются качественные и готовые к работе структуры белков и лигандов. Данная лабораторная работа посвящена ознакомлению с общими подходами к их подготовке.

Долгое время рентгеноструктурный анализ являлся единственным способом получения моделей протеинов. Однако использование РСА таит в себе несколько недостатков, в частности сложности кристаллизации белка, а также неточности и ошибки, возникающие из-за *фазовой проблемы*. В структурах белков, решенных рентгеноструктурным анализом, отсутствуют атомы водорода, связи не имеют корректного порядка, а некоторые боковые цепи белка и вовсе могут отсутствовать. Кроме того, конформация белка в кристалле – совсем не то же самое, что конформация белка в среде биологической жидкости (здесь, однако, стоит отметить, что ячейка кристалла содержит большое количество молекул растворителя, а белки удалены друг от друга, так что для большинства глобулярных протеинов серьезных конформационных отклонений наблюдаться не будет) [4]. Процесс определения структуры с помощью методов РСА будет рассмотрен более подробно.

Обязательное условие для проведения РСА – наличие кристалла, и в этом скрывается первый существенный недостаток данного метода определения структуры. На качество получаемых кристаллов оказывает влияние большое число факторов: рН, температура, концентрация соосадителей, концентрация, степень очистки и размер самого белка, тип буфера. Путь к выращиванию пригодного для анализа образца в первом приближении требует оптимизации всех этих параметров и условий, и не зря отмечают, что кристаллизация белков – больше искусство, чем наука [4]. Примеры «хороших» кристаллов представлены на рис. 1.1.

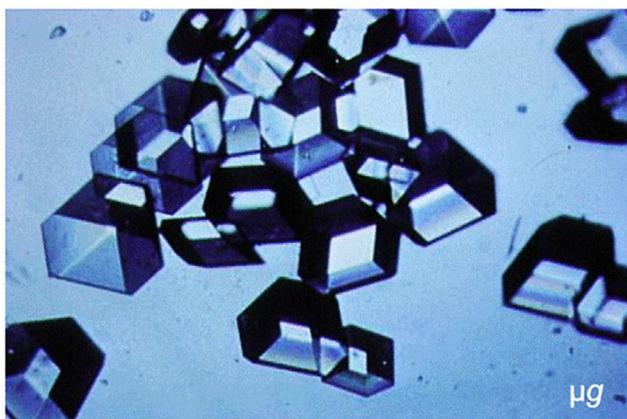


Рис. 1.1. Кристаллы человеческого инсулина, выращенные в условиях микрогравитации [5]

Облучение образца сфокусированным пучком мощного монохроматического рентгеновского излучения (получаемого на синхротронных ускорителях) приводит к дифракции лучей на кристаллической решетке. Отраженные на электронах регулярно расположенных атомов лучи формируют дифракционную картину (рентгенограмму), интенсивность рефлексов на которой фиксируют с помощью детекторов (рис. 1.2).

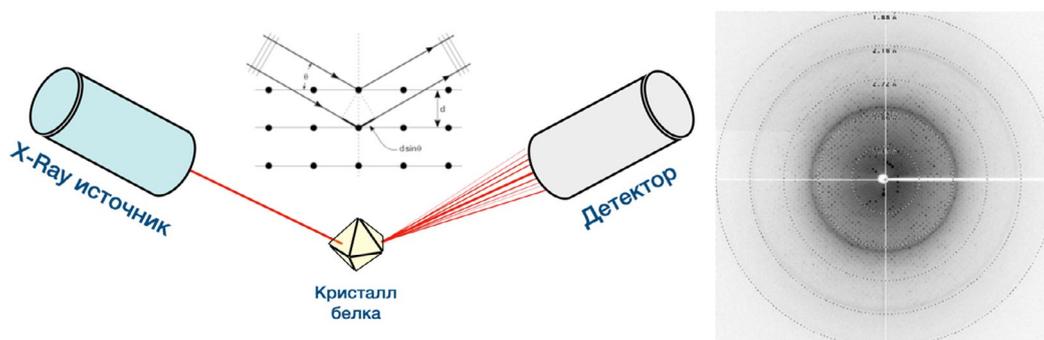


Рис. 1.2. Упрощенная схема установки и дифракционная картина мембранного белка [6]

Из обработанных данных рентгенограммы можно получить карту электронной плотности $\rho(r)$ в элементарной ячейке кристалла:

$$\rho(r) = \frac{1}{V} \sum_h \mathbf{F}_h e^{2\pi i(h,r)},$$

где $\rho(r)$ – функция электронной плотности, V – объем элементарной ячейки, r – вектор координат, $h (h, k, l)$ – вектор, описывающий расположение атомных плоскостей в кристалле (и характеризующий рефлексы на дифракционной картине), а коэффициент F_h – так называемый структурный фактор, который может быть представлен в виде

$$\mathbf{F}_h = |F_h| e^{i\varphi_h},$$

где φ_h – фаза.

Квадрат модуля структурного фактора $|F_h|^2$ связан с интенсивностью отражения, и с его помощью можно было бы вычислить электронную плотность в объеме элементарной ячейки, однако информация о фазе теряется в ходе проведения эксперимента. О методах решения фазовой проблемы можно узнать подробнее в [4].

Следующий этап заключается в интерпретации и уточнении полученной карты электронной плотности, а также попытке построения приближенной атомной модели белка на ее основе. Построение модели требует внимания со стороны оператора, и, наверное, самым важным фактором, который может привести к ошибкам, является разрешение (англ. Resolution) – мера уровня детализации, с которой рассматривается белок. При разрешении 4 Å можно различить общую форму макромолекулы и некоторые элементы вторичной структуры, при разрешении 3 Å становится очевидна главная цепь и формы боковых цепей; при разрешении 1 Å можно распознать отдельные атомы и даже идентифицировать их тип (рис. 1.3) [7]. **На практике рекомендуется использовать структуры, обладающие разрешением ≤ 2.5 Å.**

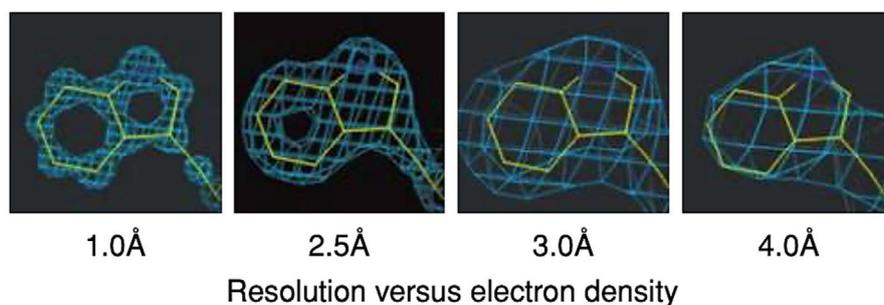


Рис. 1.3. Наглядная иллюстрация легкости интерпретации данных карт электронной плотности при различном разрешении на примере остатка триптофана [7]

Для оценки качества интерпретации на основе модели вновь рассчитывается карта электронной плотности, которая сравнивается с полученной экспериментально. Степень их соответствия друг другу описывается с помощью такой метрики качества структуры белка, как R-фактор:

$$R = \frac{\sum_h |F^{obs}(h) - F^{calc}(h)|}{\sum_h F^{obs}(h)},$$

где $F^{obs}(h)$ и $F^{calc}(h)$ – модули структурных факторов из эксперимента и рассчитанные на основе модели, соответственно.

Еще одной метрикой является R_{free} -фактор, который рассчитывается точно так же, за тем лишь исключением, что для сравнения используется та часть экспериментальных данных, которая не была использована при построении моделей (контрольная выборка).

Для совершенно случайного набора атомов R будет равен 0.63, для модели, которая идеально описывает данные эксперимента, разумеется, 0, но **на практике для хорошо решенных структур R обычно принимает значения до 0.2, а R_{free} – до 0.26** [8].

Мерой локальной конформационной подвижности атомов в полученной модели белка является **B-фактор** (или температурный фактор), отражающий степень делокализации электронной плотности атомов из-за температурно-зависимых атомных вибраций. Чем выше значение B-фактора, тем выше подвижность остатка или атома.

Очевидно, что решить структуру белка можно не только с помощью РСА. Развитие методов **двумерной спектроскопии ядерного магнитного резонанса (2D ЯМР)** позволило использовать методы магнитного резонанса для определения третичных структур протеинов, причем, в отличие от РСА – в растворе [9] (рис. 1.4).



Рис. 1.4. Ансамбль конформаций белка цинкового пальца 126 (англ. RNF126), определенный с помощью ЯМР в растворе (PDB ID: 2P2M [10])

Недостатки ЯМР в роли инструмента для определения структур белков связаны с необходимостью выращивания бактерий, используемых в качестве систем экспрессии, на изотопно-обогащенных средах, слишком большими временами релаксации и трудностью интерпретации спектров для крупных белков [9].

Криоэлектронная микроскопия (*англ.* Cryo-EM) является сравнительно новым методом, который может работать с довольно широким диапазоном размеров белков, включая очень крупные комплексы (на самом деле, с помощью Cryo-EM получают модели даже целых органоидов и вирусов). В основе метода лежит применение просвечивающей электронной микроскопии при низких температурах и последующий анализ полученных изображений. До недавнего времени ограничением была низкая разрешающая способность, однако сейчас и эта проблема является решенной [11].

Если белок трудно кристаллизовать, а методами ЯМР и Cryo-EM определить его структуру невозможно, то помочь может **моделирование *de novo*** на основе последовательности и структур известных гомологичных протеинов. Такой подход имеет место быть, однако получаемые модели содержат много ошибок [12].

Так или иначе, решенные структуры молекул помещаются в ту или иную базу данных со свободным доступом. Информация об органических и неорганических молекулах размером до 500 атомов хранится в Кембриджской базе структурных данных (*англ.* Cambridge Structural Database, CSD) [13]. Структуры белков, нуклеиновых кислот, фрагментов вирусов и т. д. содержатся в Банке данных белковых структур (*англ.* Protein Data Bank, PDB) [14].

Каждая запись PDB имеет свой идентификационный номер (PDB ID), состоящий из четырех символов, и заголовок. Страница записи содержит информацию о виде организма, из которого выделен белок, качестве модели, массе белка и числе атомов, выходные данные оригинальной публикации, в рамках исследований которой данная структура была определена и использована, данные о симметрии и стехиометрии структуры, простейшие средства визуализации (рис. 1.5).

На рис. 1.6 можно найти информацию о макромолекулах. Раздел Entity 1, в данном случае, описывает цепь А фактора свертывания крови ХА длиной 234 остатка. Имеются ссылки на соответствующие записи в базе UniProt [15]. Раздел аннотации последовательности позволяет перейти к записям PDB, содержащим белки с различными уровнями гомологии последовательности, а также данные о гидрофобности/гидрофильности остатков (HYDROPATHY) и упорядоченности структуры (DISORDER).

Малые молекулы, к которым относятся как лиганды, так и неорганические ионы, органический растворитель, содержащийся в элементарной ячейке, также описываются. Их аннотация включает в себя указание систематического названия, брутто-формулы, нотации InChi, а также имени цепи. Запись E [auth A] для молекулы Z34 на рис. 1.7 означает, что при обработке структуры авторами данной молекуле было присвоено имя «А», но при загрузке файла .pdb имя цепи будет изменено на «Е». Также страница записи PDB может содержать ссылки на данные о биологической активности (рис. 1.7).

Очевидно, что первоначально структуру белка нужно найти. Банк данных белковых структур позволяет осуществлять поиск с помощью внушительного набора атрибутов, среди которых имеются аминокислотная последовательность, ключевые слова из заголовка записи/статьи, название самого белка (гена или классификационного семейства), структуры связанного лиганда. Поисковый запрос можно уточнить с помощью указания метода определения структуры, диапазона значений разрешения, R и R_{free} , системы экспрессии и т. д. Для выполнения расширенного поиска следует раскрыть меню Advanced Search рядом с поисковой строкой на главной странице. На рис. 1.8 показан пример с указанием PCA в качестве метода определения структуры.

Модуль визуализации

PDB ID → **1FJS** (Заголовок)

CRYSTAL STRUCTURE OF THE INHIBITOR ZK-807834 (CI-1031) COMPLEXED WITH FACTOR XA

PDB DOI: <https://doi.org/10.2210/pdb1FJS/pdb>

Classification: **BLOOD CLOTTING**

Organism(s): **Homo sapiens** ← Исходный организм

Mutation(s): No → Число мутаций

Deposited: 2000-08-08 Released: 2000-10-04

Deposition Author(s): Adler, M., Whitlow, M.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.92 Å

R-Value Free: 0.256

R-Value Work: 0.196

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.252
Clashscore		11
Ramachandran outliers		0.4%
Sidechain outliers		4.5%
RSRZ outliers		1.7%

Ligand Structure Quality Assessment

Worse 0 | 1 Better

Ligand structure goodness of fit to experimental data

Parameters of structure quality

This is version 1.4 of the entry. See complete [history](#).

Macromolecule Content

- Total Structure Weight: 33.19 kDa
- Atom Count: 2,475
- Modelled Residue Count: 286
- Deposited Residue Count: 286
- Unique protein chains: 2

Стехиометрия: данная запись содержит гетеродимер (две цепи)

Данные о содержимом: масса, число атомов и т. д.

Literature

Preparation, characterization, and the crystal structure of the inhibitor ZK-807834 (CI-1031) complexed with factor Xa.

[Adler, M.](#), [Davey, D.D.](#), [Phillips, G.B.](#), [Kim, S.H.](#), [Jancarik, J.](#), [Rumennik, G.](#), [Light, D.R.](#), [Whitlow, M.](#)

(2000) *Biochemistry* **39**: 12534-12542

PubMed: [11027132](#) Search on PubMed

DOI: <https://doi.org/10.1021/bi001477q>

Primary Citation of Related Structures: 1FJS

PubMed Abstract:

Factor Xa plays a critical role in the formation of blood clots. This serine protease catalyzes the conversion of prothrombin to thrombin, the first joint step that links the intrinsic and extrinsic coagulation pathways. There is considerable interest in the... development of factor Xa inhibitors for the intervention in thrombotic disease. This paper...

Organizational Affiliation:

Berlex Biosciences, 15049 San Pablo Avenue, P.O. Box 4099, Richmond, California 94804-0099, USA.

Библиографические данные

Рис. 1.5. Запись структуры в PDB

Macromolecules

Find similar proteins by: (by identity cutoff) | 3D Structure

Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details	Image
COAGULATION FACTOR XA	A ← Цепь Имя макромолекулы	234 ← Длина	Homo sapiens	Mutation(s): 0 Gene Names: F10 EC: 3.4.21.6	

UniProt & NIH Common Fund Data Resources

Find proteins for P00742 (Homo sapiens) Explore Go to UniProtKB:

PHAROS: Ссылка на UniProt

Entity Groups

Sequence Clusters:

UniProt Group:

Sequence Annotations

Reference Sequence: 1FJ... | Аннотация последовательности

Рис. 1.6. Аннотация макромолекул, содержащихся в записи PDB

Small Molecules

Ligands: 4 Unique

ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
Z34 ← Имя/ID Query on Z34 Download Ideal Coordinates CCD File Download Instance Coordinates	E [auth A]	N-[2-[5-[AMINO(IMINO)METHYL]-2-HYDROXYPHENOXY]-3,5-DIFLUORO-6-[3-(4,5-DIHYDRO-1-METHYL-1H-IMIDAZOL-2-YL)PHENOXY]PYRIDIN-4-YL]-N-METHYLGLYCINE C ₂₅ H ₂₄ F ₂ N ₆ O ₅ NPNSVNGQJGRSNR-UHFFFAOYSA-N		Interactions Interactions & Density 3D-визуализация
GOL Query on GOL Download Ideal Coordinates CCD File Download Instance Coordinates	F [auth A] G [auth A] H [auth A] I [auth A] J [auth A]	GLYCEROL C ₃ H ₈ O ₃ PEDCQBHVMGVHV-UHFFFAOYSA-N		Interactions Interactions & Density
CA Query on CA Download Ideal Coordinates CCD File Download Instance Coordinates	C [auth A]	CALCIUM ION Ca BHPQYMZQTOCNFJ-UHFFFAOYSA-N	Ca ⁺²	Interactions Interactions & Density
CL Query on CL Download Ideal Coordinates CCD File Download Instance Coordinates	D [auth A]	CHLORIDE ION Cl VEXZGXHMUGYJMC-UHFFFAOYSA-M	Cl ⁻	Interactions Interactions & Density

Binding Affinity Annotations

ID	Source	Binding Affinity
Z34	BindingDB: 1FJS	Ki: 0.11 (nM) from 1 assay(s)
Z34	PDBBind: 1FJS	Ki: 0.11 (nM) from 1 assay(s)
Z34	Binding MOAD: 1FJS	Ki: 0.11 (nM) from 1 assay(s)

Цепь

Систематическое название, нотация

Данные о биологической активности

Рис. 1.7. Аннотация малых молекул и ионов, содержащихся в записи PDB

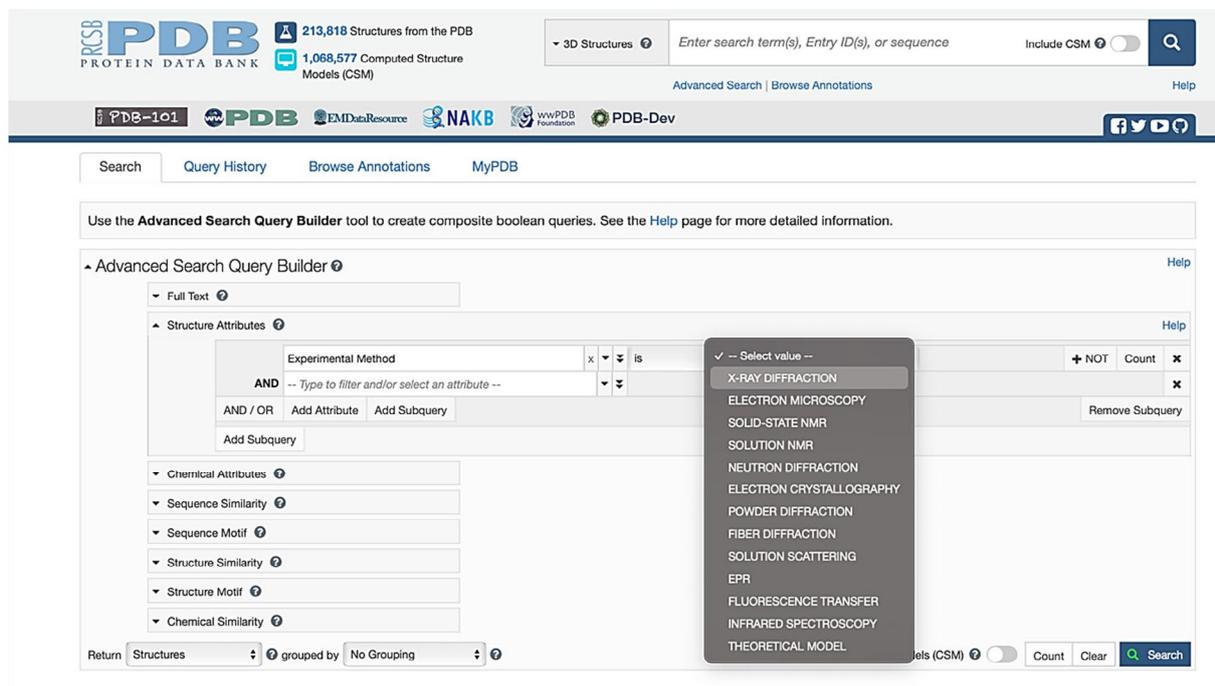


Рис. 1.8. Расширенный поиск PDB

Основная информация, хранящаяся в PDB, состоит из файлов координат, в которых перечислены атомы в каждой структуре и их 3D-местоположение в пространстве, а также сводная информация о структуре, последовательности и эксперименте. Эти файлы доступны в нескольких форматах (.PDBx/mmCIF, .PDB, .XML). Чтобы полностью изучить структуры в PDB, полезно узнать несколько понятий о файлах координат. Кроме того, эти знания помогут в использовании программ визуализации. Открыв файл .pdb в текстовом редакторе, можно увидеть следующую картину (рис. 1.9).

Первые 1000–2000 строк содержат информацию об авторах, названии, PDB ID, разрешении, эксперименте и т. д., которая не будет интересовать нас в рамках работы с данным практикумом.

Каждый атом в файле PDB имеет идентификатор (АТОМ – для атомов белка, НЕТАТМ – для атомов лиганда, ионов, растворителя), порядковый номер, тип (CA – α -атом углерода главной цепи и т. д.), код аминокислотного остатка, имя цепи, к которой он принадлежит, номер в последовательности, и, наконец, x , y , z координаты. Последние два столбца соответствуют значениям занятости (*англ.* Occupancy) и B-фактора (мера локальной подвижности) (рис. 1.10).

Стоит сказать несколько слов о таком параметре, как занятость. Кристаллы белков состоят из множества отдельных симметрично упакованных молекул. В некоторых кристаллах есть небольшие различия между конформациями макромолекул, выраженные, например, в ориентации боковых цепей на поверхности белков. Когда исследователи строят атомную модель этих частей, они могут использовать занятость для оценки количества каждой конформации. Для большинства атомов занятость задается значением 1, что указывает на то, что этот атом находится во всех молекулах в одной и той же конформации. Однако если, предположим, ион металла связывается только с половиной молекул в кристалле, исследователь увидит слабое изображение иона на карте плотности электронов и может назначить заполняемость как 0.5.

```

HEADER      LIGASE                               26-FEB-14  4PPE
TITLE       HUMAN RNF4 RING DOMAIN
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: E3 UBIQUITIN-PROTEIN LIGASE RNF4;
COMPND     3 CHAIN: A, B;
COMPND     4 FRAGMENT: HRNF4 RING DOMAIN;
COMPND     5 SYNONYM: RING FINGER PROTEIN 4, SMALL NUCLEAR RING FINGER PROTEIN,
COMPND     6 PROTEIN SNURF;
COMPND     7 EC: 6.3.2.-;
COMPND     8 ENGINEERED: YES
SOURCE     MOL_ID: 1;
SOURCE     2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE     3 ORGANISM_COMMON: HUMAN;
SOURCE     4 ORGANISM_TAXID: 9606;
SOURCE     5 GENE: RES4-26, RNF4, SNURF;
SOURCE     6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE     7 EXPRESSION_SYSTEM_TAXID: 562;
SOURCE     8 EXPRESSION_SYSTEM_STRAIN: BL21 (DE3) PRIL;
SOURCE     9 EXPRESSION_SYSTEM_VECTOR_TYPE: PGEX-6P-1
KEYWDS     RING DOMAIN, UBIQUITIN LIGASE, LIGASE
EXPDTA     X-RAY DIFFRACTION
AUTHOR     J.J.PERRY, A.S.ARVAI, C.HITOMI, J.A.TAINER
REVDAT     2 25-FEB-15 4PPE 1 JRNL
REVDAT     1 26-MAR-14 4PPE 0
JRNL       AUTH L.M.GROOCCOCK, M.NIE, J.PRUDDEN, D.MOIANI, T.WANG, A.CHELTSOV,
JRNL       AUTH 2 R.P.RAMBO, A.S.ARVAI, C.HITOMI, J.A.TAINER, K.LUGER, J.J.PERRY,
JRNL       AUTH 3 E.LAZZERINI-DENCHI, M.N.BODDY
JRNL       TITL RNF4 INTERACTS WITH BOTH SUMO AND NUCLEOSOMES TO PROMOTE THE
JRNL       TITL 2 DNA DAMAGE RESPONSE.
JRNL       REF EMBO REP. V. 15 601 2014
JRNL       REFN ISSN 1469-221X
JRNL       PMID 24714598
JRNL       DOI 10.1002/EMBR.201338369
REMARK     2
REMARK     2 RESOLUTION. 2.00 ANGSTROMS.
    
```

Рис. 1.9. Содержимое файла .pdb

Принадлежность атома к белку/лигандам	Номер атома в структуре	Тип атома	AMK	Цепь	Координаты			B-Фактор	Занятость (Occupancy)		
					Номер AMK	X	Y				Z
АТОМ	1	N	ARG	A	129	15.560	16.154	6.460	1.00	40.18	N
АТОМ	2	CA	ARG	A	129	15.044	17.088	7.456	1.00	36.21	C
АТОМ	3	C	ARG	A	129	16.101	17.906	8.216	1.00	39.10	C
АТОМ	4	O	ARG	A	129	15.748	18.692	9.099	1.00	45.76	O
АТОМ	5	CB	ARG	A	129	14.144	16.357	8.453	1.00	42.62	C
АТОМ	6	CG	ARG	A	129	12.648	16.515	8.196	1.00	42.96	C
АТОМ	7	CD	ARG	A	129	12.133	17.887	8.638	1.00	39.04	C
АТОМ	8	NE	ARG	A	129	12.594	18.958	7.759	0.50	31.79	N
АТОМ	9	CZ	ARG	A	129	11.873	19.481	6.773	0.50	30.27	C
АТОМ	10	NH1	ARG	A	129	10.647	19.037	6.536	0.50	37.26	N
АТОМ	11	NH2	ARG	A	129	12.380	20.450	6.025	0.50	29.58	N

Рис. 1.10. Содержимое файла .pdb, аннотирование атомов

В этой работе мы будем использовать структуру белково-лигандного комплекса, модель которого получена методом PCA и депонирована в Банк данных белковых структур (PDB). Основные этапы подготовки будут включать:

- добавление атомов водорода и корректировку направления водородных связей;
- назначение связям корректных порядков;
- прогнозирование состояний протонирования для гетероатомов;
- добавление отсутствующих атомов;
- очистку структуры: удаление ненужных макромолекул, молекул воды, кристаллографических артефактов;
- жесткую минимизацию структуры белка;
- конвертацию в приемлемый для дальнейшей работы формат.

Структуры лигандов также требуют подготовки. Представляемые в основном в виде линейных нотаций SMILES или графов (.sdf), они нуждаются:

- в создании стабильной конформации (получении 3D-координат атомов);
- прогнозировании состояний протонирования;
- конвертации в приемлемый для дальнейшей работы формат;
- создании стереоизомеров и таутомеров;
- нейтрализации, десольватации.

Кроме того, библиотеки лигандов нуждаются в фильтрации для отсеивания нежелательных структур, дубликатов и т. д., однако данные операции будут выполняться в рамках других лабораторных работ [1, 16].

Экспериментальная часть

1.1. Создание проекта и импорт структуры белка

1.1.1. Поиск структуры белка в PDB

Перейдите на официальную страницу PDB (<https://www.rcsb.org>). В данной работе мы попробуем подготовить структуру фактора свертывания крови Ха, относящегося к классу сериновых протеаз. В поисковой строке вбейте «**Factor XA**» и нажмите **Search**. Через несколько секунд ваш запрос будет обработан поисковой машиной PDB, результаты загрузятся в виде карточек, содержащих краткую информацию о записях. В разделе **Search Summary** вы можете увидеть, что запросу соответствует около 90 000 структур. При беглом просмотре записей видно, что сейчас результаты содержат комплексы трипсина или фактора X кроме искомым структур фактора Ха, а также имеют низкое обогащение моделями человеческих белков (рис. 1.11).

Уточним параметры поиска. Частично вы можете сделать это в левой панели **Refinements**. В разделе **Experimental Method** поставьте галочку напротив пункта **X-RAY DIFFRACTION**, в разделе **Scientific Name of Source Organism** – напротив **Homo sapiens** и нажмите на значок .

Теперь будут показаны только структуры человеческих белков, полученные методом PCA (рис. 1.12).

QUERY: Full Text = "Factor XA" MyPDB Login Search API

Advanced Search Query Builder Help

Search Summary This query matches 89,204 Structures.

Refinements Tabular Report All Selected

Structure Determination Methodology

experimental (89,204)

Scientific Name of Source Organism

- Homo sapiens (37,401)
- Mus musculus (4,817)
- synthetic construct (3,543)
- Escherichia coli (2,769)
- Rattus norvegicus (2,491)
- Bos taurus (1,912)
- Escherichia coli K-12 (1,773)
- Saccharomyces cerevisiae (1,707)
- Saccharomyces cerevisiae S288C (1,602)
- Severe acute respiratory syndrome coronavirus 2 (1,255)
- [More...](#)

Taxonomy

- Eukaryota (58,849)
- Bacteria (24,994)
- Riboviria (4,613)
- other sequences (3,576)
- Archaea (2,005)
- Duplodnaviria (976)
- Eukaryota (eukaryotes) (905)
- Varidnaviria (307)
- unclassified sequences (194)
- Monodnaviria (170)
- [More...](#)

1 to 25 of 89,204 Structures Page 1 of 3,569 25 Sort by | Score

1KIG Download File View File

BOVINE FACTOR XA

Wei, A., Alexander, R., Chang, C.-H.

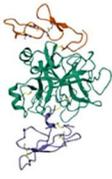
(1998) J Mol Biol 283: 147-154

Released 1998-10-28

Method X-RAY DIFFRACTION 3 Å

Organisms [Bos taurus](#)
[Ornithodoros moubata](#)

Macromolecule [ANTICOAGULANT PEPTIDE \(protein\)](#)
[FACTOR XA \(protein\)](#)

 [Explore in 3D](#)

1MTS Download File View File

FACTOR XA SPECIFIC INHIBITOR IN COMPLEX WITH BOVINE TRYPSIN

Stubbs, M.T.

(1995) FEBS Lett 375: 103-107

Released 1997-08-20

Method X-RAY DIFFRACTION 1.9 Å

Organisms [Bos taurus](#)

Macromolecule [TRYPSIN \(protein\)](#)

Unique Ligands [BX3, CA](#)

 [Explore in 3D](#)

Рис. 1.11. Результаты поиска

Refinements ▶

Structure Determination Methodology

experimental (30,946)

Scientific Name of Source Organism

- Homo sapiens (30,946)**
- synthetic construct (1,217)
- Mus musculus (818)
- Hirudo medicinalis (279)
- Human immunodeficiency virus 1 (209)
- Rattus norvegicus (191)
- Severe acute respiratory syndrome coronavirus 2 (168)
- Escherichia coli (136)
- Bos taurus (109)
- Lama glama (102)
- [More...](#)

Taxonomy

- Eukaryota (30,882)
- other sequences (1,238)
- Riboviria (696)
- Bacteria (608)
- Eukaryota (eukaryotes) (526)
- Duplodnaviria (181)
- unclassified sequences (57)
- Monodnaviria (38)
- Varidnaviria (31)
- Archaea (28)
- [More...](#)

Experimental Method

- X-RAY DIFFRACTION (30,946)**
- NEUTRON DIFFRACTION (8)
- SOLUTION NMR (1)

Рис. 1.12. Уточнение результатов поиска

Чтобы избежать выдачи ненужных структур других протеинов, уточним название белка. В поле **Advanced Search Query Builder** в разделе **Structure Attribute** нажмите **Add Attribute** и в появившемся поле вбейте «**Macromolecule Name**», установите **has exact phrase** в качестве логического оператора, а в поле значения атрибута напечатайте «**COAGULATION FACTOR XA**», затем нажмите **Search** (рис. 1.13).

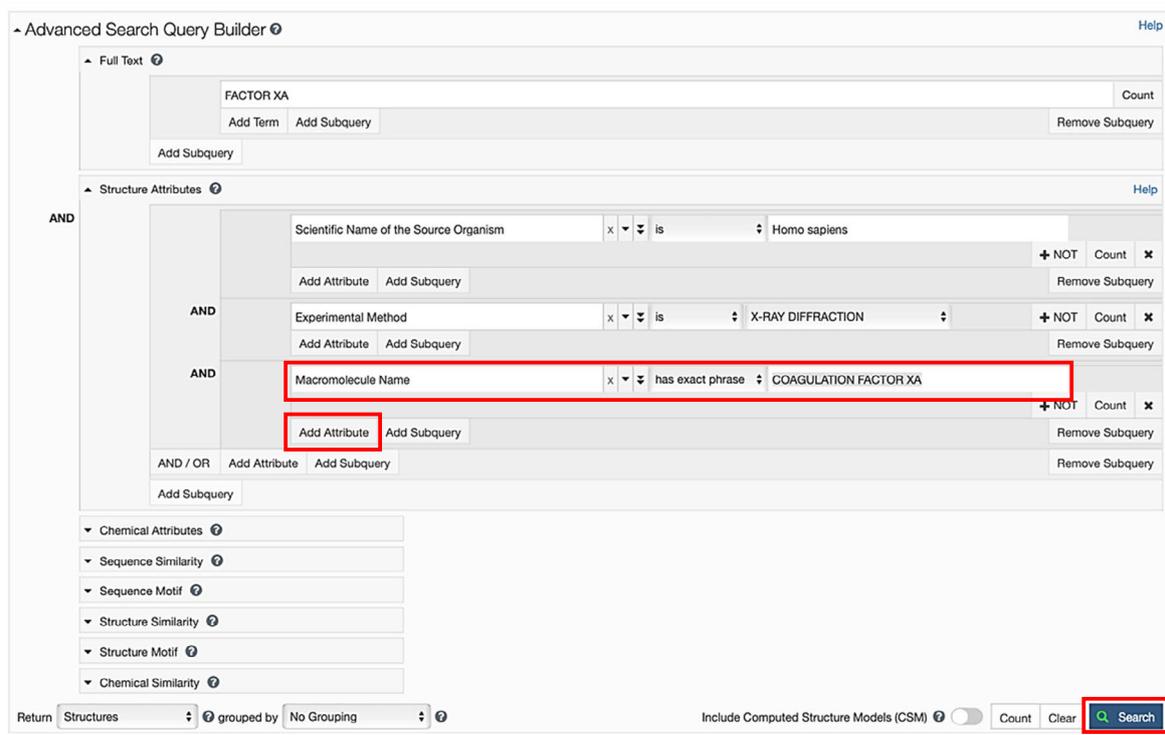


Рис. 1.13. Инструмент расширенного поиска

Важно: данная лабораторная работа будет посвящена подготовке структуры фактора свертывания крови Ха (PDB ID: **1FJS** [17]), активатора тромбина. Фактор Ха, содержащийся в данной записи PDB, включает ингибитор ZK-807834 (CI-1031) в своем активном сайте. Выбранный комплекс определен методом PCA, характеризуется приемлемыми значениями разрешения (1.75 Å) и R_{free} -фактора (0.26).

1.1.2. Создание проекта и указание рабочей директории

Создайте папку с именем LW1 в своем личном каталоге. Дважды щелкните по ярлыку Maestro на рабочем столе. Любой запуск начинается со Scratch Project (проекта по умолчанию); все расчеты и результаты выполнения заданий, произведенные в Scratch Project, не сохраняются, поэтому начинать каждую сессию Maestro следует с указания пути к своей рабочей директории и создания/открытия своего проекта. Проекты Maestro имеют расширение .prj, а в рабочей директории хранятся все импортированные/экспортированные структуры, файлы заданий и т. д.

На верхней панели перейдите на вкладку **File > Change Working Directory**. В открывшемся окне укажите путь к рабочей директории LW1 в своем личном каталоге (рис. 1.14).

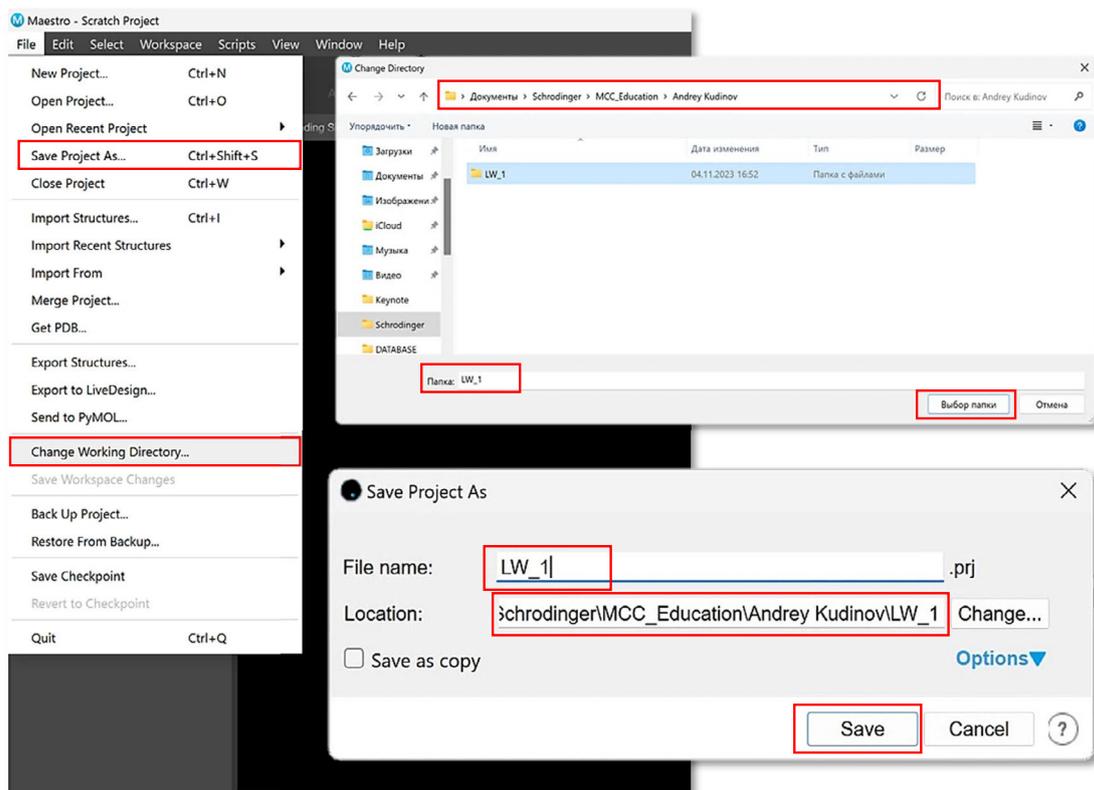


Рис. 1.14. Указание пути к рабочей директории

Затем перейдите **File > Save Project As**, задайте проекту имя LW1 в графе **File name**, укажите путь к рабочей директории в графе **Location**.

1.1.3. Знакомство с интерфейсом Maestro

Теперь, когда проект создан, директория указана и мы не рискуем потерять прогресс сессии и важные нам файлы, давайте познакомимся с рабочей средой Maestro подробнее.

Основную часть окна занимают две важные области: слева располагаются столбцы **Entry List** и **Structure Hierarchy**, а справа от них — **Workspace**. Каждой импортируемой или создаваемой структуре (будь то белок или малая молекула) соответствует отдельная запись (Entry) в **Entry List**. Нужные пользователю записи могут быть отображены средствами молекулярной графики в **Workspace** с помощью кнопки **Include**. Для того чтобы просто выбрать запись без отображения в рабочей области, нужно один раз кликнуть по соответствующей строчке. Если выбранная запись содержит несколько подструктур, как, например, в случае белок-лигандного комплекса, то в разделе **Structure Hierarchy** будут отображаться все эти подструктуры (рис. 1.15).

Ниже **Workspace** располагается окно **Sequence Viewer**, где отображается аминокислотная последовательность белка, показываемого в рабочей области (нажата кнопка **Include**). Нажатие на блок **Tasks** в верхней правой части окна откроет доступ ко всем модулям Maestro, с помощью которых можно запускать различные задания (например, молекулярный докинг или квантово-химический расчет переходного состояния). Блок **Jobs** соответствует модулю **Monitor**, через который можно наблюдать за выполнением заданий и контролировать их прогресс. **Table** ведет к **Project Table**, где отображаются не только заголовки записей проекта, но и доступные для них свойства (например, значения скоринговых функций или ADMET-параметры).

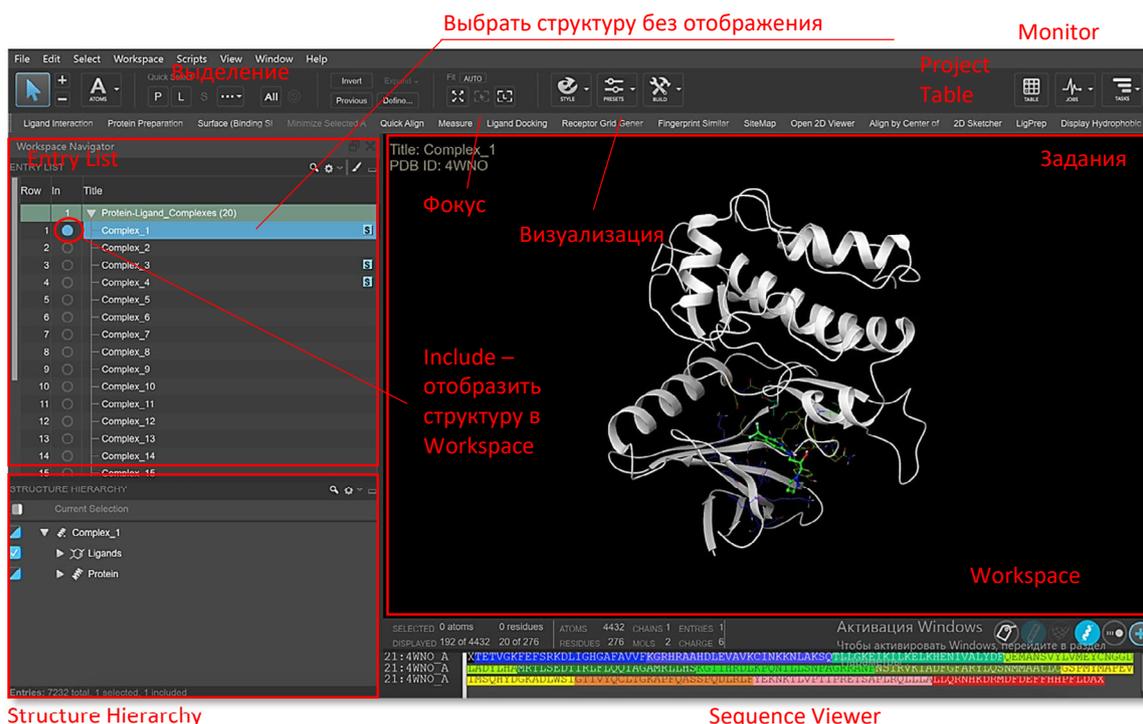


Рис. 1.15. Основные элементы рабочего окна Maestro

Панели **Style**, **Presets** и **Build** отвечают за персонализацию визуального оформления (например, изменение цвета атомов или сокрытие атомов водорода), применение predefined визуальных пресетов, а также построение/модификацию (добавление/удаление атомов и связей) отображенных в **Workspace** структур.

Панель **Fit** позволяет фокусироваться на необходимых пользователю деталях структур, например быстро приблизить камеру к лиганду, расположенному в сайте связывания белка.

Расположенная слева панель **Select** отвечает за выбор определенных элементов структуры, отображенной в **Workspace**.

Не стоит пугаться такого множества элементов интерфейса, с каждым из них мы в той или иной степени разберемся в ходе выполнения данной и всех последующих лабораторных работ.

1.1.4. Импорт структуры белка

Для того чтобы импортировать структуру белка для подготовки и дальнейшей работы, воспользуемся модулем **Protein Preparation Workflow**. Кликните на кнопку **Protein Preparation Workflow** в строке, расположенной над **Workspace** (в этой строке отображаются часто используемые и избранные модули).

Замечка: в случае необходимости вы можете импортировать отдельную цепь интересного вам белка, а также дополнительно загрузить данные об электронной плотности. Импорт происходит автоматически при подключении к сети Интернет. Однако любая структура (библиотека лигандов, файл PDB) может быть импортирована оффлайн с помощью **File > Import** (для этого необходимый файл необходимо предварительно загрузить и переместить в рабочую директорию).

Добавить любой модуль в «Избранное» вы можете самостоятельно в разделе **Task**. В открывшемся окне в разделе **Specify Protein** нажмите **Get PDB**. Введите PDB ID **1FJS** в соответствующей строке, нажмите **Download** (рис. 1.16).

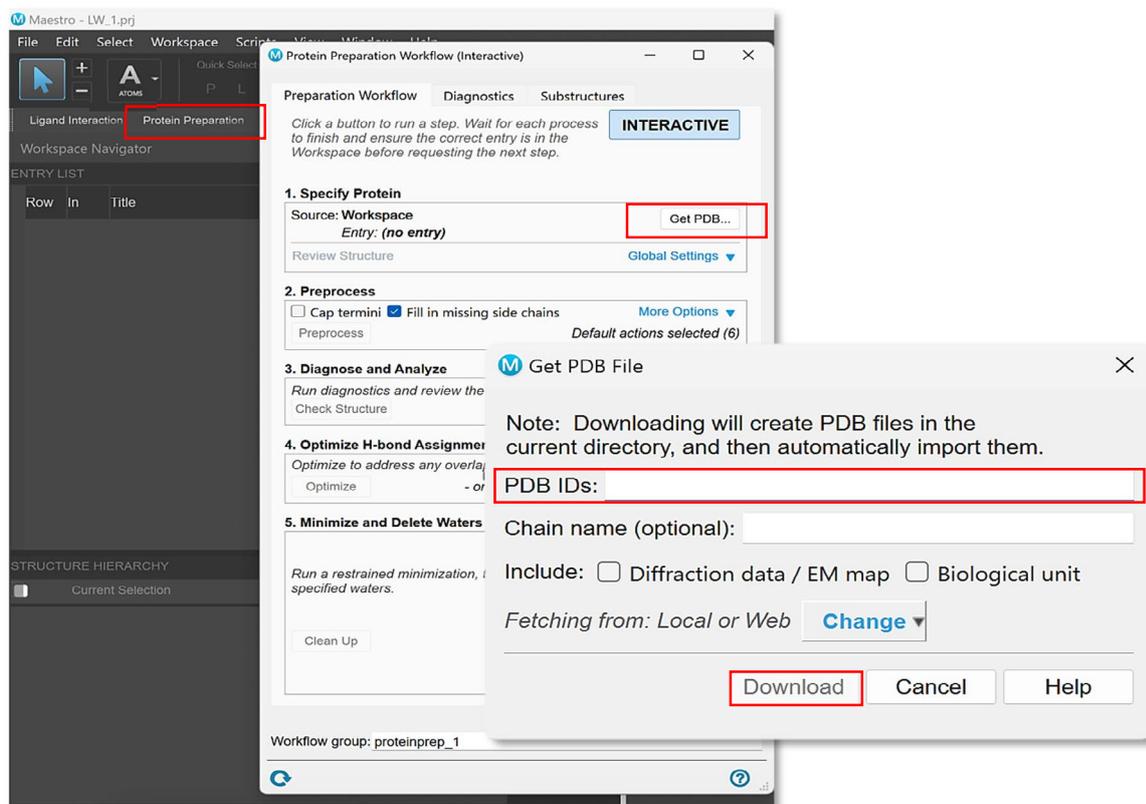


Рис. 1.16. Импорт структуры с помощью Protein Preparation Workflow

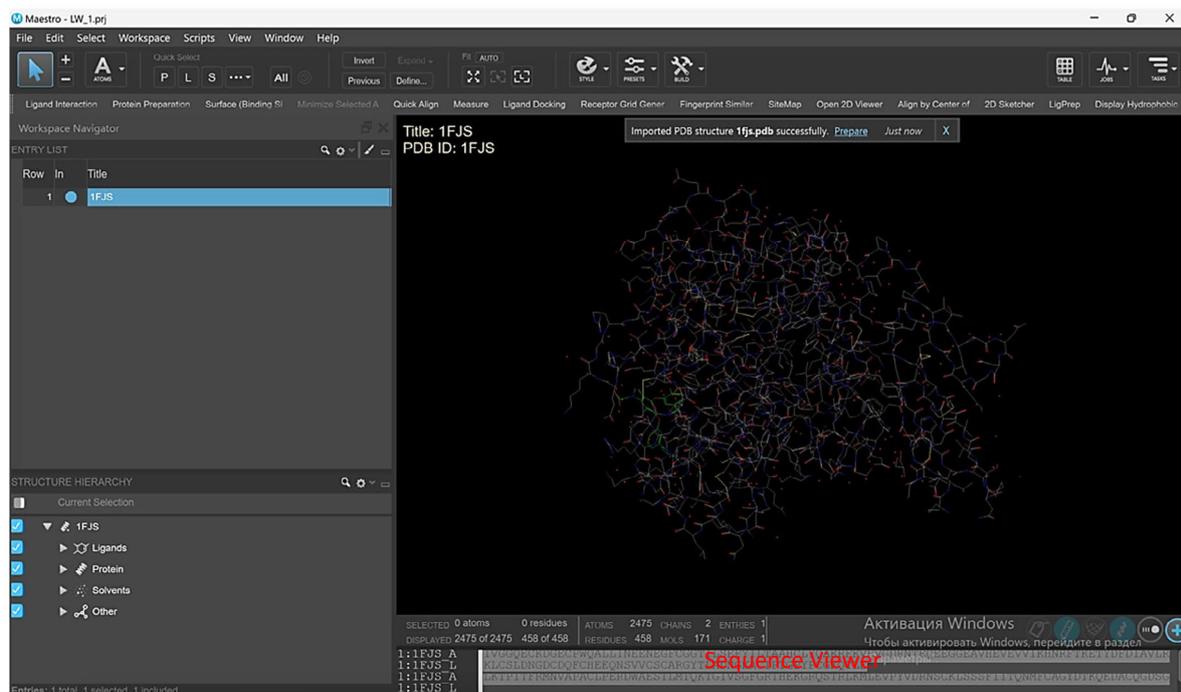


Рис. 1.17. Вид рабочего окна Maestro после импорта

В случае завершения загрузки без ошибок, структура комплекса Ха с ингибитором будет отображена в **Workspace**, а соответствующая запись появится в **Entry List**. В верхней части **Workspace** вы увидите баннер, сообщающий об успешном импорте. Обратите внимание, что в **Structure Hierarchy** теперь в виде отдельных групп содержится информация о лигандах, самом белке, молекулах воды и других атомах (ионах металлов и т. д.), которые содержатся в данной записи PDB. **Sequency Viewer** отображает последовательность фактора Ха (рис. 1.17).

1.2. Подготовка структуры белка

Панель **Protein Preparation Wizard** позволяет полностью выполнить подготовку белка для дальнейшей работы, например для выполнения молекулярного докинга. В случае нажатия кнопки **Interactive** подготовка пройдет в автоматическом режиме, однако мы не рекомендуем использовать его при работе с ранее неизвестным белком.

В разделе **Preprocess** раскройте меню **More Options**. Убедитесь, что напротив пунктов **Fill in missing side chains** (добавление отсутствующих атомов в боковых цепях), **Assign bond orders using CCD database** (уточнение порядков связей с помощью CCD (*англ.* Chemical Component dictionary)) [18], **Replace hydrogens**, **Create zero-order bonds to metals** (создание корректных координационных связей металлов), **Create disulfide bonds** (создание дисульфидных связей) проставлены галочки. Также проверьте, чтобы был выбран пункт **Generate het states**, а диапазон **pH** соответствовал 7.4 ± 2 . Последний пункт отвечает за предсказания состояния протонирования гетероатомов с помощью алгоритма Erik [19]. Пункт **Fill in missing loops** позволит предсказать конформацию отсутствующих коротких петель в структуре белка с использованием Prime [20], требует дополнительного времени для выполнения расчета и обычно не обязателен, но в этой лабораторной работе выберем также и его, поставив галочку рядом. Затем нажмите на кнопку **Preprocess**. Расчет займет ориентировочно 1–5 минут.

По завершении препроцессинга вы увидите, что теперь в вашу структуру добавлены отсутствующие ранее атомы водорода, а двойным/координационным связям назначен корректный порядок. В **Entry List** была автоматически создана группа **1FJS-proteinprep_1** и в ней сохранена соответствующая структура **1FJS - 2-preprocessed** (рис. 1.18).

Далее необходимо проверить структуру после начального этапа подготовки. В разделе **Diagnose and Analyze** нажмите на кнопку **Check Structure**. Произойдет автоматическое переключение на вкладку **Diagnostics**, в которой отображаются все обнаруженные в структуре белка проблемы.

Заметка: нарушения валентных возможностей атомов отображаются во вкладке **Valences**, отсутствующие боковые цепи и остатки основной цепи – в **Missing**, проблемы с чрезмерным перекрыванием атомных радиусов (*англ.* clashes) – в **Overlapping**, а список остатков, для которых возможны альтернативные ориентации – в **Alternates**.

Как мы можем увидеть, в нашей структуре присутствует только одна проблема: атомы водорода остатков ASN117 и THR136 находятся на расстоянии ближе допустимого. Нажатие на строчку приведет к тому, что в **Workspace** вы сможете сфокусироваться на этих проблемных атомах.

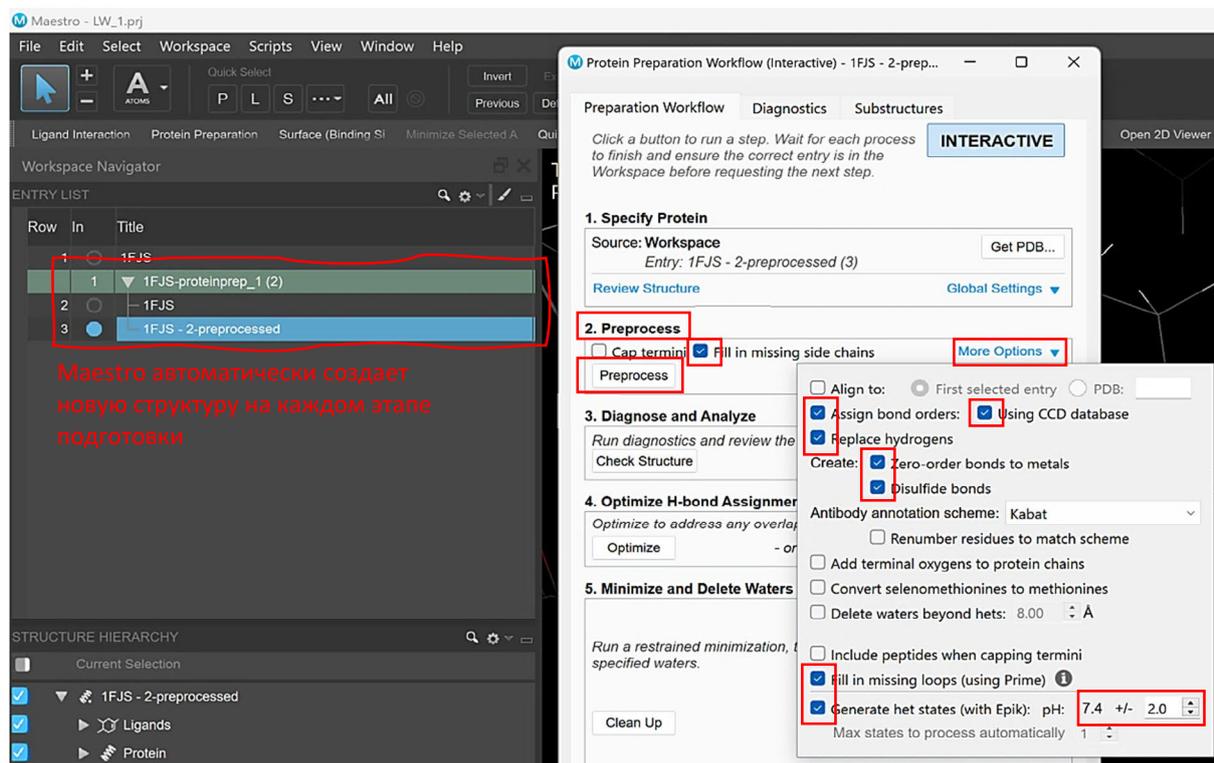


Рис. 1.18. Preprocess. Начальный этап подготовки белка

Важно: проблемы перекрывания атомных радиусов атомов водорода чаще всего не критичны и решаются на дальнейших этапах подготовки.

Из вкладки **Overlapping** перейдите на вкладку **Reports** и выберите **View: Ramachandran Plot**. Ознакомьтесь с картой Рамачандрана и сделайте выводы о том, какие элементы вторичной структуры присутствуют в данном белке. Закройте окно **Ramachandran Plot** (рис. 1.19).

Теперь перейдите на вкладку **Substructures**, чтобы подробнее ознакомиться со всем содержимым структуры PDB, которая была загружена. В верхней части окна вы можете ознакомиться со всеми малыми молекулами и ионами (напротив тех из них, которые являются лигандами, стоит соответствующая отметка), в левой нижней содержится информация о кристаллографической воде, а в правой нижней – о цепях белка.

Заметка: для каждой малой молекулы в панели **Ligands, Metals, Other** может быть сгенерировано несколько вариантов в зависимости от состояния протонирования или таутомерии. По умолчанию выбирается состояние, обладающее наибольшей стабильностью.

Однако вы можете выбрать любое из них или же сохранить сразу несколько вариантов для дальнейшего моделирования. Сейчас мы продолжим работать с первым (выбранным по умолчанию) состоянием лиганда.

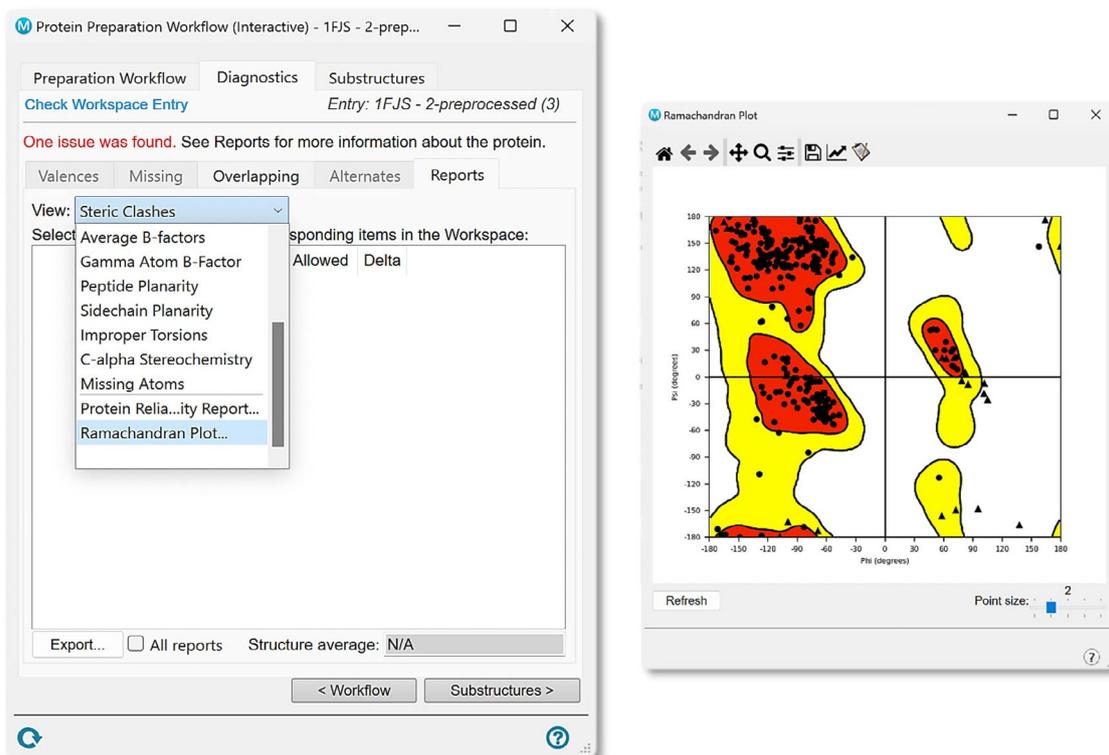


Рис. 1.19. Получение карты Рамачандрана

Молекулы глицерина (GOL), а также иона хлора (CL) явно являются кристаллографическими артефактами, никак не влияют на биологическую активность фактора Ха и могут привести к неверным результатам дальнейшего моделирования, так что должны быть удалены.

Выберите их путем нажатия ЛКМ с зажатым **Ctrl** (строчки будут подсвечены желтым), а затем нажмите **Delete from Entry**. Таким же образом поступим и с цепью L фактора Ха. При нажатии на каждую строчку на вкладке **Substructures** фокус в **Workspace** будет автоматически смещаться для удобного осмотра выбранной подструктуры (рис. 1.20).

Важно: будьте осторожны при удалении подструктур. Часто белки, депонированные в PDB, содержатся в одной элементарной ячейке кристалла в виде ди-, три-, тетра- или олигомеров, в каждой из цепей которых содержится один и тот же лиганд. Работа с такими *избыточными* структурами может занимать много машинного времени, поэтому оставляют только одну цепь. Однако необходимо четко понимать структурные особенности, механизм действия и функции мишеней при удалении цепей! Например, если вы разрабатываете ингибитор PPI и работаете с гетеродимером, удаление какой-либо цепи может быть однозначно опрочетливо. То же самое касается подготовки структур для белок-белкового докинга. Изучите информацию о белке еще до его загрузки из PDB.

Нажмите на зеленую кнопку **Workflow** для возврата к вкладке **Preparation Workflow** и перейдите в раздел **Optimize H-Bond Assignments**. Перекрытие радиусов, вызванное добавлением атомов водорода на этапе предварительной обработки, будет исправлено, а ориентация боковых цепей скорректирована при запуске этого задания.

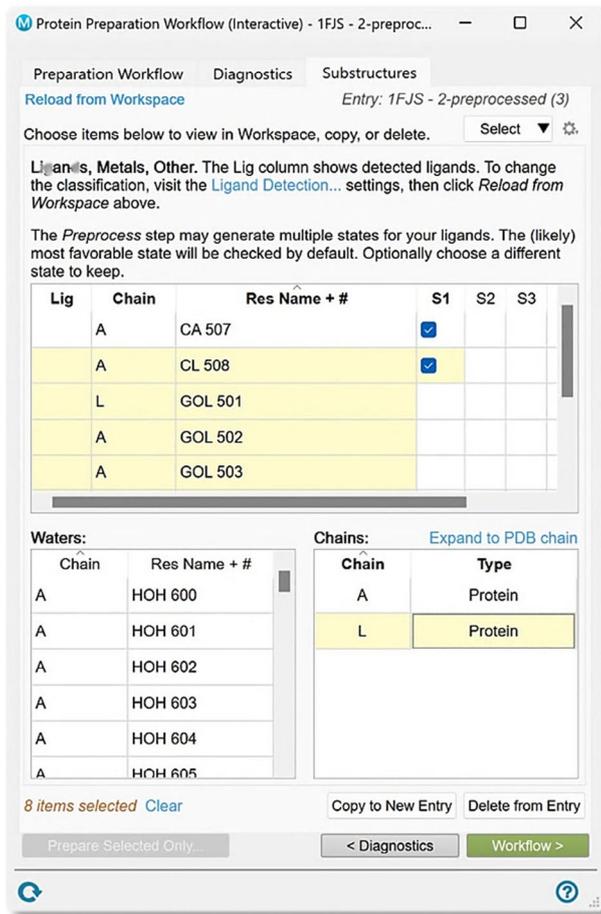


Рис. 1.20. «Очистка» структуры

Проверьте рН для оптимизации: значение должно быть зафиксировано в диапазоне рН, выбранном на этапе предварительной обработки. Нажмите **Optimize**. Расчет займет 3–6 минут (рис. 1.21).

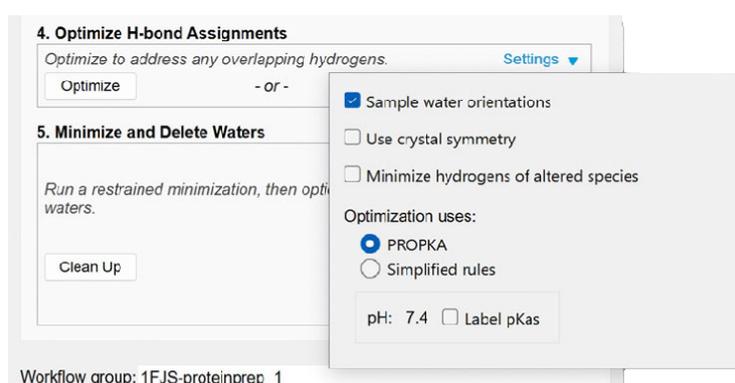


Рис. 1.21. Оптимизация сетки водородных связей

Последний этап включает ограниченную релаксацию структуры белка и удаление молекул кристаллографической воды. Перейдите в раздел **Minimization and Delete Waters**, проверьте, чтобы настройки были такими же, как указано на рис. 1.22. Запустите процесс. Расчет займет 4–6 минут.

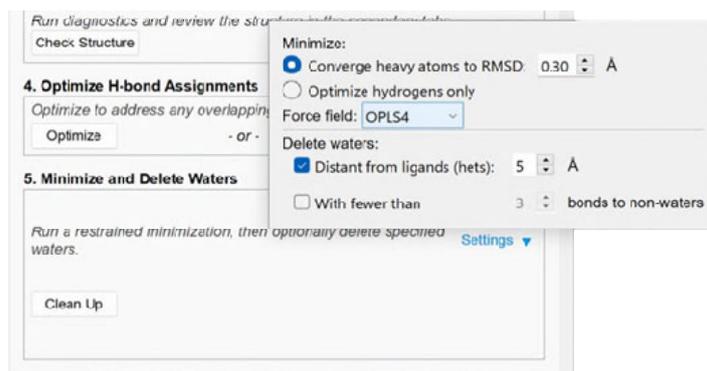


Рис. 1.22. Минимизация и удаление воды

Включите в **Workspace** структуру, полученную после минимизации. В строке избранных модулей выберите **Ligand Interaction** и дождитесь загрузки 2D-диаграммы взаимодействий лиганда в сайте связывания белка. Для доступа к легенде нажмите на значок  и выберите **LID Legend** (рис. 1.23).

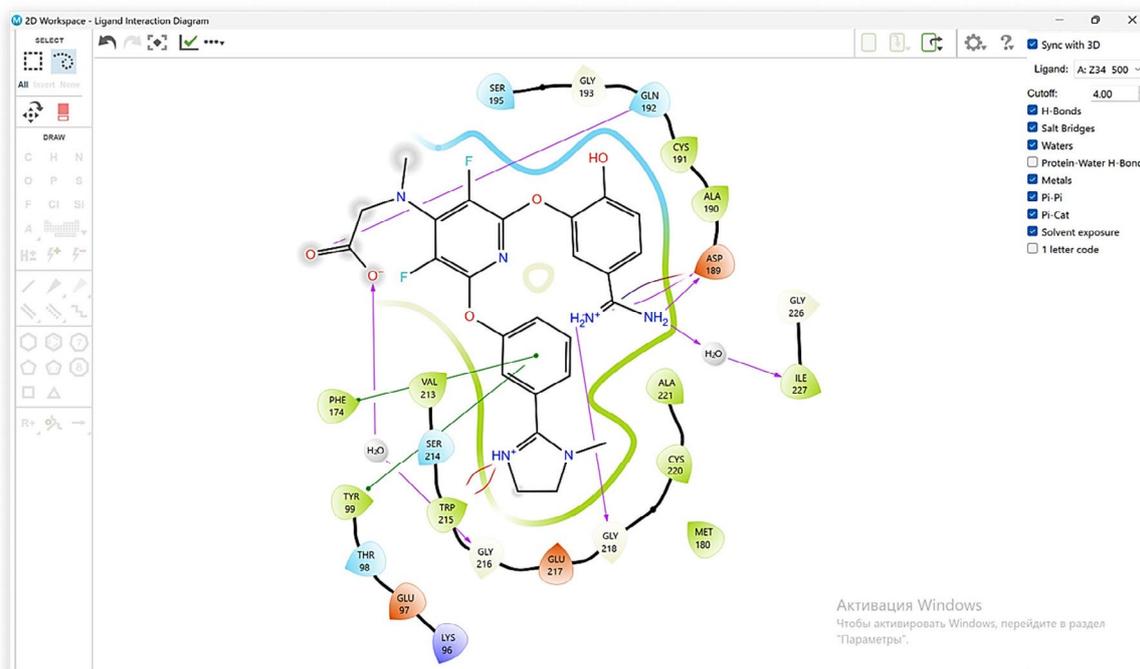


Рис. 1.23. 2D-диаграмма взаимодействий лиганда

Молекулы воды, которые участвуют в «передаче» водородных связей от лиганда к белку, в общем случае должны быть сохранены, а такой способ связывания внимательно проанализирован. Сейчас мы видим две такие молекулы и не будем их удалять. Закройте окно **Ligand Interaction Diagram**. Не забудьте также закрыть окно **Protein Preparation Workflow**.

Подробнее познакомимся со способом связывания ингибитора в сайте Ха. Выберите **Presets > Legacy > Ligand Site Cartoon**. Давайте скроем неполярные атомы водорода для облегчения визуального осмотра. В панели **Quick Select** нажмите на ..., выберите **Binding Sites+Ligands**. Обратите внимание, что теперь стала активна строка **Current Selection** в панели **Structure Hierarchy**.

Справа от **Current Selection** нажмите на значок  и в меню управления отображением атомов водорода выберите **Show Polar Hydrogens** (рис. 1.24).

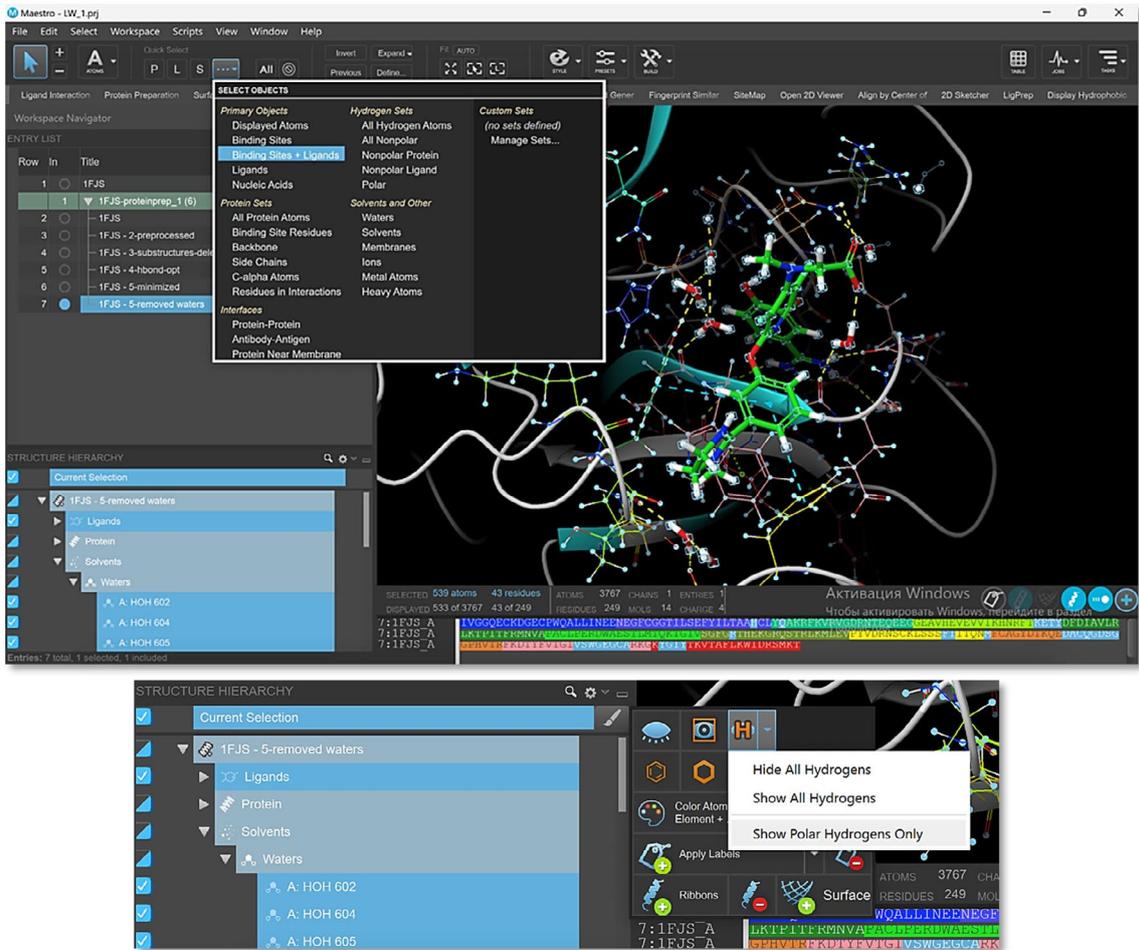


Рис. 1.24. Настройки визуального представления

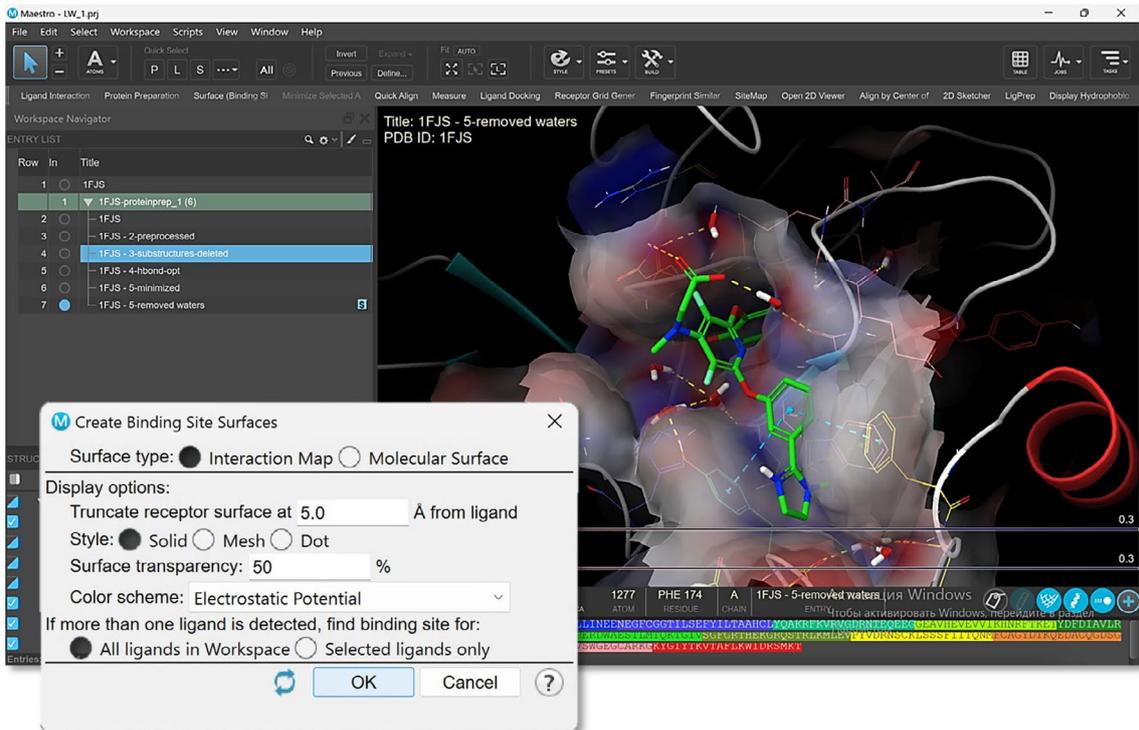


Рис. 1.25. Генерация поверхности электростатического потенциала сайта

Давайте отобразим поверхность молекулярного электростатического потенциала сайта. Выберите **Surface (Binding Site)** в строке избранных заданий, в открывшемся окне выберите **Interaction Map** напротив **Surface Type**, а также **Solid** напротив **Style**. Раскройте список рядом с **Color Scheme** и выберите **Electrostatic Potential**, затем **ОК**. Расчет займет 2–3 минуты. Когда поверхность будет отображена, изучите способ связывания лиганда в сайте. Колесико мыши может быть использовано для приближения/отдаления, а движение мышью при зажатом **Alt** – для вращения. Скрыть молекулярные поверхности всегда можно нажатием на значок **S** рядом с записью в **Entry List** (рис. 1.25).

1.3. Подготовка структуры лиганда

Нажатием **ПКМ** на строку **1FJS - 5-removed waters** в **Entry List** откройте контекстное меню, в котором выберите **Split > Into ligands, Waters, Other**. Затем включите структуру **1FJS - 5-removed waters_ligand** в **Workspace**, а в строке избранных заданий выберите **LigPrep** – модуль для подготовки лигандов (рис. 1.26) [1].

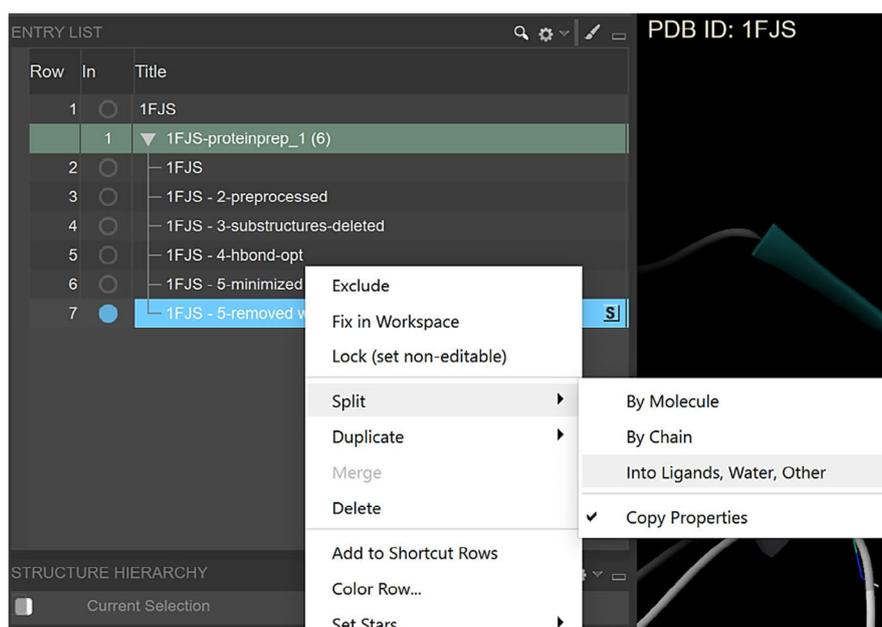


Рис. 1.26. Сплитирование структуры комплекса

Модуль **LigPrep** позволяет подготовить для работы структуры лигандов: произвести конвертацию форматов, сгенерировать наиболее выгодную конформацию, получить таутомерные формы и т. д.

В графе **Use structures from:** выберите **Workspace**. Проверьте, чтобы используемым силовым полем было именно поле **OPLS4** (англ. Optimized Potentials for Liquid Simulations) [21]. В разделе **Ionization** выберите **Generate possible states** в диапазоне **pH: 7.4±0.2**, используйте алгоритм **Epik**. Поставьте галочки напротив пунктов **Desalt** и **Generate tautomers**. Выберите пункт **Determine chiralities from 3D structure** в разделе ниже. В качестве формата вывода выберите **Maestro** (.mae). Графа **Job Name** отвечает

за название задания. Назовите задание «**ligprep_LW1**». Нажмите **Run**. Расчет займет 2–5 минут (рис. 1.27).

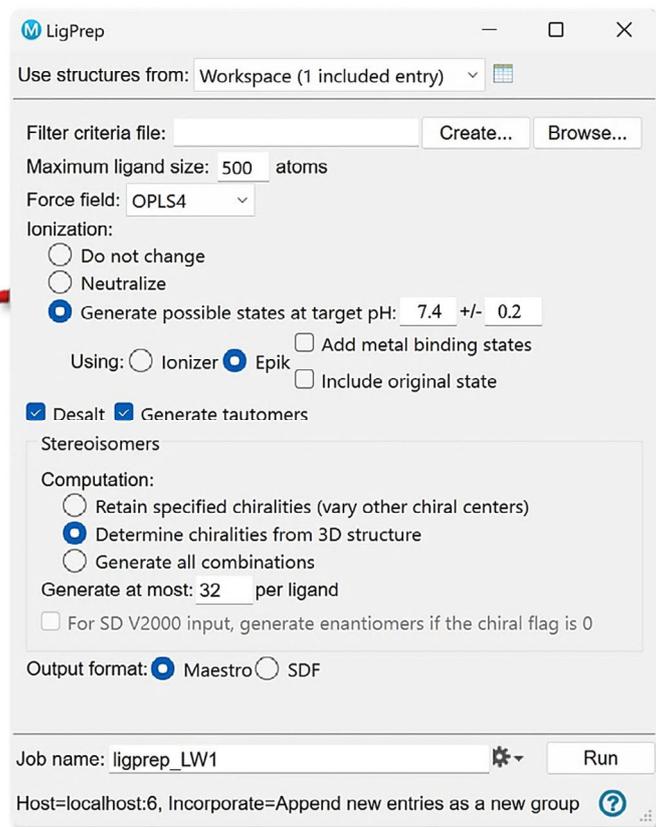


Рис. 1.27. Подготовка структуры лиганда с помощью LigPrep

Заметка: выбранные нами настройки **LigPrep** являются стандартным способом подготовки лигандов. Раздел **ionization** отвечает за назначение зарядов и учет состояний протонирования. Учет хиральности и создание стереоизомеров также возможно. Нами был выбран путь создания такой конфигурации, как указано в структурной формуле (см. модуль **2D Sketcher**). Однако в конкретном случае хиральность, очевидно, не важна: в молекуле отсутствуют центры хиральности.

В случае успешного запуска задания в панели **Jobs** появится зеленая индикация, сообщающая о наличии активной задачи. Модуль **Monitor** позволит вам контролировать выполнение задач. Двойной щелчок по строчке с нужной работой (или нажатие кнопки **Monitor**) открывает доступ ко вкладкам **Details** и **File** для конкретного задания (рис. 1.28).

Двойной щелчок по имени того или иного файла позволяет посмотреть его содержимое (доступно не для всех файлов). Полезным может являться просмотр log-файлов. Щелкните по файлу **ligprep_LW1.log** для открытия. Как показано ниже, файл log содержит всю информацию об имени задания, используемых подпрограммах, директориях, занимаемых расчетом процессорах, ошибках, а также текущей стадии.

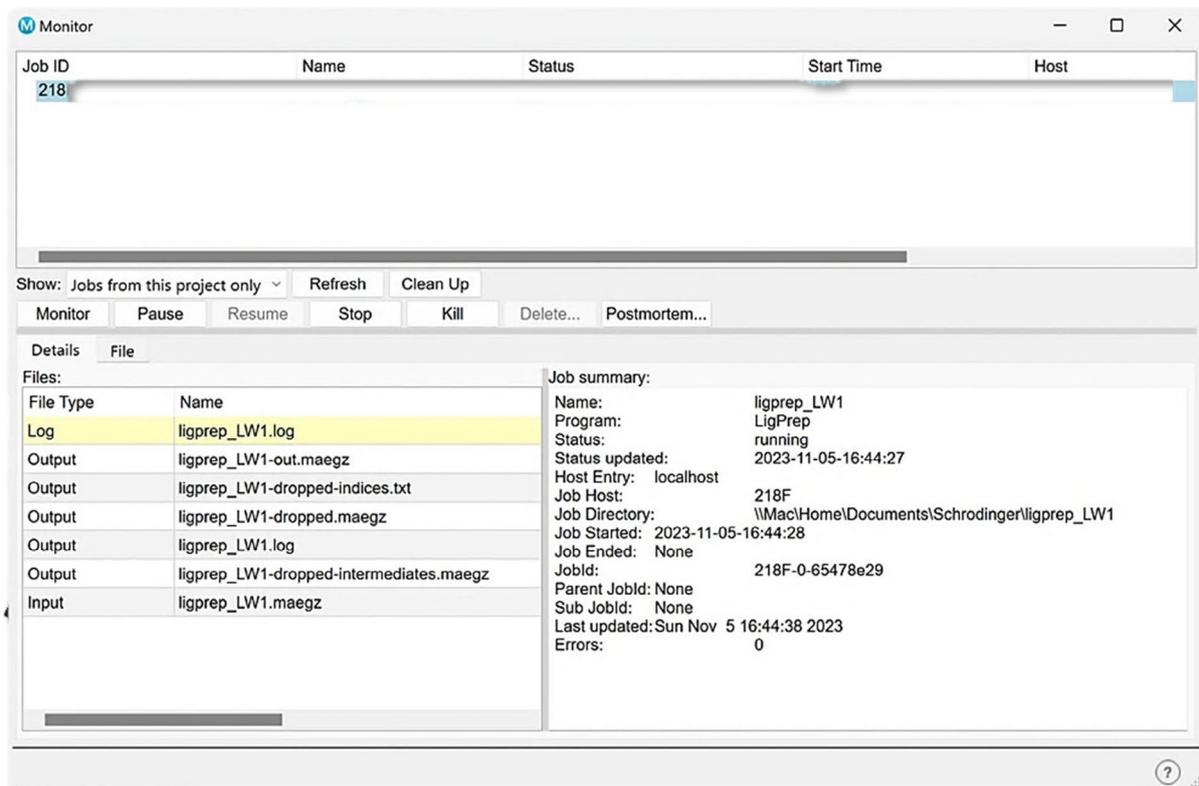


Рис. 1.28. Панель Monitor

Monitoring job: ligprep_LW1

```

-----
JobId : 218F-0-65478e29
Name : ligprep_LW1
Program : LigPrep
Version : 2023-1 build 128
Host : 218F
Dir : \\Mac\Home\Documents\Schrodinger\ligprep_LW1
HostEntry : localhost
JobHost : 218F
JobDir : C:\Users\andrejkudinov\AppData\Local\Schrodinger\tmp\ligprep_LW1
JobSchrodinger : C:\Program Files\Schrodinger2023-1
Commandline : C:\Program Files\Schrodinger2023-1\ligprep.exe -inp ligprep_LW1.inp -HOST
localhost:6 -NJOBS 1 -PROJ "\\mac\Home\Documents\Schrodinger\MCC_Education\Andrey
Kudinov\LW_1\LW_1.prj" -DISP append -VIEWNAME ligprep_gui.LigPrepPanel -TMPLAUNCHDIR
StartTime : 2023-11-05-16:44:28
-----

```

LigPrep: preparing subjobs

```

Running subjobs on hosts:
localhost (Max: 6)
Number of jobs: 1
Max retries per job: 2
Max allowed failures: unlimited
Verbosity: normal

```

Starting JobDJ...

Keep one job on driver host: False

JobDJ columns:

C: Number of completed subjobs

A: Number of active subjobs (e.g., submitted, running)

W: Number of waiting/pending subjobs

C A W | Activity JobId JobName JobHost

--- | -----

0 1 0 | launched 218F-0-65478e31 ligprep_LW1-1 localhost [218F]

После завершения расчета Maestro автоматически импортирует полученные структуры в группу **ligprep_LW1-out**. Давайте сравним нативную конформацию лиганда из белка и конформации, выданные **LigPrep**. Для этого выделите все элементы в группе **ligprep_LW1-out**. После чего нажмите на значок кастомизации в верхней правой части **Entry List**. Выберите **All Atoms** > **Color Atoms** > **Element + Custom Carbons** > **Faded Orange** (рис. 1.29).

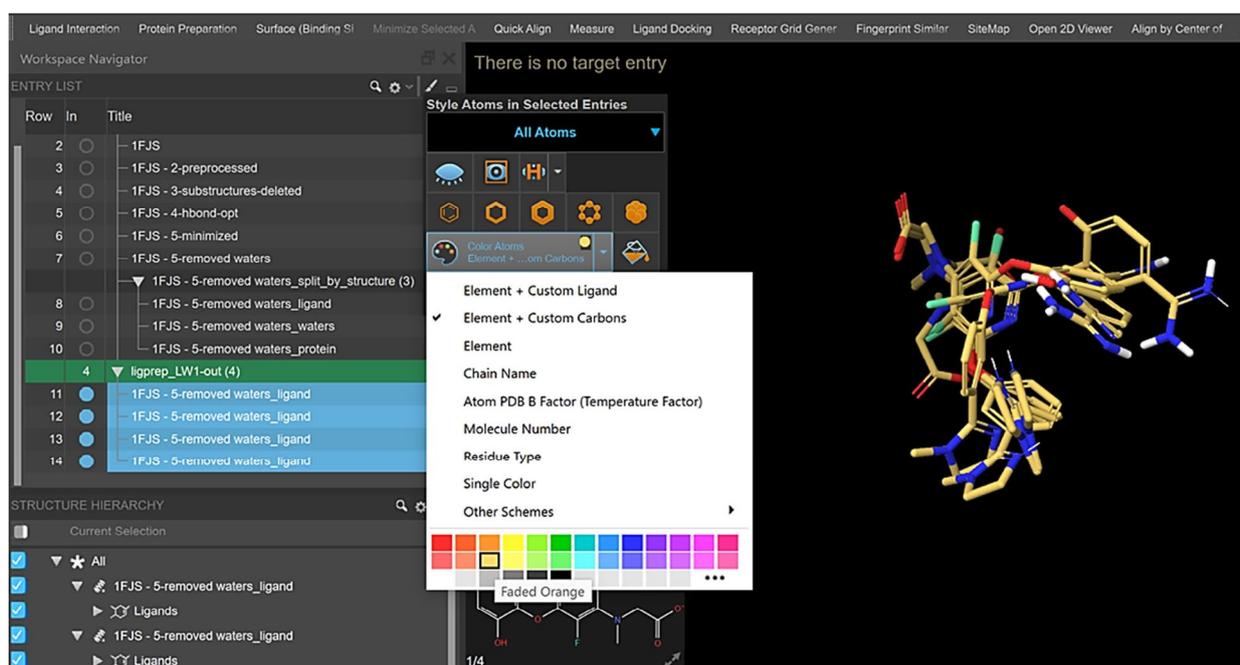


Рис. 1.29. Изменение цвета атомов углерода в молекуле

Теперь конформации надо наложить друг на друга. Выделите также **1FJS - 5-removed waters_ligand**. Перейдите в **Task** > **Superposition**. В открывшемся окне выберите нативный лиганд в качестве шаблона, а тип наложения – согласно максимальной общей подструктуре. Нажмите **Superimpose Structures**. Расчет займет примерно минуту (рис. 1.30).

Дважды щелкните по кнопке **Include** напротив строки нативного лиганда, закрепив его в **Workspace**. Используя **Ctrl**, включите остальные лиганды. Оцените, насколько иногда генерируемые методами конформационного поиска/минимизации конформации могут отличаться от биологически активной!

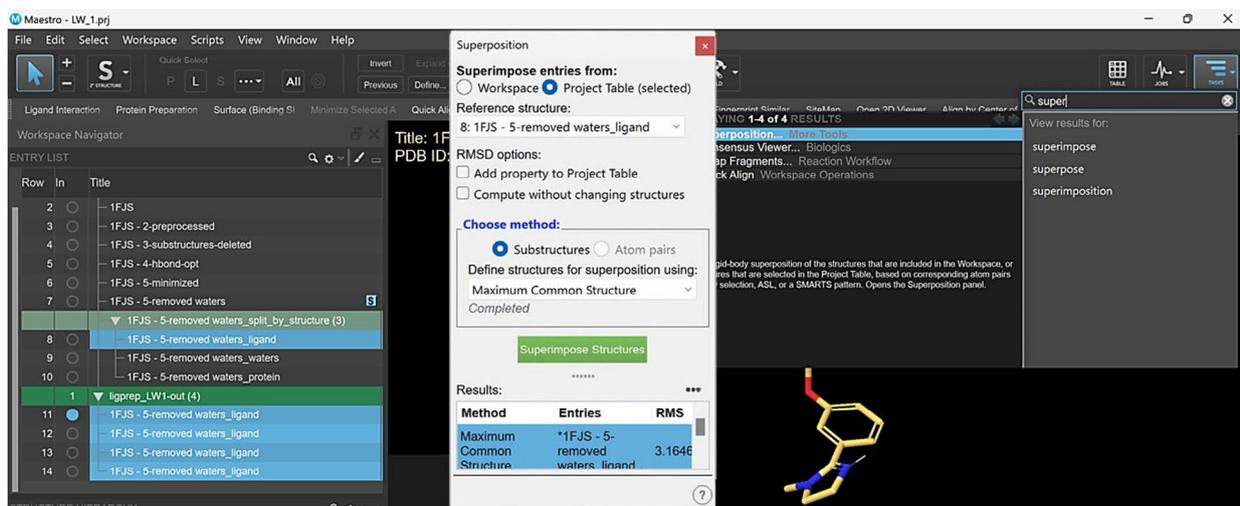


Рис. 1.30. Наложение молекул с помощью модуля Superposition

Отчет к данной лабораторной работе не требуется, однако в качестве тренировки выполните подготовку белковых структур, представленных в таблице ниже, с помощью модуля **Protein Preparation Workflow**. Вариант уточните у преподавателя (табл. 1.1).

Таблица 1.1

Комплексы для подготовки

	PDB ID, Комплекс 1	Дальнейшее использование подготовленной структуры	PDB ID, Комплекс 2	Дальнейшее использование подготовленной структуры
Вариант I	6BOY	Молекулярный докинг молекул PROTAC в комплекс POI-E3	5DXH	Поиск АТФ-конкурентных ингибиторов
Вариант II	5FL5	Разработка селективных ингибиторов CA IX	6BVB	Разработка пептидомиметика на основе дегрона HIF, связанного с E3 лигазой VHL
Вариант III	4WNP	Поиск АТФ-конкурентных ингибиторов	5UBR	Поиск АТФ-конкурентных ингибиторов

Вопросы к защите

1. Какие уровни структурной организации белков вам известны?
2. Какие методы определения трехмерной структуры белков существуют? Перечислите их основные достоинства и недостатки, область применимости.
3. Что описывает фактор R_{free} , определяемый для структур, полученных методом PCA, и чем отличается от R -фактора?
4. Какую структуру – обладающую большим или меньшим значением R -фактора – следует выбрать для работы?
5. Что такое разрешение кристаллографической структуры?
6. Какие проблемы могут быть обнаружены в структуре, полученной методом PCA?

7. Углам между какими атомами основной цепи белка соответствуют углы φ и ψ ?
8. Каким областям на карте Рамачандрана соответствуют право- и левозакрученные альфа-спирали?
9. Перечислите основные функции модуля LigPrep.
10. Перечислите основные функции модуля Protein Preparation Wizard.
11. Почему не всегда стоит удалять «лишние» цепи из структуры белков?
12. Какие молекулы воды следует сохранять при подготовке белково-лигандных комплексов?
13. Какие свойства атомов белка характеризует В-фактор?

Рекомендуемая литература

1. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments /G. M. Sastry, M. Adzhigirey, T. Day [et al.] //Journal of Computer-Aided Molecular Design. – 2013. – Vol. 27, № 3. – P. 221–234.
2. Schrödinger Release 2023-4: Protein Preparation Wizard ; Epik. – New York : Schrödinger, 2023.
3. Schrödinger Release 2023-4: LigPrep. – New York : Schrödinger, 2023.
4. Drenth, J. Principles of Protein X-Ray Crystallography /J. Drenth. – 3rd Ed. – New York : Springer, 2007. – XIV, 332 p. – ISBN 978-0-387-33334-2.
5. Protein Crystals – 1g vs Microgravity : image article //NASA : [сайт] /Page Ed. D. Mas-sengill. – URL: www.nasa.gov/image-article/protein-crystals-1g-vs-microgravity/ (дата обращения: 30.09.2025).
6. The New York Consortium on Membrane Protein Structure (NYCOMPS): A high-throughput platform for structural genomics of integral membrane proteins /J. Love, F. Mancía, L. Shapiro [et al.] //Journal of Structural and Functional Genomics. – 2010. – Vol. 11, № 3. – P. 191–199.
7. Drug Design : Structure- and Ligand-Based Approaches /Edited by M. M. Kenneth [et al.]. – Cambridge : Cambridge University Press, 2010. – X, 276 p. – ISBN 978-0-521-88723-6.
8. Kleywegt, G. J. Model Building and Refinement Practice /G. J. Kleywegt, T. A. Jones //Methods in Enzymology. – 1997. – Vol. 277. – P. 208–230.
9. Protein NMR Spectroscopy : Principles and Practice /J. Cavanagh, W. J. Fairbrother, A. G. Palmer III [et al.]. – 2nd Ed. – Amsterdam [et al.] : Elsevier [et al.], 2007. – XXV, 885 p. – ISBN 978-0-121-64491-8.
10. Structural insight into SUMO chain recognition and manipulation by the ubiquitin ligase RNF4 /Y. Xu, A. Plechanovová, P. Simpson [et al.] //Nature Communications. – 2014. – Vol. 5. – Article number 4217.
11. Benjin, X. Developments, applications, and prospects of cryo-electron microscopy /X. Benjin, L. Ling //Protein Science. – 2020. – Vol. 29, № 4. – P. 872–882.
12. Pan, X. Recent advances in de novo protein design: Principles, methods, and applications /P. Xingjie, T. Kortemme //Journal of Biological Chemistry. – 2021. – Vol. 296. – Article number 100558.
13. CCDC. Structural Chemistry Data, Software, and Insights : [сайт]. – URL: www.ccdc.cam.ac.uk/ (дата обращения: 20.12.2024).
14. RCSB PDB : Protein Data Bank : [сайт]. – URL: www.rcsb.org/ (дата обращения: 09.11.2023).
15. UniProt : [сайт]. – URL: www.uniprot.org/ (дата обращения: 09.11.2024).

16. Molecular Modeling : Basic Principles and Applications / H.-D. Holtje, W. Sippl, D. Rognan, G. Folkers. – 3rd, Revised and Expanded Ed. – Weinheim : Wiley-VCH, 2008. – X, 310 p. – ISBN 978-3-527-31568-0.
17. Preparation, Characterization, and the Crystal Structure of the Inhibitor ZK-807834 (CI-1031) Complexed with Factor Xa / M. Adler, D. D. Davey, G. B. Phillips [et al.] // Biochemistry. – 2000. – Vol. 39, № 41. – P. 12534–12542.
18. The Chemical Component Dictionary: Complete Descriptions of Constituent Molecules in Experimentally Determined 3D Macromolecules in the Protein Data Bank / J. D. Westbrook, C. Shao, Z. Feng [et al.] // Bioinformatics. – 2015. – Vol. 31, № 8. – P. 1274–1278.
19. Epik: a software program for pKa prediction and protonation state generation for drug-like molecules / J. C. Shelley, A. Cholleti, L. L. Frye [et al.] // Journal of Computer-Aided Molecular Design. – 2007. – Vol. 21, № 12. – P. 681–691.
20. What do all the Prime MM-GBSA energy properties mean? // Schrödinger : [сайт]. – URL: www.schrodinger.com/kb/1875 (дата обращения: 15.11.2024).
21. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space / C. Lu, C. Wu, D. Ghoreishi [et al.] // Journal of Chemical Theory and Computation. – 2021. – Vol. 17, № 7. – P. 4291–4300.



Лабораторная работа 2

КОНФОРМАЦИОННЫЙ АНАЛИЗ МАЛЫХ МОЛЕКУЛ

Цель работы: освоение основных подходов к конформационному поиску малых молекул.

Вам предстоит: выполнить систематический и стохастический поиск конформаций малых молекул с помощью таких инструментов, как MacroModel и ConfGen, а также проанализировать результаты.

Теоретическое введение

Конформационный поиск — это важный метод в молекулярном моделировании, который направлен на генерацию различных конформаций молекулы (исследование конформационного пространства), определение их энергии и нахождение низкоэнергетических конформеров. Области применения конформационного поиска представлены ниже:

- **определение наиболее энергетически стабильной конформации:** конформационный поиск позволяет выявить набор самых энергетически стабильных конформаций. Это полезно для понимания фундаментальных свойств структуры и определения ее активности;
- **определение способа связывания лиганда с мишенью:** конформационный поиск может использоваться для определения оптимальной ориентации/конформации лиганда в связанном состоянии с белком в ходе молекулярного докинга;
- **уточнение при определении структуры белка:** конформационный поиск играет важную роль в уточнении результатов определения структуры макромолекул такими методами, как рентгеноструктурный анализ или ядерно-магнитный резонанс. Сравнение экспериментальных данных с конформациями, найденными с помощью конформационного поиска, помогает сопоставить экспериментальные и модельные структуры.

Зависимость энергии молекулы от тех или иных геометрических параметров атомов удобно представлять графически в виде кривой или поверхности потенциальной энергии. В качестве таких геометрических параметров могут выступать декартовы (x , y , z) или так называемые внутренние координаты (значения длин связей, валентных и диэдральных углов). На рис. 2.1 представлены примеры кривых и поверхностей потенциальной энергии различных молекул. Как вы можете увидеть, реальный энергетический ландшафт молекулы довольно сложен и характеризуется большим числом стационарных точек — наиболее интересными для нас будут являться минимумы энергии.

По способу «перемещения» в пространстве конформаций выделяют методы **систематического** и **стохастического** поиска.

Процедура **систематического** поиска заключается в последовательном изменении какого-либо геометрического параметра, чаще всего вращение вокруг торсионных углов с заданным шагом. Если он достаточно мал, то достигим полный охват конформационного пространства.

Однако расчет конформационного пространства для структуры, содержащей всего шесть вращающихся связей, предполагает создание и минимизацию около трех миллионов конформаций с учетом шага в 30° . Данная ситуация называется «комбинаторным взрывом», а такой конформационный поиск стал бы чрезвычайно долгой и затратной процедурой [1–3].

В ходе **стохастического** поиска новые конформации создаются путем случайного изменения геометрических параметров (внутренних координат). Энергия каждой новой конформации сравнивается с энергией предыдущей. Если энергия ниже, то новая структура становится «текущей» и минимизируется, в противном случае вероятность принятия конформации в качестве «текущей» пропорциональна $\exp\left[-\frac{\Delta E}{kT}\right]$. Таким образом алгоритм моделирует тепловое поведение молекулы, стремясь отдавать предпочтение стабильным конформерам, но допуская случайные флуктуации в состояния с более высокой энергией, что затем может привести к глобальному минимуму. Обычно используются эвристические подходы, такие как генетические алгоритмы или метод Монте-Карло. Он быстрее, чем систематический поиск, но не гарантирует полный обход всего пространства конформаций (рис. 2.2) [1].

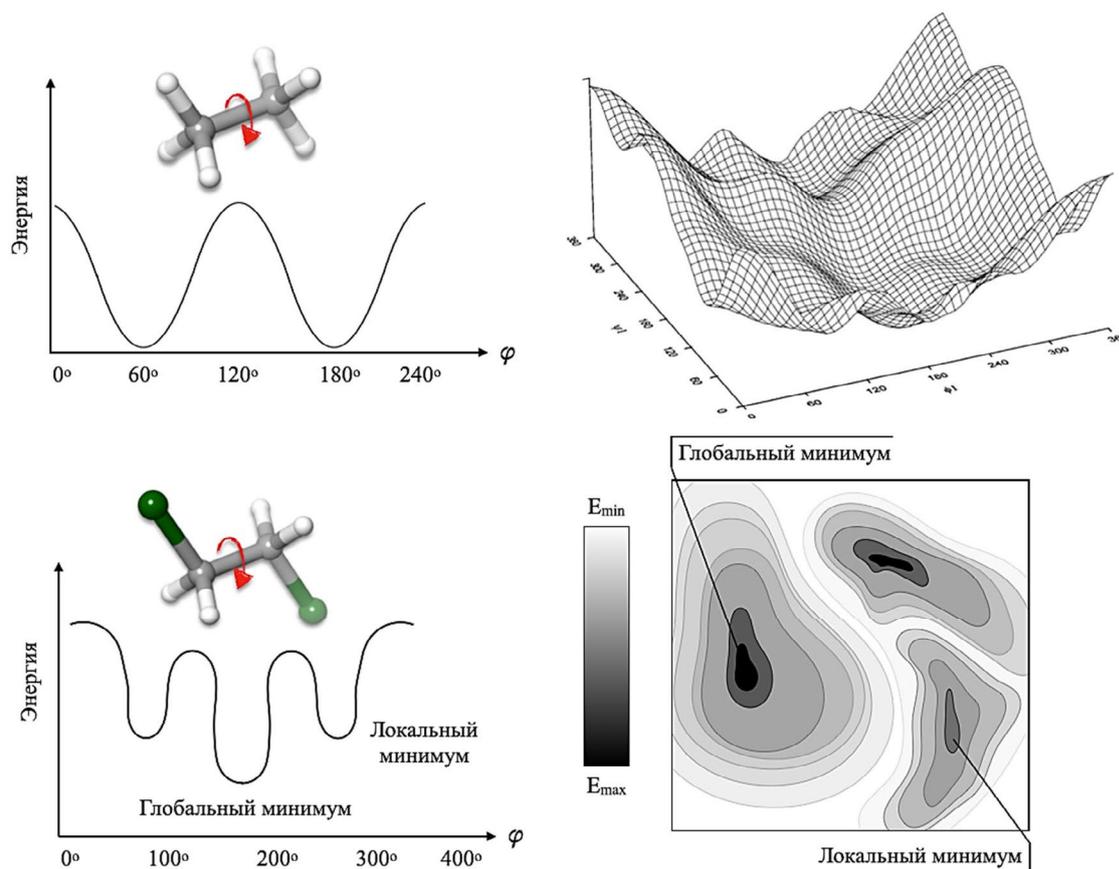


Рис. 2.1. Кривые и поверхности потенциальной энергии

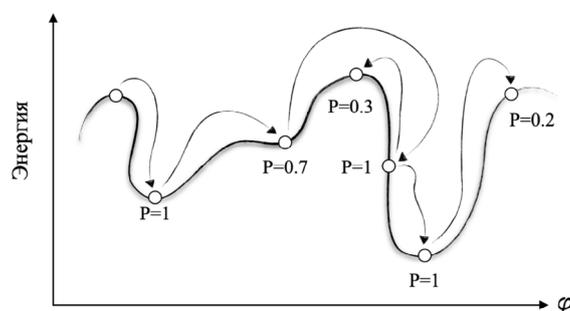


Рис. 2.2. Стохастический поиск

В данном практикуме не будут подробно описаны такие методы, как моделирование отжига, генетические алгоритмы и геометрия расстояний. Настоятельно рекомендуем вам ознакомиться с материалами лекций и источниками из списка литературы к данной работе, чтобы подробнее узнать об этих подходах.

Большое значение в вопросе конформационного поиска играют также подходы к минимизации энергии конформаций. Весь процесс минимизации сводится к простой математической задаче – поиску точек минимума функции потенциальной энергии $U(q)$, где q – геометрические параметры системы.

Метод скорейшего спуска относится к градиентным методам минимизации. В общем случае градиент некоторой функции $\varphi(x, y, z)$ является вектором, направление которого совпадает с направлением возрастания функции, а его компоненты равны частным производным по переменным.

Иногда градиент удобно представлять в матричном виде \mathbf{g} (особенно актуально для метода сопряженных градиентов):

$$\mathbf{q} = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_p \end{pmatrix} \quad \mathbf{g} = \begin{pmatrix} \frac{\partial U}{\partial q_1} \\ \frac{\partial U}{\partial q_2} \\ \vdots \\ \frac{\partial U}{\partial q_p} \end{pmatrix}.$$

Движение против градиента, соответственно, приводит к уменьшению значения функции. На этом принципе и основан метод скорейшего спуска. Алгоритм производит: (1) вычисление энергии начальной конфигурации p_0 ; (2) расчет градиента в данной точке; (3) линейный поиск против градиента в попытках найти точку с минимальным значением энергии p_1 путем изменения значений координат атомов (декартовых или внутренних); (4) расчет градиента уже в точке p_1 и т. д. (рис. 2.3).

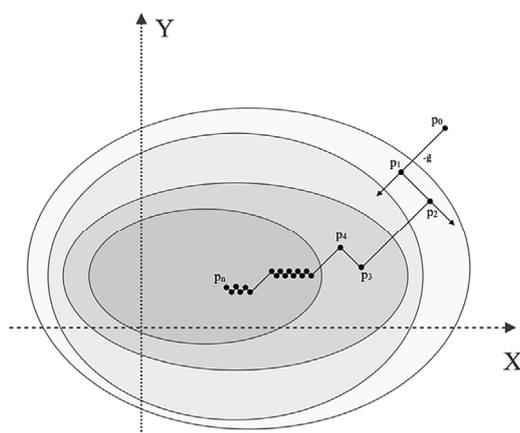


Рис. 2.3. Градиентный спуск

Метод наискорейшего спуска интуитивно понятный, не требовательный с аппаратной точки зрения, однако обладает плохой сходимостью в окрестностях минимума. Метод **сопряженных градиентов** является современной модификацией наискорейшего спуска и сейчас повсеместно используется в соответствующем ПО; суть его заключается в уточнении направления поиска в текущей точке относительно направления на предыдущем шаге:

$$\mathbf{V}^{(k)} = -\mathbf{g}^{(k)} + \gamma^{(k)} \mathbf{V}^{(k-1)},$$

где k – текущий шаг, $\mathbf{V}^{(k)}$ – направление на текущем шаге k , $\mathbf{g}^{(k)}$ – градиент, рассчитанный в точке $p^{(k)}$, $\mathbf{V}^{(k-1)}$ – направление на предыдущем шаге, а $\gamma^{(k)}$ – корректирующий коэффициент скаляр, рассчитанный как

$$\gamma^{(k)} = \frac{(\mathbf{g}^{(k)})^T \mathbf{g}^{(k)}}{(\mathbf{g}^{(k-1)})^T \mathbf{g}^{(k-1)}}.$$

Метод сопряженного градиента позволяет достичь минимума за кратно меньшее число шагов, чем наискорейший спуск (рис. 2.4).

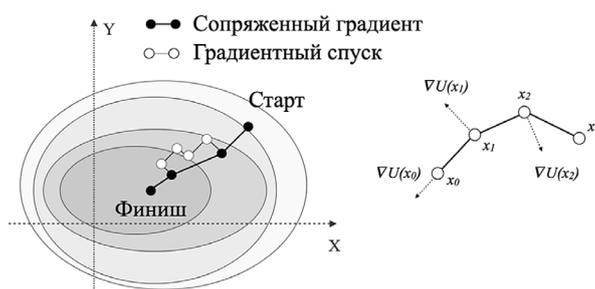


Рис. 2.4. Метод сопряженных градиентов в сравнении с наискорейшим спуском

Метод **Ньютона-Рафсона** позволяет приблизиться к минимуму с помощью вычисления первых и вторых частных производных. Как мы уже знаем, взятие частных производных энергии по координатам соответствует операции градиента, а компоненты градиента удобно размещаются в матрице \mathbf{g} ; матрицу же вторых производных размерности $p \times p$ (где p – число переменных) принято называть Гессианом \mathbf{H} [4]:

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - (\mathbf{H}^{(k)})^{-1} \mathbf{g}^{(k)}.$$

Конформационный поиск в Maestro представлен модулями MacroModel [5, 6] и ConfGen [7, 8]. Задание MacroModel может быть запущено из **Task > MacroModel > Conformational Search**. Алгоритм MacroModel включают в себя следующие методы поиска (табл. 2.1).

Таблица 2.1

Методы поиска, реализованные в MacroModel

Метод поиска	Комментарий
Monte Carlo Multiple Minimum (англ. МСММ, Стохастический поиск множественных минимумов методом Монте-Карло)	Данный метод очень эффективен при выполнении глобального поиска. В ходе расчета производятся случайные изменения торсионных углов. Хотя количество углов, изменяющихся при поиске, не ограничено, наличие их в количестве 10 или 15 значительно увеличивает сложность/время и ухудшает сходимость поиска [9]
Systematic Pseudo-Monte Carlo (англ. SPMC, Систематический поиск методом псевдо-Монте-Карло)	Если задача состоит в том, чтобы найти все возможные конформационные минимумы на поверхности, систематический метод MacroModel сможет это обеспечить. SPMC демонстрирует высокую эффективность на протяжении всего поиска, даже в конце, когда другие методы имеют тенденцию повторно обращаться к уже найденным конформерам. Метод поиска SPMC наиболее эффективен для небольших молекул [10]

Метод поиска	Комментарий
Low-Mode Conformational Search Methods (англ. LMCS, Конформационный поиск методом низкочастотных нормальных колебаний)	Данный подход позволяет обнаруживать седловые точки и минимумы на гиперповерхности потенциальной энергии путем градиентного слежения за значениями собственного вектора молекулярной системы (следованием за модой, англ. mode-following), полученными с помощью анализа нормальных колебаний. Если у вас мало или вообще нет предварительных знаний о системе, которую нужно найти, то этот метод позволяет выполнять автоматический поиск без необходимости определения параметров, которые нужно варьировать [11]

В этой работе мы не будем выполнять полноценный поиск с использованием MacroModel, а ограничимся систематическим сканированием нескольких торсионных углов небольшой молекулы с помощью модуля Coordinate Scan.

ConfGen, другой оригинальный алгоритм Schrödinger, на первом этапе разделяет молекулу на фрагменты, разрывая экзоциклические вращающиеся связи, позволяет получать конформационные массивы для всех фрагментов с проведением грубой минимизации и затем в определенном порядке соединяет их в единую систему, которая финально оптимизируется. ConfGen параметризован таким образом, чтобы как можно чаще приводить к биологически активным конформациям.

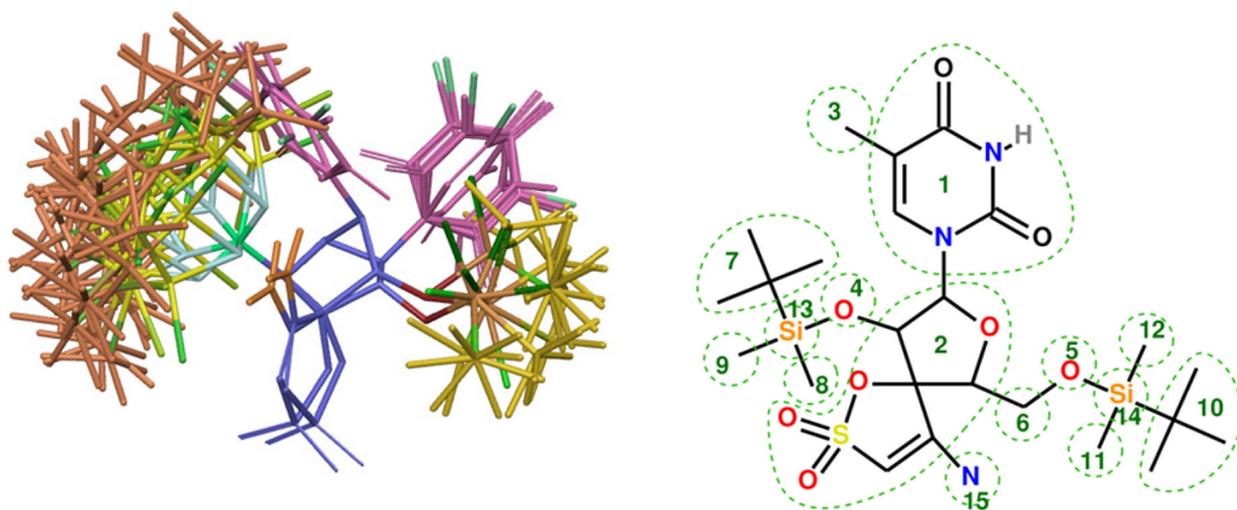


Рис. 2.5. Разделение молекул на фрагменты с помощью ConfGen

Показано, что в режиме высокой производительности алгоритм позволяет обрабатывать 12–15 лигандов в секунду без минимизации и порядка 0.4 лиганда в секунду с оптимизацией геометрии с помощью силового поля OPLS (рис. 2.5).

Экспериментальная часть

2.1. Создание проекта

В своем личном каталоге создайте папку LW2. Дважды щелкните по ярлыку Maestro на рабочем столе. На верхней панели перейдите на вкладку **File > Change Working Directory**. В открывшемся окне укажите путь к рабочей директории LW2 в своей личной папке.

Затем перейдите **File > Save project as**, задайте проекту имя LW2 в графе **File name**, укажите путь к рабочей директории в графе **Location**.

2.2. Конформационный анализ с помощью MacroModel

2.2.1. Подготовка структур

Откройте модуль **2D Sketcher**, для этого перейдите **Task > 2D Sketcher**. Открывшееся окно представляет собой обычный химический редактор. Нарисуйте представленную ниже структуру (если данное пособие открыто на ПК, просто скопируйте и вставьте SMILES). После чего нажмите **Save as New**, назовите новую запись «**Ddrive**» (рис. 2.6).

SMILES: Cc1c(=O)[nH]c(=O)n(COCC)c1Sc2ccccc2

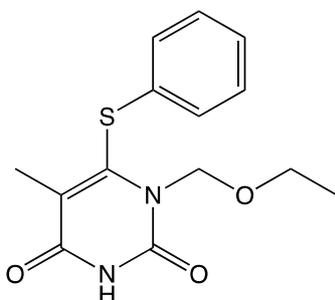


Рис. 2.6. Структура для работы

Выделите структуру **Ddrive** в **Entry List**. Откройте модуль **LigPrep** (кликните на значок модуля **LigPrep** в строке избранных заданий или перейдите **Task > LigPrep**), выставьте настройки по умолчанию (**OPLS4, Desalt, Generate tautomer, Generate possible states at target pH 7.4±0.2, Epik, Determine chiralities from 3D structure**) и выполните подготовку структуры **Ddrive**.

2.2.2. Выполнение сканирования торсионных углов

Выделите в **Entry List** и включите в **Workspace** структуру, полученную в результате расчета **LigPrep**. Отобразите в **Workspace** номера атомов в виде меток: в верхней части окна нажмите на панель **Style**, далее **Apply Labels > Element + Atom Number** (рис. 2.7).

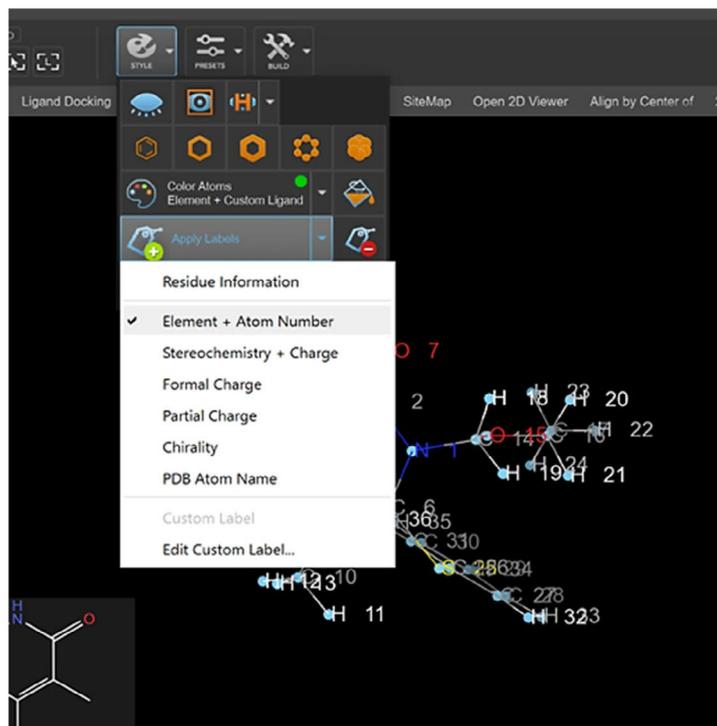


Рис. 2.7. Добавление меток атомам

Откройте инструмент конформационного поиска **MacroModel**. Перейдите **Task > Macromodel > Coordinate Scan**. В открывшемся окне выберите **Workspace (Inducted Entry)** напротив **Use structure from**, затем перейдите на вкладку **Scan**, где в качестве типа координат (**Coordinate Type**) выберите **Dihedral** (рис. 2.8).

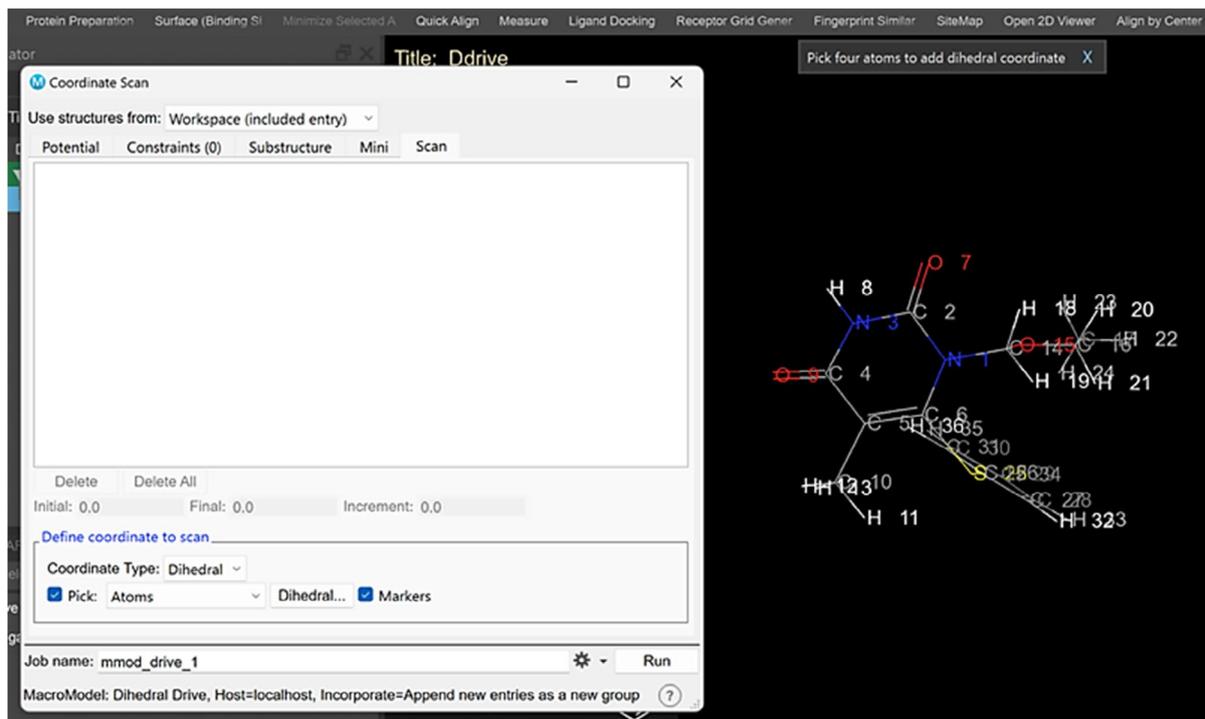


Рис. 2.8. Настройки конформационного поиска

Появившийся баннер в верхней части **Workspace** сообщает, что теперь необходимо определить искомые торсионные углы. Используя **Alt** для вращения, выделите атомы **1, 14, 15** и **16**. Вокруг выделенных атомов будут появляться метки в виде кубиков. Торсионный угол № 1 будет определен (связи, его образующие, подсвелятся зеленым). В зависимости от версии ПО и настроек нумерации номера атомов могут не совпадать – выделите атомы, показанные на рис. 2.9.

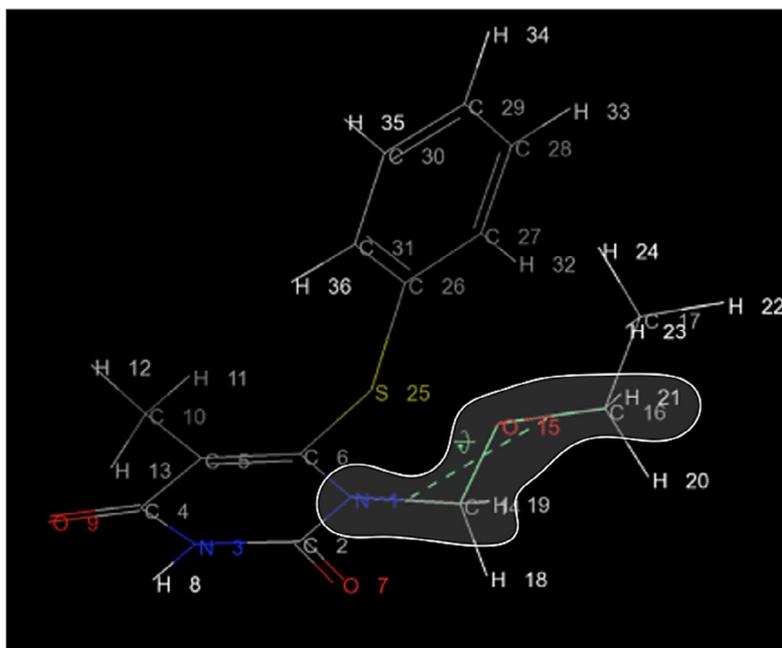


Рис. 2.9. Определение торсионного угла № 1

Используя **Alt** для вращения, выделите атомы **1, 6, 25** и **26**. Вокруг выделенных атомов будут появляться метки в виде кубиков. Торсионный угол № 2 будет определен (связи, его образующие, подсвелятся зеленым) (рис. 2.10).

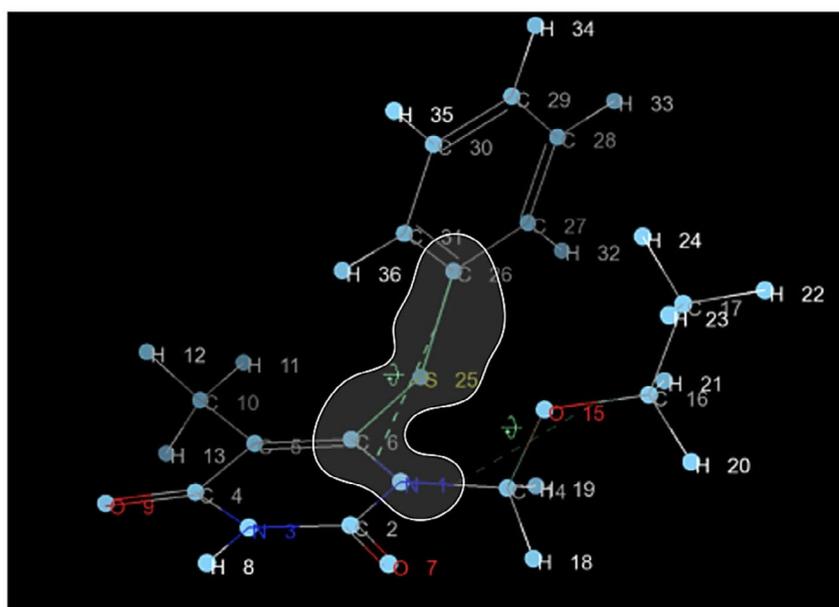


Рис. 2.10. Определение торсионного угла № 2

Вернитесь к окну **Coordinate Scan**. Обратите внимание, что выбранные углы появились на вкладке **Scan**. Перейдите на вкладку **Potential**. Проверьте, чтобы **OPLS4** было выбрано в качестве силового поля. Установите **None** напротив типа растворителя (**Solvent**) (в этом случае расчет будет происходить в вакууме). Остальные настройки автоматически выставляются по умолчанию (как показано на рис. 2.11).

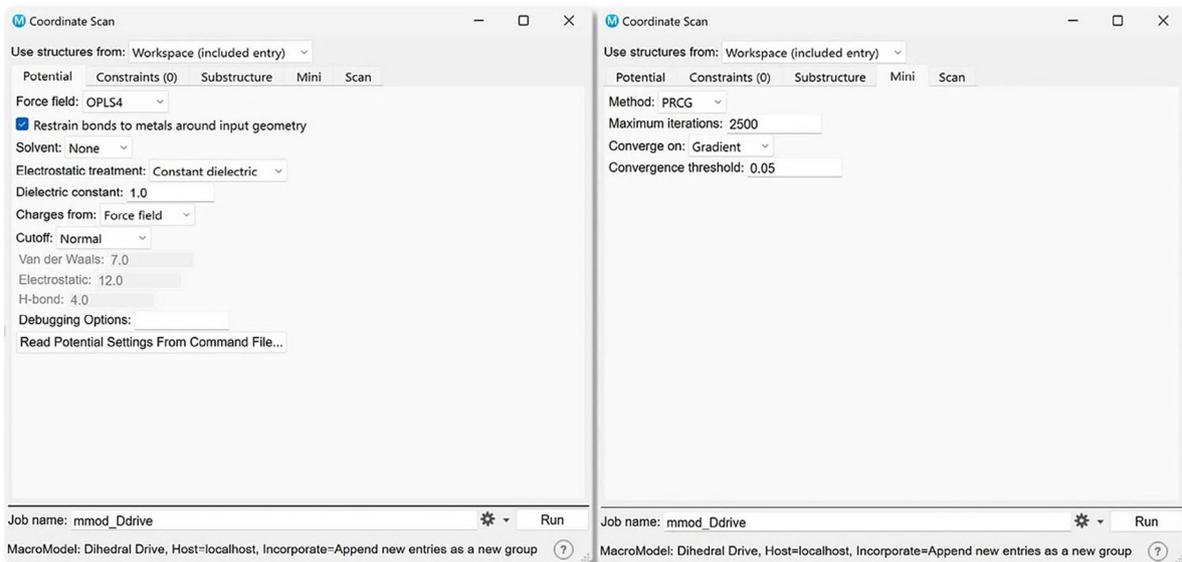


Рис. 2.11. Запуск задания

Перейдите на вкладку **Mini**, установите **PRCG** в качестве метода минимизации. Назовите работу «**mmod_Ddrive**» и нажмите **Run**. Расчет на одном процессоре займет примерно минуту.

Заметка: в этой работе мы запустим расчет без учета растворителя. Вы можете запустить эту работу в воде, а затем в хлороформе, чтобы сравнить внутримолекулярные взаимодействия в конформациях. **MacroModel** также имеет существенные возможности для управления силовыми полями, ограничениями и настройками, связанными с навигацией по поверхности потенциальной энергии.

2.2.3. Анализ поверхности потенциальной энергии

Отобразите рассчитанные значения потенциальной энергии в **Entry List**. Для этого в верхней правой части **Entry List** нажмите **Show Property > Choose > Relative Energy-S-OPLS**, потом **Choose > Potential Energy-S-OPLS**, затем **Choose** и **ОК**. Относительные энергии конформаций малых молекул играют решающую роль в определении формы и активности лиганда (рис. 2.12).

Последняя структура в группе **mmod_Ddrive-out** — **mmod_Ddrive-out Minimum E. structure during drive** — конформация с самой низкой энергией, обнаруженная в ходе сканирования пространства по двум торсионным углам. Переименуйте начальную структуру **Ddrive** (находящуюся в группе **ligprep_Ddrive-out**): дважды щелкните ЛКМ по строчке в **Entry list** и введите имя «**Ddrive Starting Structure**».

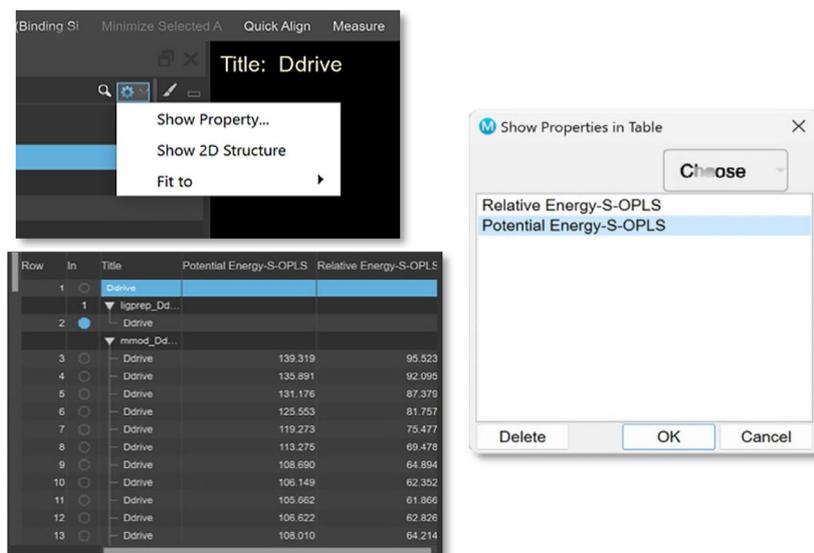


Рис. 2.12. Отображение рассчитанных значений энергии в Entry List

Щелкните ПКМ по переименованной структуре, в контекстном меню выберите **Duplicate > As Ungrouped Entry**. **Ddrive Starting Structure** теперь содержится в **Entry List** как отдельная запись в конце (строка 1373). Переместим ее в группу **mmod_Ddrive-out** на первую позицию: либо просто перетащите запись, зажав ЛКМ, либо вызовите контекстное меню нажатием ПКМ > **Move to > Other row** и в открывшемся окне напротив номера строки введите **3**. Структура будет перемещена.

Наложите структуры друг на друга с помощью модуля **Superposition** (не забудьте выделить их все; выделить всю группу можно, щелкнув ЛКМ по заголовку группы). В окне **Superposition** напротив строчки **Superimpose entries from** выберите **Project Table (selected)**, в качестве референсной выберите структуру **Ddrive Starting Structure**, поставьте галочку у пункта **Add property to Project Table** в **RMSD options**. Используйте метод совмещения подструктур (**Substructures**) и нотацию **SMARTS** в качестве способа их определения (рис. 2.13).

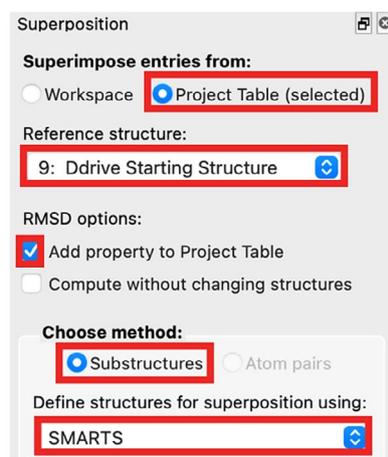


Рис. 2.13. Настройки наложения структур

Теперь необходимо получить выражение SMARTS. Нажмите **A** на клавиатуре: режим выделения атомов будет активирован. Включите первую структуру в **Workspace**

и с зажатым **Ctrl** выделите все атомы пиримидинового ядра. Выражение SMARTS после выделения должно принять вид:

SMARTS: [#7]1:[#6]:[#7H]:[#6]:[#6]:[#6]:1

В окне **Superposition** нажмите **Get from selection**, затем — **Superimpose structures**. Расчет займет меньше минуты (рис. 2.14).

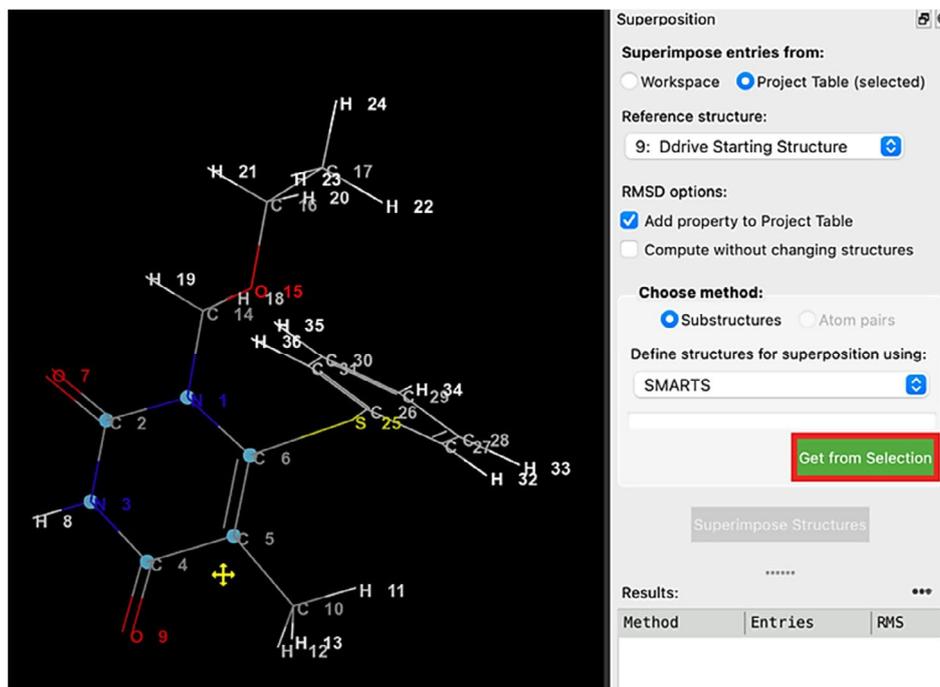


Рис. 2.14. Выбор шаблона SMARTS для наложения структур

Не снимая выделения со всех структур в группе, дважды щелкните по кнопке **Include** напротив **Ddrive Starting Structure**, закрепив ее в **Workspace**. Используя клавиши «стрелка вправо» и «стрелка влево», вы будете поочередно включать в **Workspace** выделенные структуры. Также вы можете включить любое количество структур в **Workspace**, чтобы оценить результаты генерации конформаций. На рис. 2.15 показаны 300 самых низкоэнергетических конформаций.

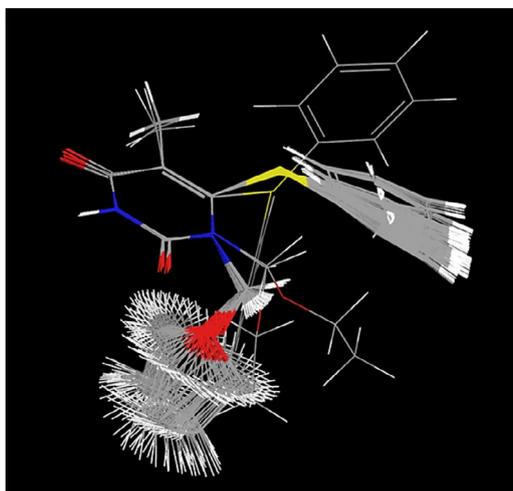


Рис. 2.15. 300 самых низкоэнергетических конформаций

Получим поверхность потенциальной энергии для сгенерированных конформаций. Проследуйте **Task > MacroModel > Plot Coordinate Scan**. В открывшемся окне нажмите **Load Results** для загрузки результатов конформационного поиска. Выберите файл **mmod_Ddrive-out.grd**. Раскройте меню напротив **Settings** и выберите **Scale: Absolute**. Дождитесь загрузки карты (рис. 2.16).

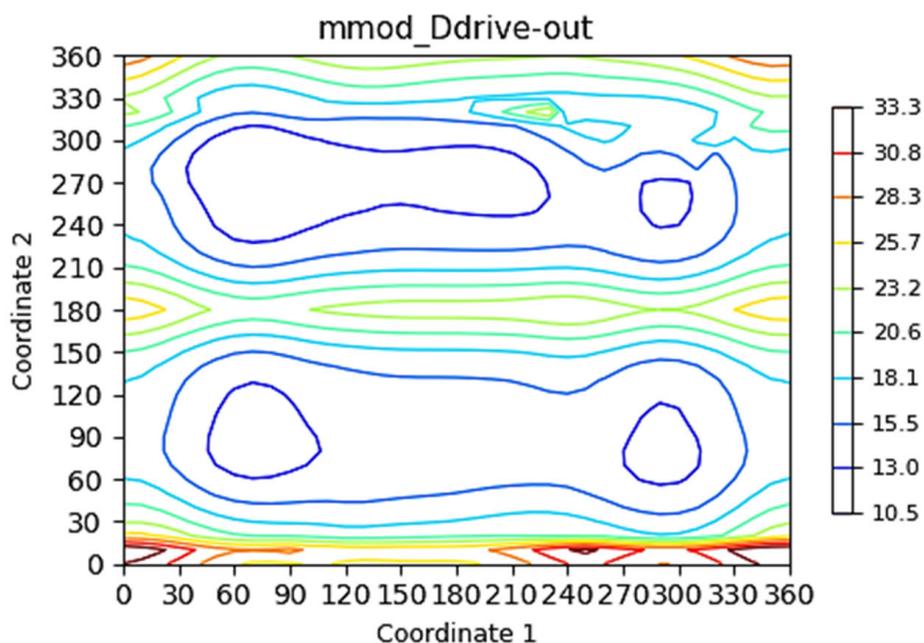


Рис. 2.16. 2D-карта потенциальной энергии

Отразите в отчете:

- 1) структуру исследуемого соединения;
- 2) снимок **Workspace** с наложением десяти самых низкоэнергетических конформаций (выделите серым цветом) и конформации с наибольшей энергией (выделите желтым цветом);
- 3) ответьте на вопрос: какие значения торсионных углов соответствуют наиболее низкоэнергетическим конформациям? Добавьте изображение 2D-карты в отчет. Включите в **Workspace** структуру с самой низкой энергией, используя инструмент **Measure > Dihedrals**, отобразите значения углов в **Workspace**. Сохраните снимок **Workspace** (**Workspace > Save Image As**), добавьте его в отчет;
- 4) ответьте на вопрос: чем обусловлено различие в энергиях между самой низко- и самой высокоэнергетической конформациями? Для ответа – включите обе в **Workspace**, затем **Show Workspace Configuration Panel > Workspace Layout > Title By > Entry**. В меню кастомизации (**Style** или значок кастомизации справа от **Current Selection** в **Structure Hierarchy**) выберите тип отображения атомов **CPK representation**. Через меню управления свойствами добавьте в **Entry List** значения **Van der Waals Energy-S-OPLS** и **Bend Energy S-OPLS** (энергия изгиба валентного угла), щелкните по заголовку столбца **ПКМ > Show property in Workspace > Add** (рис. 2.17).

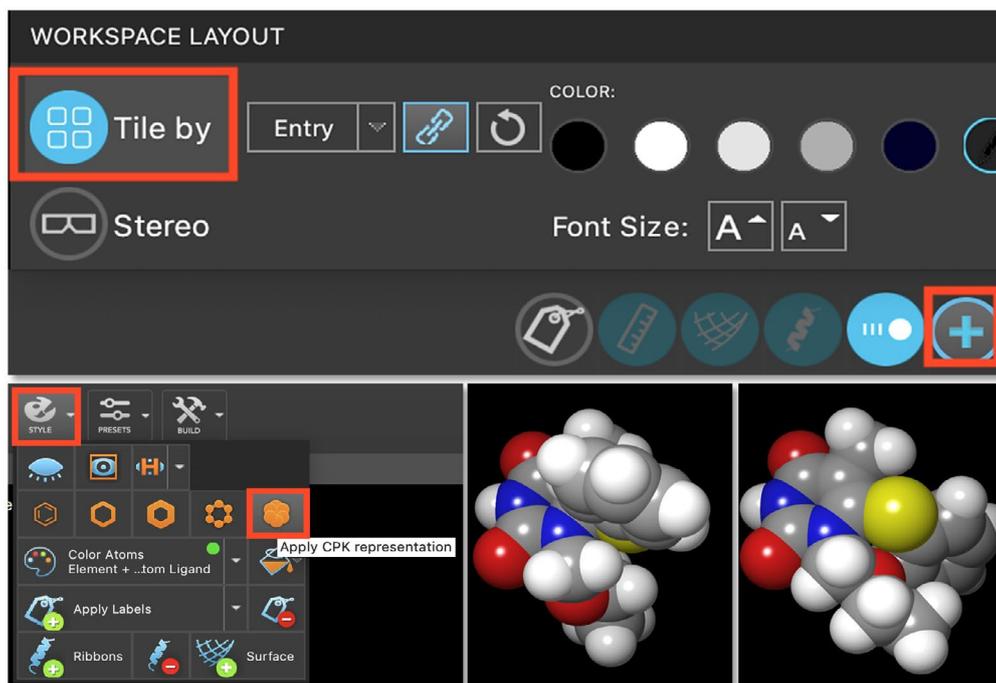


Рис. 2.17. Настройки Workspace

Снимок **Workspace** вставьте в отчет. Покажите на рисунке области, которые могут объяснять различия в энергии.

2.3. Конформационный анализ с использованием алгоритма ConfGen

Заметка: часто при SAR-оптимизации структуры полезно варьировать линкеры, соединяющие циклические (особенно ароматические) ядра для оценки влияния на активность. В этом упражнении вам будет предложено оценить различия в энергиях конформаций для следующих молекул (рис. 2.18).

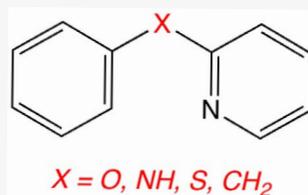


Рис. 2.18. Структуры, используемые в упражнении

Алгоритм **ConfGen** будет использован для выполнения поиска, так как он значительно быстрее, чем любой из методов **MacroModel** и, кроме того, позволяет генерировать конформации, близкие к биоактивным. Без сомнения, то же позволяет делать и **MacroModel**, к тому же обуславливая гораздо более широкий спектр тонких настроек поиска и минимизации. Однако в целях экономии времени и CPU-ресурсов мы не будем применять его. Вы можете самостоятельно выполнить тот же поиск, используя **MacroModel**: **Task > MacroModel > Conformational Search**.

2.3.1. Подготовка структур

Импортируйте файл **ConfGen_Linkers.mae** из своей рабочей директории (**Ctrl+I** или **File > Import Structures**).

2.3.2. Конформационный поиск

Раскройте группу **Phenyl-linker-Pyridyl** и включите все структуры в **Workspace**. Затем откройте модуль **ConfGen: Task > Bioactive Search**. В открывшемся окне выберите **Workspace** напротив **Use structures from**, установите «50» как число генерируемых конформаций, назовите работу так, как показано на рис. 2.19.

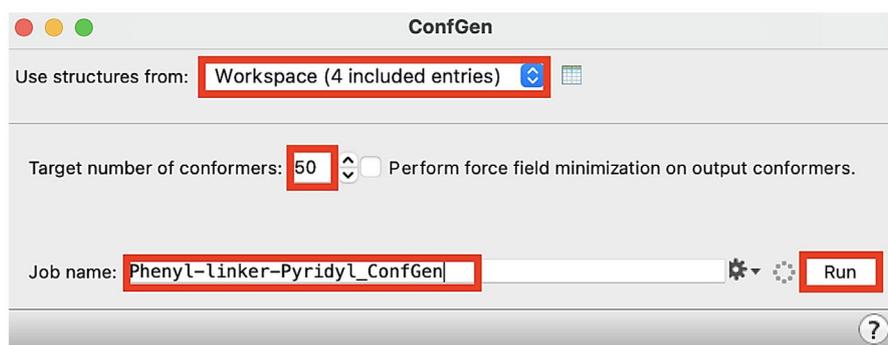


Рис. 2.19. Запуск поиска ConfGen

Нажмите **Run**. Расчет займет меньше минуты.

2.3.3. Анализ результатов конформационного поиска

Скройте свойства **Potential Energy-S-OPLS/Relative Energy-S-OPLS** и добавьте **energy**, **relative energy** через меню управления свойствами в **Entry List**. Затем выберите все группы среди результатов расчета **ConfGen**, щелкните ПКМ на заголовке столбца **energy > Sort Selected (Ascending)** (рис. 2.20).

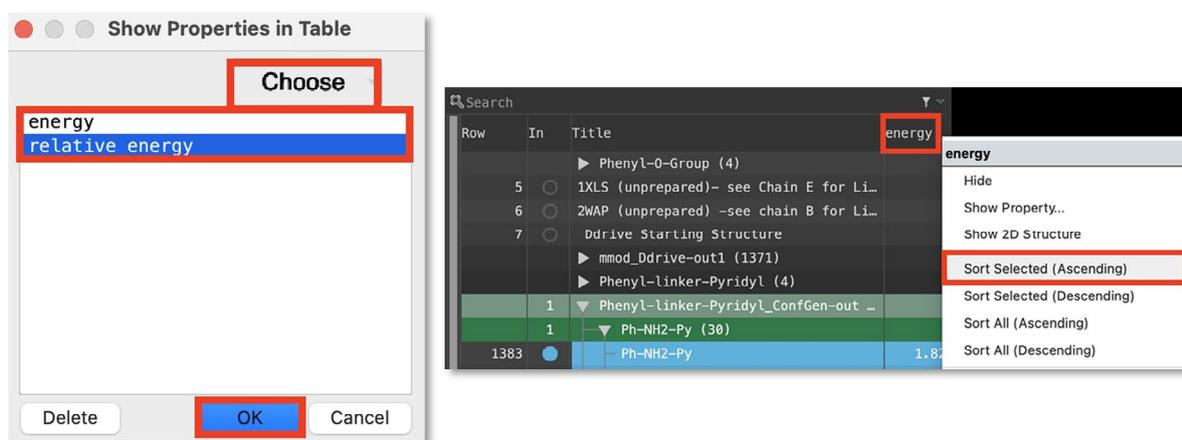


Рис. 2.20. Сортировка

Щелкните ПКМ на любом месте в **Workspace** и выберите **Clear Workspace**. Включите в **Workspace** самые низкоэнергетические конформеры для каждой из групп. На-

жмите **Ctrl+A > Presets > Apply Default Custom Presets**. Выполните выравнивание молекулы с помощью **Superposition** с использованием **Ph-NH2-Py** в качестве референсного лиганда, шаблон SMARTS определите как атомы углерода пиридинового ядра.

Наложите лиганды друг на друга, чтобы увидеть различия. Щелкните ПКМ по заголовку столбца **energy > Show property in Workspace > Add**.

Отразите в отчете:

- 1) структуры исследуемых соединений;
- 2) снимок Workspace с наложением конформаций с самой низкой энергией для всех четырех структур, а также значениями energy;
- 3) ответьте на вопрос: какой линкер позволяет получить конформацию с наименьшей энергией и почему?

Вопросы к защите

1. Дайте определение понятиям «конформация», «конформационный поиск» и «минимизация энергии».
2. В чем заключаются суть, достоинства и недостатки систематического конформационного поиска?
3. В чем заключаются суть, достоинства и недостатки стохастического конформационного поиска?
4. Чему пропорциональна вероятность принятия конформации с энергией E_{k+1} в стохастическом поиске Монте-Карло, если $E_{k+1} > E_k$?
5. Приведите общее уравнение силового поля и охарактеризуйте каждый энергетический термин.
6. Каким образом можно учесть влияние растворителя при выполнении конформационного поиска?
7. Какой недостаток метода наискорейшего спуска решается использованием подхода «сопряженных градиентов»?
8. Какой метод поиска используется модулем Coordinate Scan?
9. Для чего применяются подходы конформационного поиска в CADD?

Рекомендуемая литература

1. Molecular modeling : Basic Principles and Applications / H.-D. Holtje, W. Sippl, D. Rognan, G. Folkers. — 3rd, Revised and Expanded Ed. — Weinheim : Wiley-VCH, 2008. — X, 310 p. — ISBN 978-3-527-31568-0.
2. Leach, A. R. A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules // Reviews in Computational Chemistry / Eds: K. B. Lipkowitz, D. B. Boyd. — Weinheim, 1991. — Vol. 2. — P. 1–55.
3. Systematic search in conformational analysis / D. D. Beusen, E. F. Berkley Shands, S. F. Karasek [et al.] // Journal of Molecular Structure : THEOCHEM. — 1996. — Vol. 370, № 2–3. — P. 157–171.
4. Hinchliffe, A. Molecular Modelling for Beginners / A. Hinchliffe. — 2nd Ed. — Chichester : Wiley-VCH, 2008. — XVIII, 410 p. — ISBN 978-0-470-51314-9.
5. The Conformations of Taxol in Chloroform / J. P. Snyder, N. Nevins, D. O. Cicero, J. Jansen // Journal of the American Chemical Society. — 2000. — Vol. 122, № 4. — P. 724–725.
6. Schrödinger Release 2023-4: MacroModel. — New York : Schrödinger, 2023.

7. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers / K. S. Watts, P. Dalal, R. B. Murphy [et al.] // Journal of Chemical Information and Modeling. – 2010. – Vol. 50, № 4. – P. 534–546.
8. Schrödinger Release 2023-4: ConfGen. – New York : Schrödinger, 2023.
9. Chang, G. An Internal Coordinate Monte-Carlo Method for Searching Conformational Space / G. Chang, W. C. Guida, W. C. Still // Journal of the American Chemical Society. – 1989. – Vol. 111, № 12. – P. 4379–4386.
10. Goodman, J. M. An unbounded systematic search of conformational space / J. M. Goodman, W. C. Still // Journal of Computational Chemistry. – 1991. – Vol. 12, № 9. – P. 1110–1117.
11. Kolossváry, I. Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides / I. Kolossváry W. C. Guida // Journal of the American Chemical Society. – 1996. – Vol. 118, № 21. – P. 5011–5019.



Лабораторная работа 3

ФАРМАКОФОРНЫЙ ПОИСК

Цель работы: Освоение основных подходов к фармакофорному моделированию и скринингу.

Вам предстоит:

- 1) провести построение и валидацию фармакофорных гипотез с помощью лиганд- и структурно-ориентированных подходов;
- 2) выполнить фармакофорный скрининг с помощью полученных гипотез;
- 3) проанализировать его результаты.

Теоретическое введение

Согласно официальному определению International Union of Pure and Applied Chemistry (*англ.* IUPAC), **фармакофор** – это совокупность стерических и электронных свойств, которые необходимы для обеспечения оптимальных супрамолекулярных взаимодействий со специфической биологической мишенью для индукции (или ингибирования) ее биологического ответа [1].

Фармакофор (фармакофорную гипотезу) следует рассматривать как **общий знаменатель** группы молекул, которые обладают сходным фармакологическим профилем и распознаются одним и тем же участком белка-мишени. В фармакофорном моделировании каждый тип атома (или группы атомов) в молекуле, проявляющий определенные свойства, сводится к *фармакофорным признакам* (фармакофорным группам или *сайтам*). Эти структуры могут быть обозначены как доноры (D) или акцепторы (A) водородных связей, катионные (P), анионные (N), ароматические/циклические (R) или гидрофобные (H) центры. На рис. 3.1 вы можете увидеть пример фармакофорной модели для ингибитора ВТК пиртобрутиниба (PDB ID: 8FLN [2]), полученного с помощью ПО Phase [3–5].

Алгоритм Phase располагает сайт A на атоме, который несет одну или несколько неподеленных электронных пар, а векторный атрибут присваивается каждой идеализированной оси водородной связи в соответствии с гибридизацией акцепторного атома (сайт D характеризуется аналогично).

Циклы, галогены, алкильные цепи длиной до четырех атомов рассматриваются как одна гидрофобная группа. Цепочки из пяти и более атомов углерода разбиваются на более мелкие фрагменты и каждый обозначается как отдельный сайт, местоположение которого представляет собой средневзвешенное положение неводородных атомов во фрагменте.

Отрицательно и положительно ионизируемые фрагменты моделируются как одна точка, расположенная на формально заряженном атоме или в центре группы атомов (возникающие ионизируемые центры автоматически распознаются Phase, представленные структуры не требуют явной ионизации).

Наконец, если пользователь того пожелает, ароматические кольца можно отличить от других гидрофобных групп и обозначить как отдельный тип фармакофорного признака (т. е. R, а не H). В этом случае в центре каждого ароматического кольца помещается точка сайта, а с ним связывается двунаправленный вектор, нормальный к плоскости кольца (рис. 3.2) [3].

Фармакофорная модель представляется в непривязанном к конкретной структуре виде, описывается совокупностью сфер определенного радиуса и набором попарных расстояний между ними. Радиус сферы является *толерантностью* фармакофорного признака (рис. 3.3).

Любой фармакофор может быть также представлен в виде битовой строки; подобные представления называются *фармакофорными отпечатками пальцев* (фармакофорными ключами). Длина ключа зависит от числа уникальных попарных связей между признаками и заданных диапазонов расстояний (например, $>1 \text{ \AA}$, $1-3 \text{ \AA}$, $3-6 \text{ \AA}$, $<10 \text{ \AA}$ и т. д.). Как можно увидеть на рис. 3.4, признак R16 находится относительно признака D5 на дистанции в 9.55 \AA – значит ячейка отпечатка, соответствующая диапазону $9-13 \text{ \AA}$, будет содержать значение «1» (иначе 0) [6].

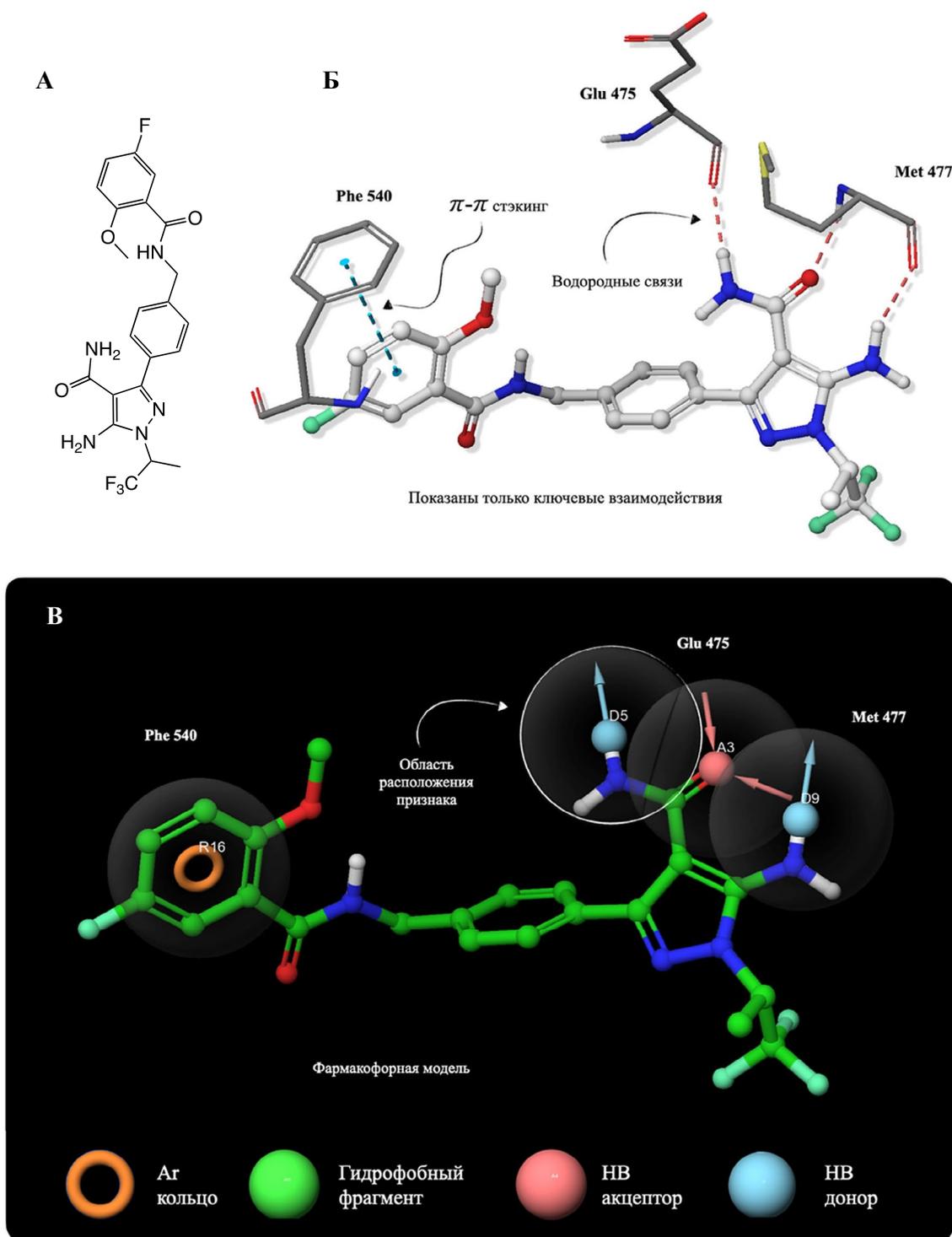


Рис. 3.1. Фармакофорная модель пиртобрутиниба: А – структура пиртобрутиниба; Б – способ связывания пиртобрутиниба в активном сайте ВТК; В – фармакофорная гипотеза, основанная на взаимодействиях, наложенная на структуру пиртобрутиниба; красный пунктир – водородные связи, синий пунктир – стэкинговые взаимодействия

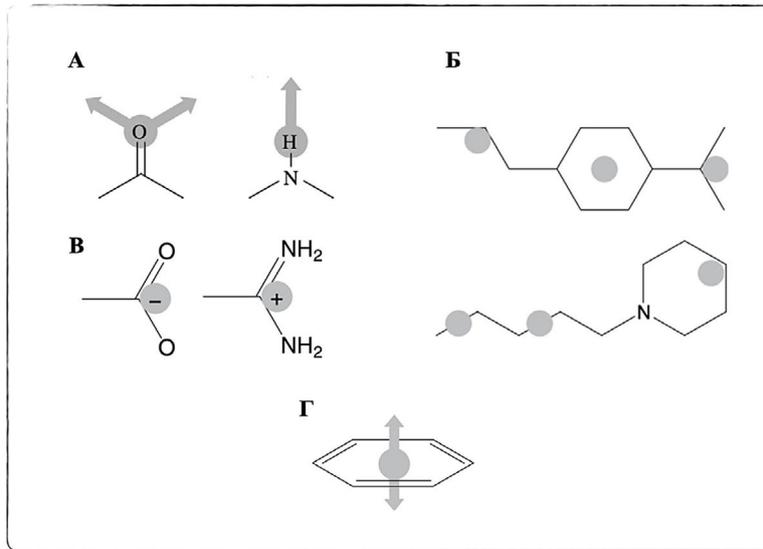


Рис. 3.2. Примеры представления фармакофорных признаков

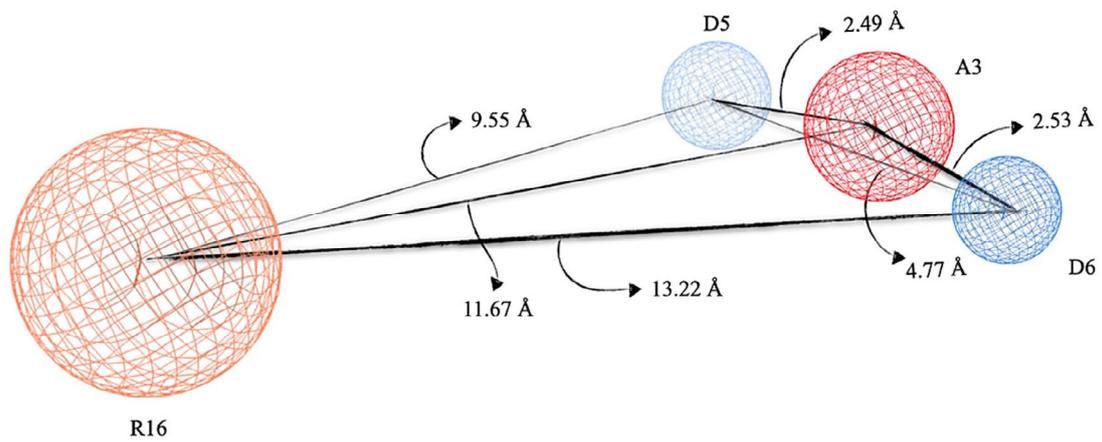


Рис. 3.3. Представление фармакофорной гипотезы

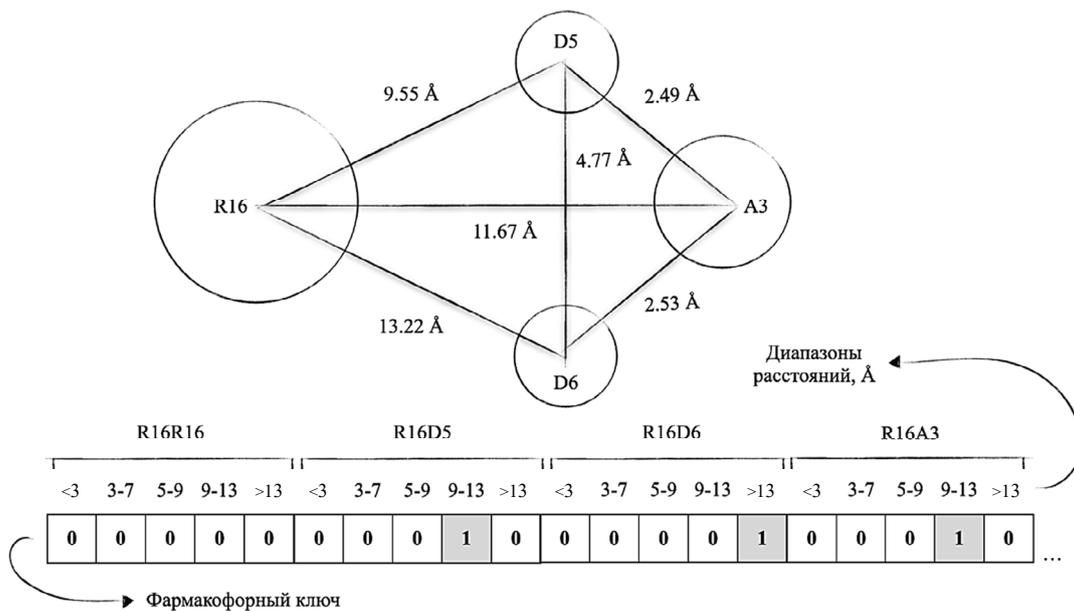


Рис. 3.4. Фармакофорный ключ

Фармакофорное моделирование чаще всего применяется для виртуального скрининга с целью идентификации молекул, вызывающих желаемый биологический эффект. Для этого исследователи создают фармакофорную модель (гипотезу), которая, скорее всего, кодирует правильную трехмерную организацию требуемых паттернов взаимодействий. В зависимости от того, насколько много известно о конкретном белке-мишени, доступны варианты построения такой гипотезы как на основе лигандов, так и на основе структуры рецептора. Затем происходит сравнение с фармакофорами соединений из библиотеки для скрининга. Те структуры, которые продемонстрировали наилучшее соответствие своих сайтов сайтам гипотезы, идентифицируются в качестве хитов (рис. 3.5).

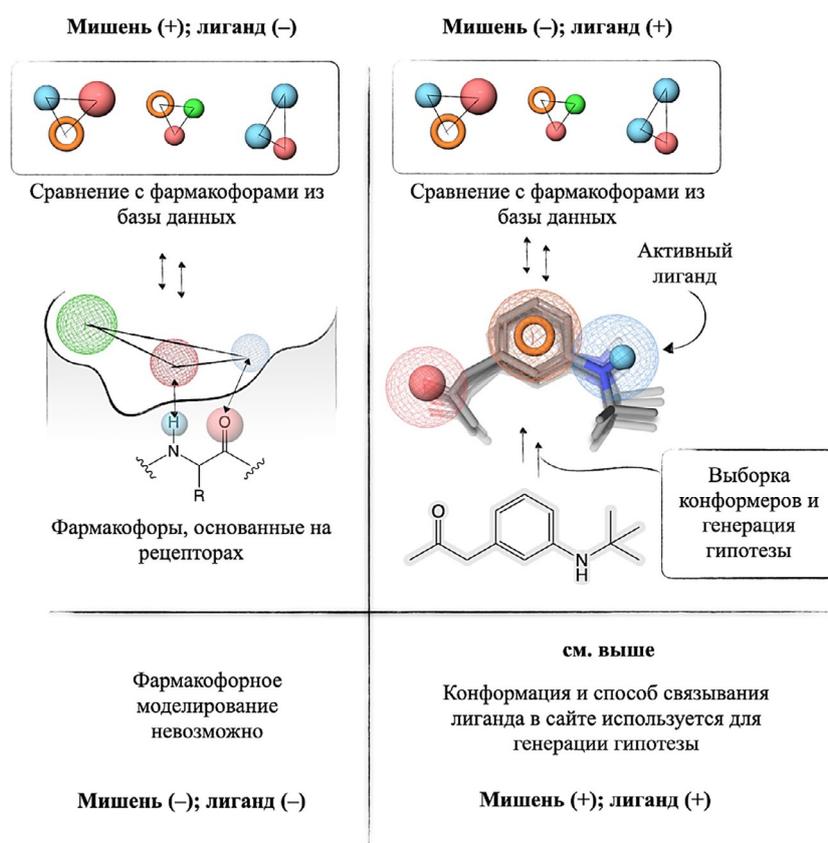


Рис. 3.5. Виды фармакофорного моделирования

Если структура мишени и лиганды, способные связываться с ней, неизвестны, использование подходов фармакофорного моделирования невозможно. Однако фармакофорные отпечатки пальцев могут быть использованы для расчетов метрик разнообразия (несходства) при создании всевозможных виртуальных библиотек.

В случае наличия информации о потенциальной мишени интереса (и/или комплекса «лиганд – мишень») можно применить подход **моделирования фармакофоров, основанных на рецепторе (RBPM, англ. Receptor-based pharmacophore modeling)**. Типичный рабочий процесс в данном случае включает в себя подготовку структуры белка, анализ способа связывания в активном сайте (или идентификацию потенциальных активных сайтов), построение гипотезы и ее применение в ходе скрининга лигандов, стыковки, сравнении и оценки сходства сайтов связывания. Важными фармакофорными признаками, которые не рассматривались нами до этого, являются *исключенные объемы*

(англ. Excluded volumes) — области пространства, в которых запрещено расположение атомов биологически активных соединений при выполнении моделирования. Это естественное продолжение концепции, согласно которой фармакофор отображает межмолекулярные взаимодействия. Молекула, образующая потенциальный солевого мостик с каким-либо остатком белка, не должна занимать ту же область пространства, что и боковая цепь. Исключенные объемы представляются в виде жестких сфер заданного радиуса и являются чрезвычайно полезными при RBPM (что, однако, не исключает возможности их использования и при моделировании фармакофоров, основанных на лиганде).

В случае **фармакофорного моделирования, основанного на лигандах** (LBPM, англ. Ligand-based pharmacophore modeling), первоначально следует озадачиться поиском набора активных лигандов, на основе которых и будет строиться будущая гипотеза. Необходимо включать в первичный пул также неактивные соединения для облегчения идентификации важных фармакофорных признаков и валидации модели. Такой набор соединений называется обучающим.

Следующий этап заключается в проведении конформационного анализа, направленного на получение низкоэнергетических трехмерных структур. Практически никогда не удастся обойтись одной конформацией для одного соединения и приходится иметь дело с ансамблями конформеров. В этом шаге кроется существенная проблема: биологически активная конформация может не соответствовать ни одному из локальных минимумов, отобранных алгоритмом. Другой негативный момент заключается в том, что исчерпывающий поиск и особенно последующее многокомпонентное выравнивание являются чрезвычайно дорогостоящими и нетривиальными процедурами. Обычно конформации, находящиеся близко друг к другу по значениям энергии или RMSD атомов, объединяются или часть из них устраняется.

В любом случае для полученного массива конформаций генерируются фармакофорные сайты и выполняется процедура наложения. Суть этого процесса состоит в том, чтобы соединить набор активных лигандов, каждый из которых *одинаково связывается с одним и тем же белком* неизвестной трехмерной структуры для того, чтобы общие черты истинного фармакофора стали очевидными. Набор фармакофорных свойств структуры какого-либо соединения выбирается в качестве *эталона* и относительно него выравниваются все остальные. Фармакофор, соответствующий *эталону*, считается отдельной фармакофорной гипотезой. Затем *эталон* меняется и наложение выполняется снова до тех пор, пока не будут получены все возможные модели, т. е. потенциально любой лиганд из набора может дать начало фармакофорной гипотезе. Упрощенно можно сказать, что лучшей гипотезой является та, на которую лучше всех накладываются фармакофорные сайты лигандов набора. Однако в действительности оценка гипотез представляет собой интересную процедуру, которая в каждом ПО реализована по-разному (как и собственно процесс идентификации гипотез) [7].

Процедуры генерации трехмерных структур и получения набора фармакофорных свойств повторяются и для лигандов из скрининговой библиотеки. Затем каждый из них выравнивается и сравнивается с полученной ранее гипотезой. Концепция идентификации хитов в таком скрининге похожа на таковую для ранжирования гипотез: чем лучше накладывается фармакофор какого-либо соединения с гипотезой-эталон, тем больше вероятность того, что это соединение будет обладать должной биологической активностью (рис. 3.6) [6].

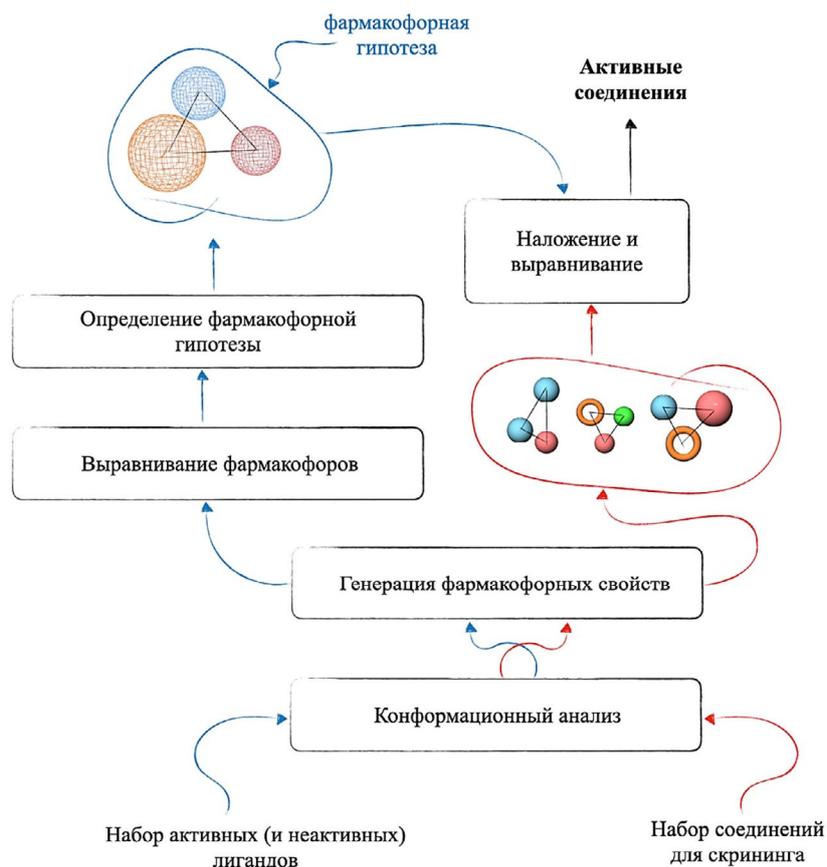


Рис. 3.6. Рабочий процесс LBPM

Модуль Maestro Schrödinger Phase, используемый для задач фармакофорного моделирования в этой работе, основывается на алгоритмах MacroModel для предварительной генерации конформаций. Полученные структуры фильтруются через энергетическое окно, как правило 5–10 ккал/моль, и проверку избыточности, где любые две структуры в пределах одной ккал/моль друг от друга считаются эквивалентными, если все соответствующие пары тяжелых атомов в двух структурах находятся в пределах заданного пользователем расстояния (как правило, 1–2 Å). Любой k -точечный фармакофор представляется вектором n расстояний, где $n = (k \cdot (k-1))/2$ (см. п. 2.3.1). Похожие фармакофоры объединяются в соответствии с их межсайтовыми расстояниями d , каждое из которых фильтруется через двоичное дерево решений, как показано на рисунке ниже. Таким образом, определенный фармакофор назначается n -мерному боксу, стороны которого равны по длине ширине терминального узла дерева, который обозначается как ϵ (рис. 3.7).

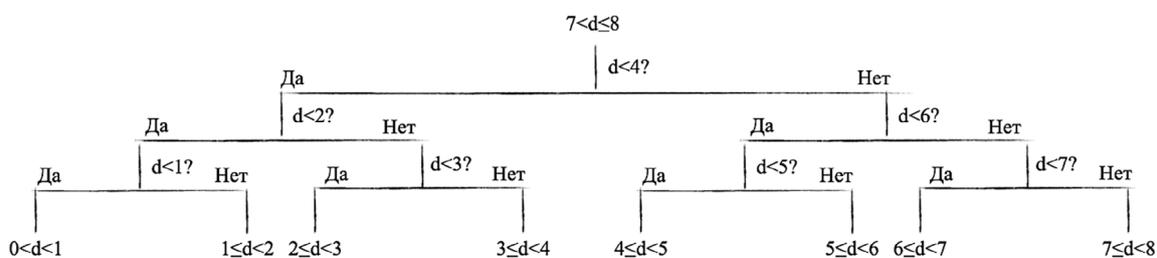


Рис. 3.7. Иллюстрация метода древовидного разделения, используемого для идентификации общих фармакофоров на основе межсайтовых расстояний. Это дерево имеет глубину четыре и разбивает расстояния в интервале (0, 8] Å на конечные узлы шириной 1 Å

Каждый бокс в Phase содержит свою фармакофорную гипотезу. Для ее получения один из лигандов и его фармакофор среди общего набора принимается за эталонный, а все остальные выравниваются относительно него. Выравнивание в большинстве программных решений осуществляется с помощью метода наименьших квадратов, который минимизирует RMSD между парами точек в пространстве.

Качество выравнивания Phase измеряется тремя способами: (1) показателем выравнивания S_{site} , который представляет собой среднеквадратичное отклонение (RMSD) координат точек сайтов; (2) показателем S_{vector} , который представляет собой средний косинус углов, образованных соответствующими парами векторных признаков (акцепторы, доноры и ароматические кольца) в выровненных структурах; и (3) показателем объема S_{vol} , основанным на перекрытии объемов Ван-дер-Ваальса неводородных атомов в каждой паре структур:

$$S_{site} = 1 - \frac{RMSD_i}{cutoff_{RMSD}}$$

где $cutoff_{RMSD}$ — определяемое пользователем значение отсечки RMSD (по умолчанию равно 1.2); S_{site} всегда принимает значение от 0 до 1 ($RMSD_i > cutoff_{RMSD}$);

$$S_{vector} = \frac{1}{n_v} \sum_{j=1}^{n_v} \cos \theta_{ij}$$

где n_v — количество векторных признаков в гипотезе, θ_{ij} — угол между j -м векторным признаком в нереперентном фармакофоре и соответствующим векторным признаком в эталонном фармакофоре;

$$S_{vol} = \frac{Vol_{common,i}}{Vol_{total,i}}$$

где $Vol_{common,i}$ представляет собой перекрывающийся объем между лигандом i и эталонным лигандом, а $Vol_{total,i}$ представляет собой общий объем, занимаемый обоими лигандами. Общий S_{vol} для гипотезы представляет собой среднее значение, полученное в результате применения приведенной выше формулы ко всем неэталонным лигандам.

Суммарная же оценка гипотезы Phase S (в оригинальной литературе называемая *Survial Score*) определяется формулой

$$S = \omega_{site} S_{site} + \omega_{vector} S_{vector} + \omega_{vol} S_{vol} + \omega_{sel} S_{sel} + \omega_{match}^{M-1} - \omega_E \Delta E + \omega_{act} A,$$

где ω — настраиваемые весовые коэффициенты, S_{sel} — селективность — эмпирическая оценка редкости гипотезы, т. е. того, какая доля молекул с большой вероятностью соответствует модели, независимо от их активности по отношению к рецептору. Селективность определяется в логарифмическом масштабе, поэтому значение 2 означает, что 1 из 102 молекул, как ожидается, будет соответствовать гипотезе. Желательна более высокая селективность, поскольку она указывает на то, что гипотеза с большей вероятностью будет уникальной для лигандов активного набора. Селективность — это лишь приблизительная оценка редкости, поэтому вам следует быть осторожными и не придавать ей слишком большого значения при общем рейтинге гипотез. ω_{match} — термин, позволяющий награждать гипотезы, которые соответствуют большему количеству активных соединений (M — число активных лигандов, соответствующих гипотезе).

$\omega_{match}\Delta E$ – штраф, назначаемый в том случае, когда эталонный лиганд не соответствует самому низкоэнергетическому конформеру, а A – активность (все слагаемое $\omega_A A$ позволяет штрафовать гипотезы, эталон которых не является самым активным лигандом в наборе).

В том случае, когда в обучающий набор включены неактивные лиганды, можно провести дополнительную валидацию гипотез. Алгоритм может штрафовать те модели, которым соответствуют неактивные соединения, вычитая из уже рассчитанного суммарного S для каждой гипотезы оценку $S_{inactive}$. Расчет происходит точно так же, как и в формулах до этого, за исключением изменения показателя совпадения S_{site} :

$$S_{inactive} = \omega_{site} \left(1 - \left[\frac{m}{k} RMSD_{i,m}^2 + \frac{k-m}{k} cutoff_{RMSD}^2 \right]^{\frac{1}{2}} \right) + \omega_{vector} S_{vector} + \omega_{vol} S_{vol},$$

где k – общее число сайтов неактивной молекулы, m – число сайтов неактивной молекулы, совпавших с таковыми в гипотезе.

Если гипотеза содержит n сайтов, а неактивный представитель обучающего набора только k , то эффективность наложения рассчитывается на основе выравнивания только по k -точкам.

Финальная оценка PhaseHypoScore (S_{adjust}) представляет собой выражение:

$$S_{adjust} = S - \omega_{inactive} S_{inactive}.$$

Функция оценки PhaseScreenScore, используемая при скрининге библиотек соединений против гипотез, представляет собой просто усеченную до первых трех терминов оценку PhaseHypoScore [3, 4].

Экспериментальная часть

3.1. Создание проекта

В своем личном каталоге создайте папку LW3. Дважды щелкните по ярлыку Maestro на рабочем столе. На верхней панели перейдите на вкладку **File > Change Working Directory**. В открывшемся окне укажите путь к рабочей директории LW3 в своей личной папке.

Затем перейдите **File > Save project as**, задайте проекту имя LW3 в графе **File name**, укажите путь к рабочей директории в графе **Location**.

3.2. Фармакофорное моделирование, основанное на лигандах

3.2.1. Подготовка структур

Скопируйте файл **at1.mae** в рабочую директорию и импортируйте его в Maestro (**File > Import** или **Ctrl+I**). Файл содержит 25 подготовленных с помощью **LigPrep** лигандов, которые будут использоваться для получения гипотез.

3.2.2. Генерация фармакофорных гипотез

Включите все 25 структур в **Workspace**: для этого выделите группу, затем, щелкнув **ПКМ** по ее заголовку, выберите **Include** в открывшемся контекстном меню. Игнорируйте предупреждения программы (нажмите **Continue** во всплывающем окне). Как вы можете увидеть, лиганды не выровнены.

Заметка: несмотря на то, что Phase производит выравнивание структур при генерации гипотез, иногда полезно выполнять предварительное выравнивание с помощью инструмента **Flexible Ligand Alignment** (для идентификации возможных биологически активных конформаций). В этой работе мы не будем выполнять предварительное выравнивание.

Откройте модуль генерации фармакофорных гипотез **Phase: Task > Ligand-Based Virtual Screening > Develop Pharmacophore Hypotheses**. В открывшемся окне раскройте список под **Create pharmacophore model using** и выберите **Multiple ligands (selected entries)** (не забудьте, что все лиганды должны быть выделены в **Entry List**). Программа будет производить расчет некоторое время (1–4 минуты), затем все фармакофорные сайты будут отображены в **Workspace**.

Для определения активных и неактивных соединений в наборе нажмите **Define** в разделе **For Actives/Inactives Split (selected entries)** (рис. 3.8).

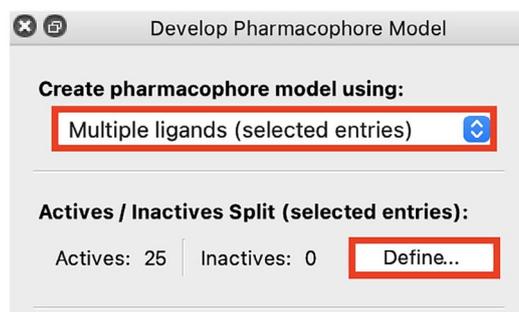


Рис. 3.8. Создание фармакофорных сайтов и определение активных соединений

Выберем свойства и пороговые значения, которые будут использованы для классификации соединений набора как активные/неактивные. В открывшемся окне **Define Actives and Inactives** в правой его части среди доступных свойств выберите **pIC50-Exp**, щелкнув по заголовку **ЛКМ**. Установите **Set Active if activity как ≥ 7.30** и **Set Inactive if activity как ≤ 5.00** . Нажмите **Apply > OK** (рис. 3.9).

Заметка: вы также можете вручную указать лиганды, которые должны соответствовать гипотезе, установив метки в столбце **Required Match** для желаемых молекул.

Выберите пункт **Find Best Alignment and Common Features** в разделе **Pharmacophore Method**. Теперь **Phase** будет использовать только лучшие наложения при создании гипотез. Пункт **Use Prealigned Ligands** используется в случае, если вы выполнили выравнивание до импорта структур в модуль создания гипотез.

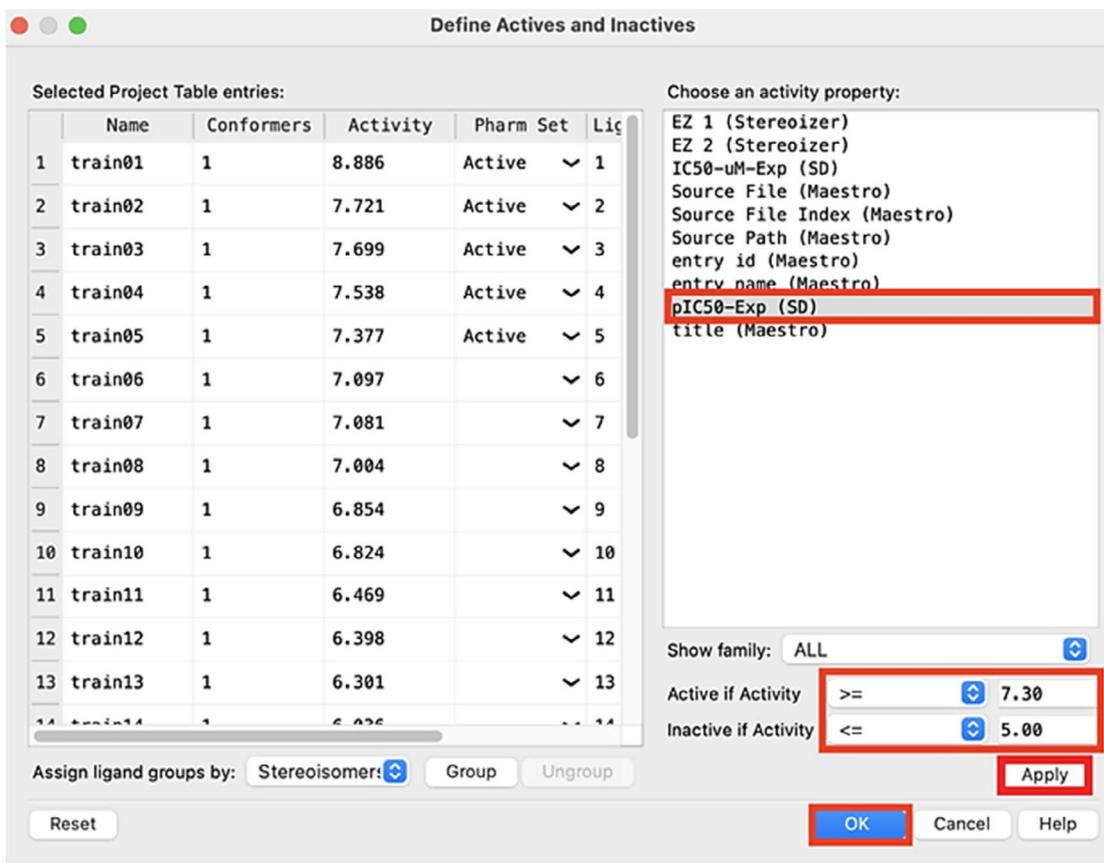


Рис. 3.9. Классификация соединений

Откройте панель настройки гипотез, нажав на кнопку **Hypothesis Settings**. Во вкладке **Features** в разделе **Number of features in the hypothesis** установите диапазон сайтов в гипотезе как «5 из 6», а **Preferred minimum number of features** (предпочитаемое минимальное число сайтов) как «5».

В таблице, расположенной в нижней части, в столбце **Minimum** установите значение «1» напротив (D) **Donor** и (N) **Negative**. В разделе **Feature presets** установите **Make acceptor and negative equivalent**. Теперь **Phase** будет сохранять только те гипотезы, где есть как минимум один донор водородной связи и/или один акцептор/анионный центр (рис. 3.10).

Заметка: опция **Exclude features from hypothesis** (исключить сайты из гипотезы) позволяет скрыть любые сайты, которые вы не хотите видеть в гипотезе **Phase**. Это может быть использовано для маскировки признаков в консервативном ядре, чтобы «заставить» алгоритм «сосредоточиться» на более переменных областях серии соединений.

Создадим исключенные объемы. Перейдите на вкладку **Excluded Volume** и нажмите **Create excluded volume shell**. Выберите **Actives and Inactives** в разделе **Create shell from** и нажмите **Save**. Теперь **Phase** создаст исключенные объемы в тех областях, где отсутствуют атомы активных структур, но присутствуют атомы неактивных (рис. 3.11).

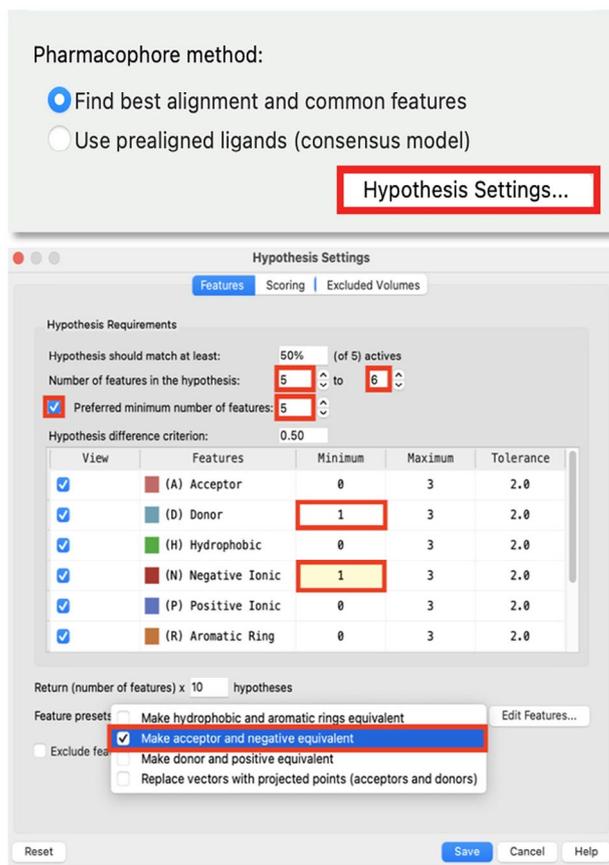


Рис. 3.10. Настройки количества сайтов для генерации гипотез

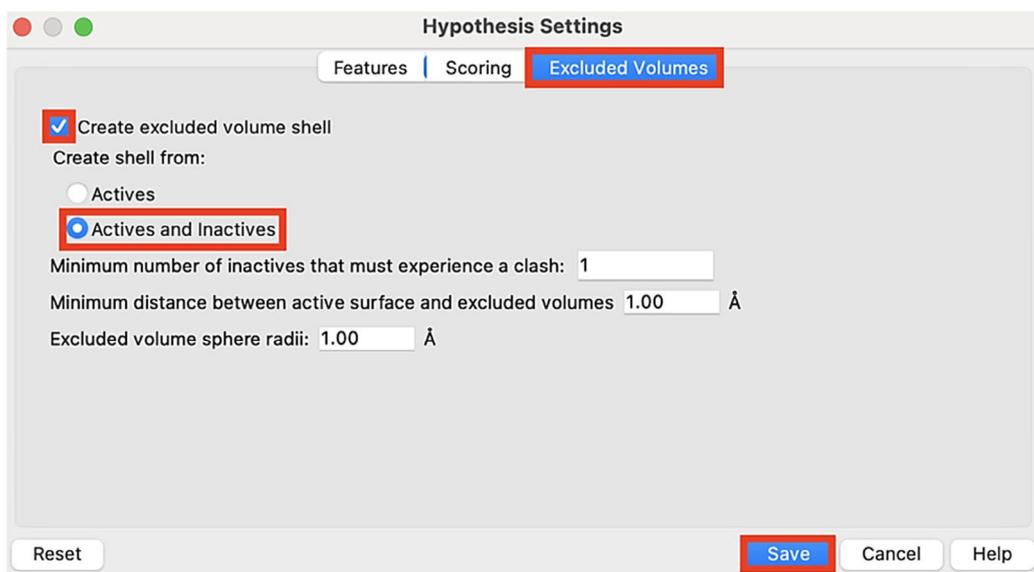


Рис. 3.11. Создание исключенных объемов

Вернитесь в окно **Develop Pharmacophore Model**, измените имя работы на «**at1_dcph**» и нажмите **Run**. Расчет занимает примерно 6–10 минут на четырех процессорах.

Когда расчет завершится, в **Entry List** будет добавлена группа **at1_dcph**, содержащая 20 различных гипотез.

Заметка: буквы в названии каждой гипотезы соответствуют тем сайтам, которые в ней содержатся. Например, **ADRR** – это гипотеза, включающая акцептор и донор водородной связи, два ароматических кольца.

Раскройте группу **DHNRRR_1** и закрепите запись **DHNRRR_1** в **Workspace** двойным щелчком ЛКМ по кнопке **Include**. Включите все лиганды из подгруппы **Active**. Обратите внимание на соответствие всех сайтов гипотезы наложенным лигандам. Кроме того, активные соединения не занимают области Excluded Volumes. Выключите подгруппу **Active**. Аналогичным образом раскройте группу **Inactive Partial Match** и осмотрите лиганды (рис. 3.12).

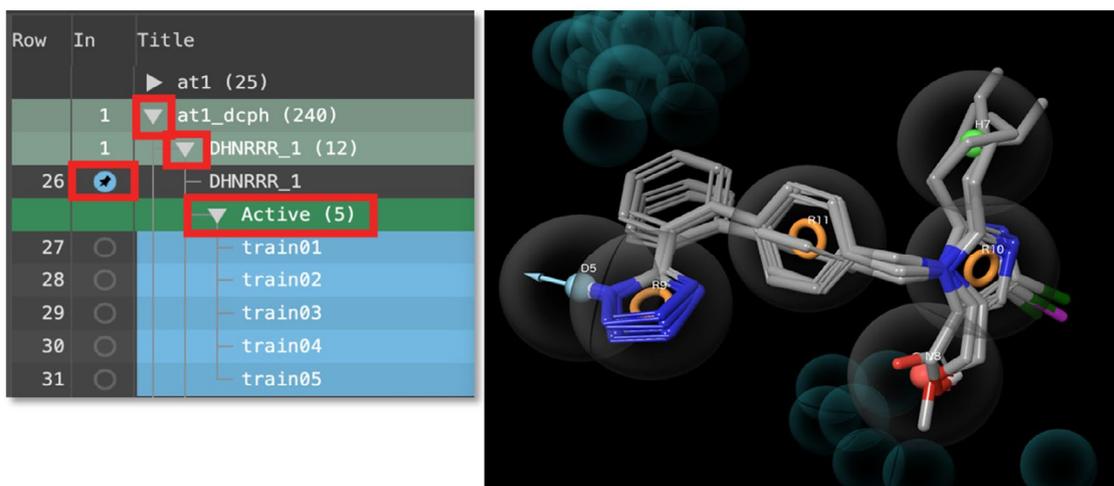


Рис. 3.12. Наложение лигандов, соответствующих активной гипотезе **DHNRRR_1**

Отразите в отчете:

- 1) 2D-карту фармакофора **DHNRRR_1**. Для этого оставьте в **Workspace** только запись самой гипотезы **DHNRRR_1**, затем в строке избранных заданий выберите инструмент **Measure** и отобразите все расстояния между сайтами в **Workspace**. Для измерения расстояния кликните ЛКМ по сайтам, между которыми его надо определить. В отчете приведите нарисованную карту и снимок экрана **Workspace** с показанными расстояниями;
- 2) таблицу со значениями **PhaseHypoScore** для всех сгенерированных гипотез. Для этого откройте панель **Project Table**. В **Property Tree** (справа) снимите флажок с пункта **All**. Через строку поиска найдите свойство **PhaseHypoScore** и поставьте рядом с ним флажок. Выделите все группы с гипотезами. В верхней панели нажмите **Data > Export > Spreadsheet**. Выберите формат **Excel** для экспорта и назовите файл по своему усмотрению. В таблице для отчета должны быть только гипотезы и значения скоринга для них.

3.2.3. Создание базы данных Phase

Импортируйте файл **DUD-E_act_3chp_trainc.mae**, содержащий лиганды, которые будут использованы для создания базы (**Ctrl+I** или **File>Import**). Затем выберите группу **DUD-E_act_3chp_trainc** в **Entry List** и откройте модуль создания скрининговых наборов **Phase: Tasks > Ligand-Based Virtual Screening > Create Database**.

Заметка: на этом этапе нам будет необходимо создать базу данных, содержащую соединения для фармакофорного поиска с помощью Phase. В ходе создания такого скринингового набора Phase автоматически генерирует низкоэнергетические конформации, выполняет создание фармакофорных сайтов и выравнивание, а затем удобно складывает данные в файлах формата .db. Вы можете использовать одну и ту же базу данных для поиска против различных гипотез. Phase способен работать и с непредобработанными лигандами, однако предварительное создание базы очень сильно сокращает время скрининга.

В открывшемся окне **Create Phase Database** перейдите на вкладку **Input**. Напротив **Choose task** выберите **Create new database** для создания новой базы. Если вы планируете дополнить существующий набор, следует выбрать **Append to Existing database**. Определите место сохранения базы данных, указав путь к своей рабочей директории в разделе **Specify the database location on the job**.

Выберите **Project Table (selected entries)** напротив **Use ligands from**, а в разделе **Conformation Generation** выберите **Generate ligand conformers**. Проверьте, чтобы число генерируемых конформаций было равно «50» (рис. 3.13).

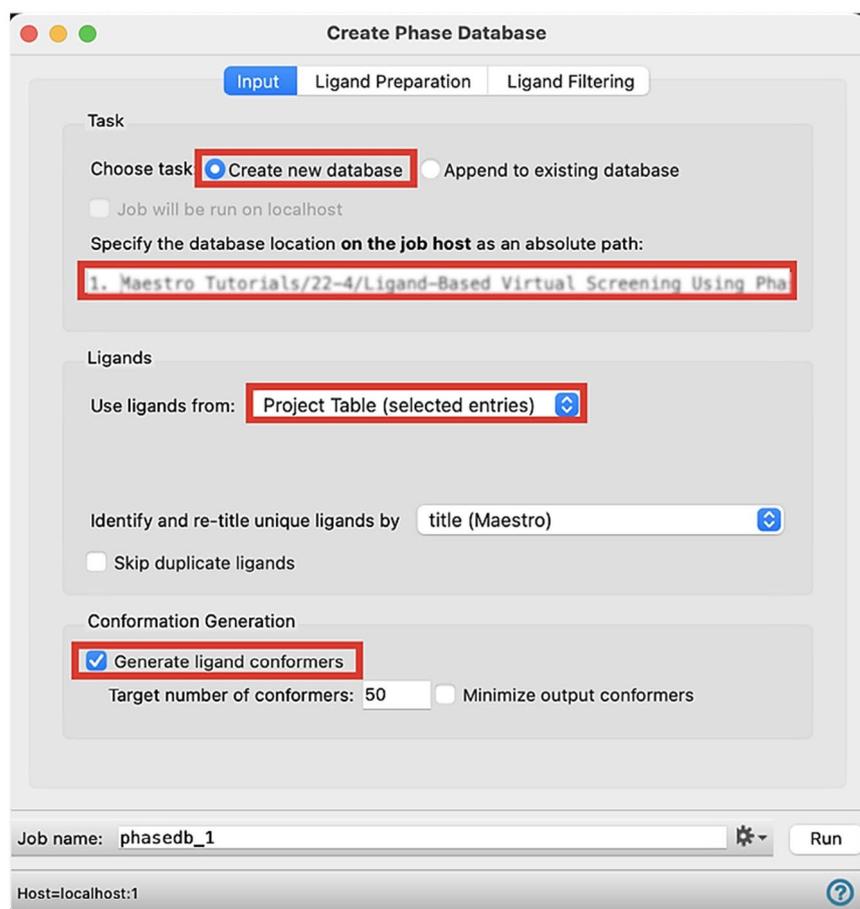


Рис. 3.13. Установка входных настроек модуля создания базы данных Phase

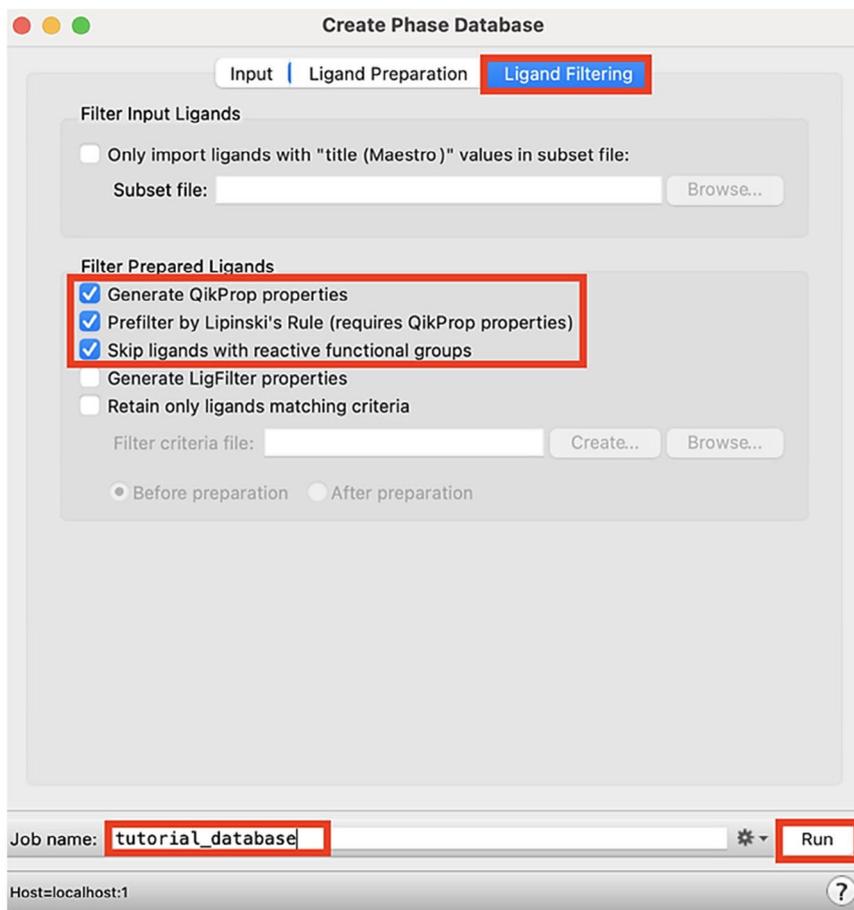


Рис. 3.14. Установка настроек фильтрации лигандов модуля создания базы данных Phase

Перейдите на вкладку **Ligand Filtering**. Выберите первые три пункта в разделе **Filter Prepared Ligands**. Теперь для всех соединений в наборе будут сгенерированы физико-химические свойства с помощью алгоритма **QikProp**, и соединения, содержащие реакционно-способные группы, а также не соответствующие правилам Липински, будут исключены. Вкладка **Ligand Preparation** отвечает за параметры подготовки лигандов с помощью интегрированного в **Phase** алгоритма **QikProp**. Все настройки автоматически устанавливаются по умолчанию и не будут изменяться в этой лабораторной работе. Оставьте название задания «**Phasebd_1**» и нажмите **Run**. Расчет на 4 процессорах занимает 5 минут (рис. 3.14).

3.2.4. Скрининг базы данных Phase

С помощью **Ctrl** включите в **Workspace** гипотезы **DHNRRR_1**, **DHNRRR_2** и **ADHNR_1** из группы **at1_dcp**. Откройте модуль скрининга **Tasks > Ligand-Based Virtual Screening > Ligand Screening**. В разделе **Ligands to screen** выберите **File**, затем **Browse** и укажите файл, содержащий созданную вами базу данных. Нажмите **Add Hypothesis > Workspace**. В таблицу ниже будут автоматически добавлены включенные вами в **Workspace** гипотезы. Вы можете настроить число сайтов, которые должны совпасть, а также учесть исключенные объемы в соответствующих столбцах. В этой работе оставим без изменений эти параметры (рис. 3.15).

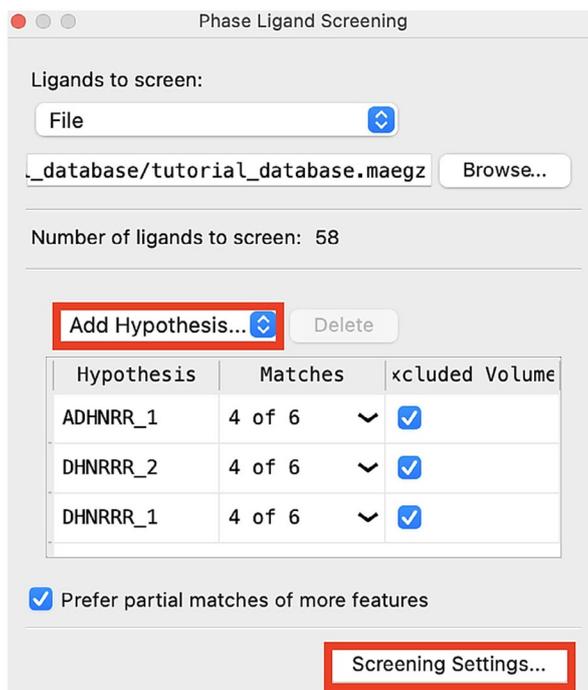


Рис. 3.15. Импорт базы данных и гипотез в модуль скрининга

Откройте окно настроек скрининга, нажав **Screening Settings**. Перейдите на вкладку **Constraints** и нажмите **Pick features**. В Workspace включите только **DHNR_1** и выберите сайты **R10, R11** с помощью **Ctrl**. Затем нажмите **Add Constraint**. Повторите это для остальных гипотез (рис. 3.16).

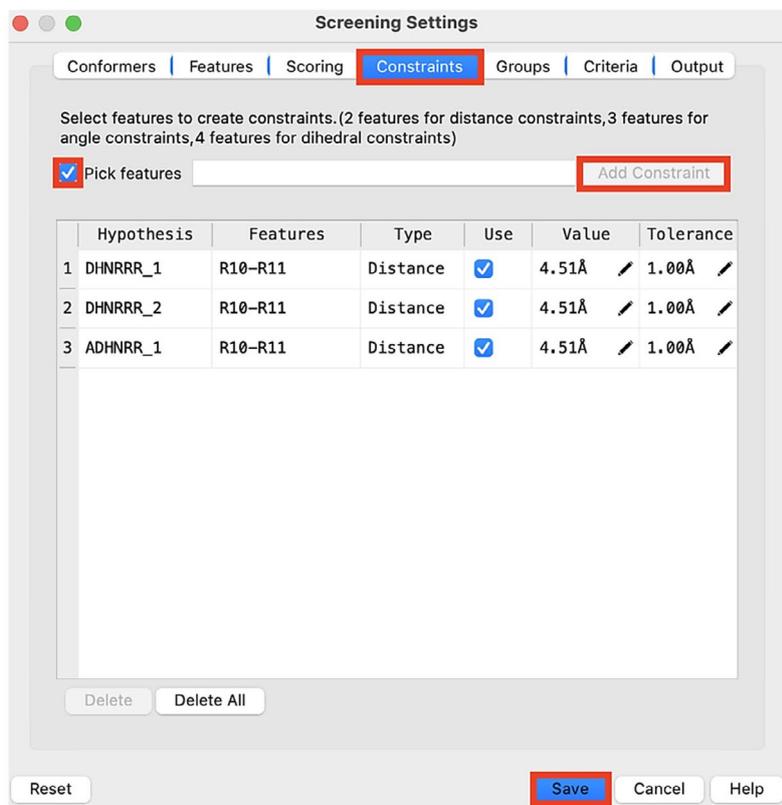


Рис. 3.16. Настройка ограничений

Заметка: на вкладке **Constraints** можно задать ограничения на расстояние, значения валентных и торсионных углов для комбинаций отдельных признаков в гипотезах, используемых при скрининге. Чтобы добавить ограничения к гипотезе, гипотеза должна отображаться в рабочей области. Выделение двух сайтов определяют ограничение расстояния, трех – ограничение угла, а четырех – торсионное ограничение.

Перейдите на вкладку **Conformers** и проверьте, чтобы был выбран пункт **Use Existing Conformers**. Нажмите **Save**. Назовите работу «**at1_dcpH_screen**», затем нажмите **Run**. Расчет на одном процессоре займет примерно 3 минуты.

Заметка: вкладка **Conformers** отвечает за настройки генерации конформаций. Так как мы используем готовую предобработанную базу данных, нам незачем еще раз проводить конформационный поиск, однако вы все равно можете провести его в режиме **on-the-fly** прямо во время фармакофорного скрининга, нажав **Generate Conformation During Search**.

Настройки, доступные на вкладке **Features**, по наполнению и смыслу аналогичны таковым в модуле создания гипотез. Вкладка **Scoring** позволяет провести калибровку значений вкладов оценочной функции (рис. 3.17).

Scoring function: Custom

Consider atom types when computing volume scores

Scoring function

Scoring formula = 1.00 × (1.0 - alignment score / 1.20 Å)
 + 1.00 × vector score
 + 1.00 × volume score
 + 1.00 × included volume score

Alignment score is RMSD between the hypothesis sites and the corresponding matches in ligand sites, and can range from 0Å (best) to typically 1~2Å (there is no upper limit). Cutoff value of 1.2 (default) is recommended; scores higher than this value will cause this term to be negative.
 Vector scores typically range from -1.0 to 1.0 (best).
 Volume scores typically range from 0.0 to 1.0 (best).
 Included volume scores typically range from 0.0 to 1.0 (best).

Рис. 3.17. Параметризация оценочной функции

Вкладка **Groups** позволяет провести группировку сайтов в гипотезе, а **Criteria** – настроить пороговые значения компонентов оценочной функции Phase, определяющие отбраковку лигандов.

После завершения расчета раскройте группу **at1_dcpH_screen** и подгруппу **AHNRRR_1**. Включите и зафиксируйте саму гипотезу **AHNRRR_1** в **Workspace**, затем выберите всю подгруппу **AHNRRR_1**. С помощью клавиш «стрелка вправо» и «стрелка

влево», поочередно включая в **Workspace** лиганды, оцените степень их соответствия гипотезе. Добавьте в **Entry List** свойства **PhaseScreenScore** и **Matched Ligand Sites**. Ознакомьтесь с результатами.

Отразите в отчете: ответьте на вопрос — почему лиганд **CHEMBL480909** в группе **DHNRRR_1** имеет значение **PhaseScreenScore** выше, чем **CHEMBL14502**, хотя демонстрирует худшее соответствие сайтов?

3.2.5. Создание фармакофорной гипотезы на основе структурно-разнообразного набора лигандов

Очистите **Workspace**. Выберите все лиганды группы **DUD-E_act_3chp_trainc**, использовавшейся в предыдущем упражнении, и включите первую запись в **Workspace**.

Заметка: до этого мы генерировали гипотезы, основываясь на выравнивании структурно схожих соединений. Нетривиальную задачу представляет собой получение гипотез из наборов, отличающихся разнообразием. Еще более сложной задачей является выбор лучшей гипотезы из этого множества.

Откройте панель **Develop Pharmacophore Hypothesis: Tasks > Ligand Based Virtual Screening > Develop Pharmacophore Hypothesis**. Выберите **Multiple ligands (selected entries)** в разделе **Create pharmacophore model using**. Фармакофорные сайты будут отображены в **Workspace**. Выберите **Find best alignment and common features**, затем откройте настройки гипотезы. На вкладке **Features**, установите **Hypothesis should match at least** как «25%» и **Number of features in the hypothesis** как «4 to 7». Проверьте, чтобы напротив **Preferred minimum number of features** было установлено значение «5». Установите критерий различий между гипотезами (**Hypothesis difference criterion**) как «1.00» и нажмите **Save** (рис. 3.18).

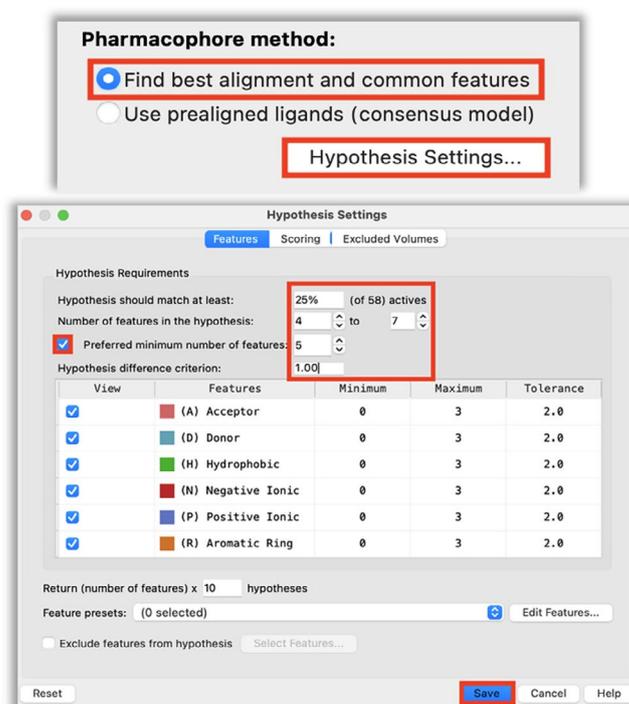


Рис. 3.18. Настройки количества сайтов гипотез и различия гипотез

В окне **Develop Pharmacophore Hypothesis** в разделе настроек конформационного поиска поставьте галочку у пункта **Generate conformers**, нажмите **Options** и установите «50» напротив **Target number of conformers**. Назовите задание как «lkha4_common» и нажмите **Run**. Расчет на 4 процессорах занимает ориентировочно 30 минут (рис. 3.19).



Рис. 3.19. Запуск задания

Для того чтобы просмотреть гипотезы, щелкните дважды по значку **H**. После завершения задания выберите группу **lkha4_common**. Нажмите на значок **H** рядом с заголовком группы. Выберите **Analysis > Detect binding modes**. Установите количество предполагаемых способов связывания (**Proposed number of binding modes**) в открывшемся окне равное «3». Лиганды и гипотезы будут сгруппированы по способам связывания. Познакомьтесь с результатами кластеризации.

3.3. Структурно-ориентированный фармакофорный поиск

3.3.1. Подготовка структур

Загрузите и подготовьте структуру комплекса лейкотриен-А-4 гидролазы с (3*S*)-3-амино-4-оксо-4-[(4-фенилметоксифенил)амино]бутановой кислотой (PDB ID: 3CHP) [9]. В разделе **Preprocess**:

- замените (дополните) атомы водорода с помощью **Replace Hydrogens**;
- уточните порядки связей с использованием базы **CCD**;
- создайте корректные дисульфидные связи и координационные связи с металлами;
- спрогнозируйте состояния протонирования с использованием алгоритма **Epik (pH = 7.0±2)**.

Если препроцессинг белка был выполнен правильно, никаких проблем не будет идентифицировано после нажатия кнопки **Check Structure** в разделе **Diagnose and Analysis**. Не удаляйте подструктуры (YB801, ACT820, IMD850, 4BO901, ZN701) из состава комплекса. Выполните оптимизацию направлений водородных связей и минимизацию согласно стандартному протоколу, описанному в лабораторной работе 1.

3.3.2. Автоматическая генерация фармакофорных гипотез

Включите в **Workspace** запись **3CHP – removed water**, затем дважды щелкните ЛКМ по панели **Presets**. Структура в **Workspace** обновится и станет более удобной для работы.

Откройте панель разработки фармакофорных гипотез: **Tasks > Browse > Ligand-Based Virtual Screening > Develop Pharmacophore Hypothesis**. Выберите **Receptor-ligand complex (Workspace)** в разделе **Create pharmacophore model using**. В качестве метода определения сайтов (**Method**) выберите **Auto (E-Pharmacophore)**, затем в разделе выбора лиганда: **A:4BO (901)**. Откройте панель настройки гипотезы (рис. 3.20).

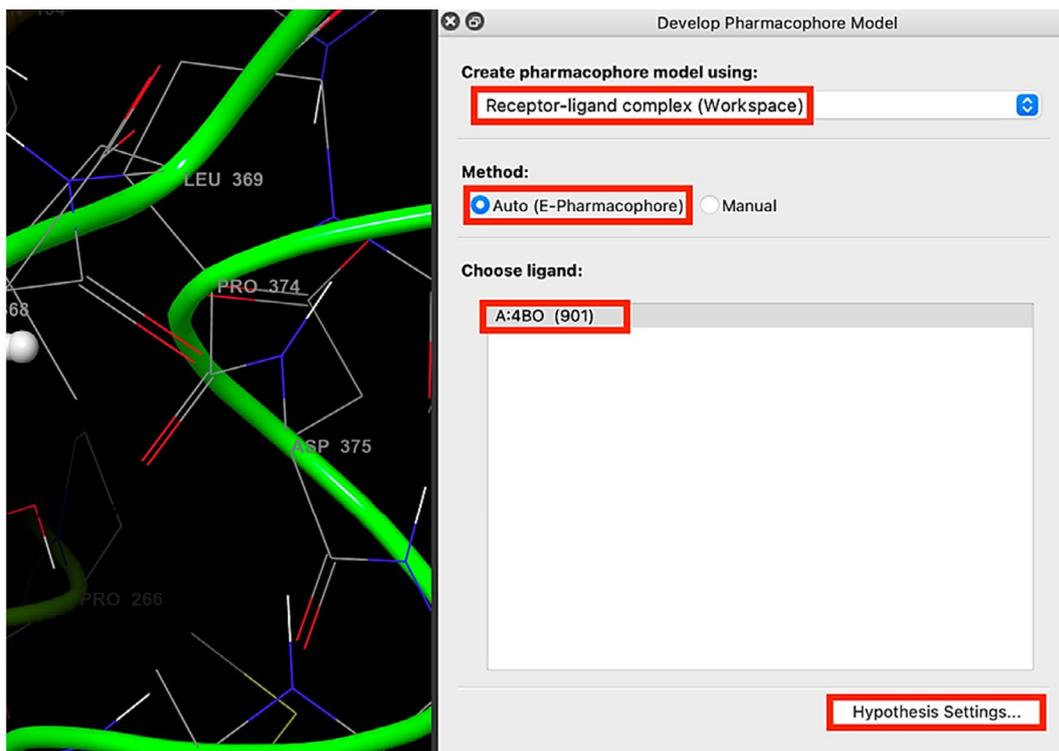


Рис. 3.20. Создание гипотезы

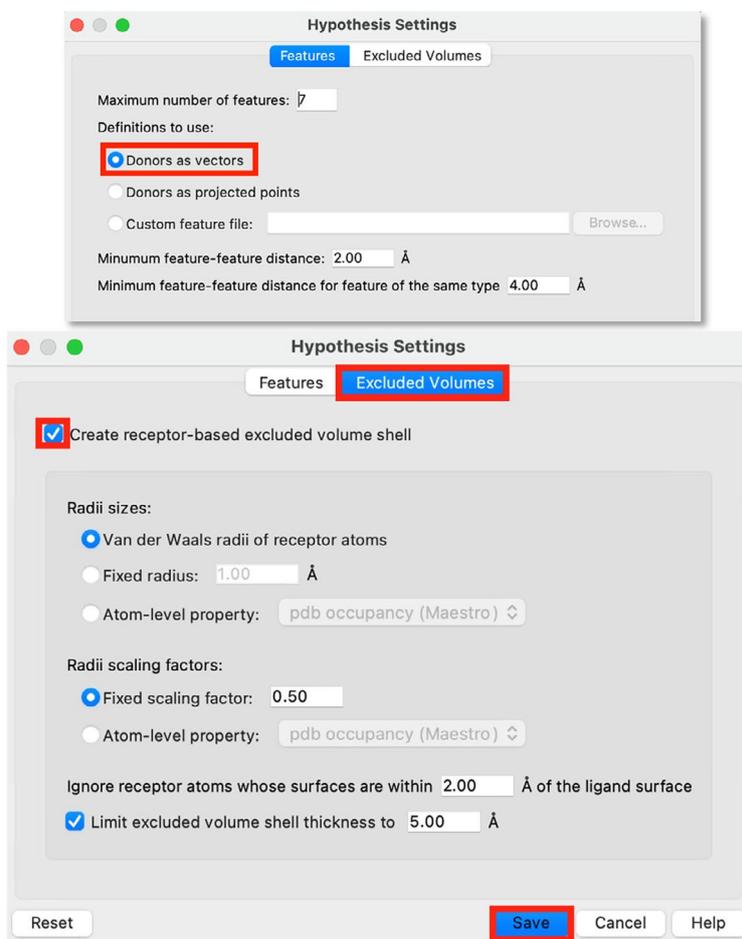


Рис. 3.21. Настройки сайтов гипотезы и создание исключенных объемов

Заметка: автоматический метод включает в себя использование терминов скоринговой функции **Glide XP** [8] для определения сайтов, в наибольшей степени способствующих связыванию. Вы можете вручную построить гипотезу, руководствуясь имеющимися взаимодействиями с рецептором и собственными соображениями.

На вкладке **Features** выберите **Donors as vectors**. Затем создайте исключенные объемы, перейдя на соответствующую вкладку, где выберите **Create receptor-based excluded volume shell**, оставив остальные настройки так, как они установлены по умолчанию. Нажмите **Save** (рис. 3.21).

Измените название задания на «**lkha4_epharm_exvol**» и нажмите **Run**. Расчет займет 3–5 минут.

3.3.3. Ручная корректировка фармакофорных гипотез

Включите в **Workspace** запись самой гипотезы **lkha4_epharm_exvol**, а также запись комплекса **ЗЧР – removed water**. Осмотрите сайты и взаимодействия, определенные **Phase**.

Заметка: в ряде публикаций отмечается взаимодействие между основной цепью остатка TRP311 и связанным лигандом [9, 10]. Поскольку этот сайт отсутствует в ранее созданной гипотезе фармакофора, мы можем добавить его вручную.

В разделе **Method** окна **Develop Pharmacophore Model** выберите **Manual**. Программе может потребоваться время (1–3 минуты) для перехода в режим ручной настройки (рис. 3.22).

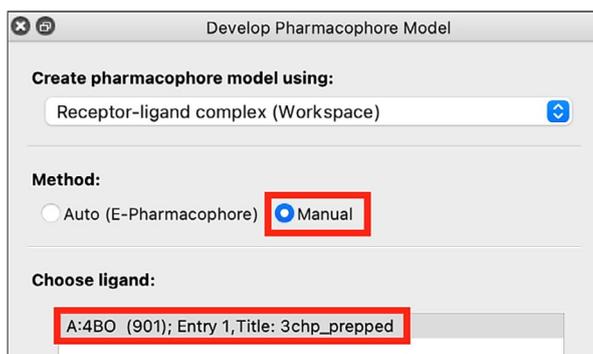


Рис. 3.22. Переход в ручной режим

В появившемся ниже разделе **Show Features** перетащите ползунок до упора вправо (**More**). Откройте окно настройки гипотезы и создайте исключенные объемы, затем нажмите **Save** (рис. 3.23).

Отобразите ароматические водородные связи. Для этого в панели настройки конфигурации **Workspace** нажмите на три точки над значком . Справа в разделе **Non-covalent bonds** выберите **Ligand-Receptor**, затем кликните по **Aromatic H-Bonds**. В разделе **Contacts/Clashes** отключите отображение **Bad Clashes** для удобства работы в **Workspace** (рис. 3.24).

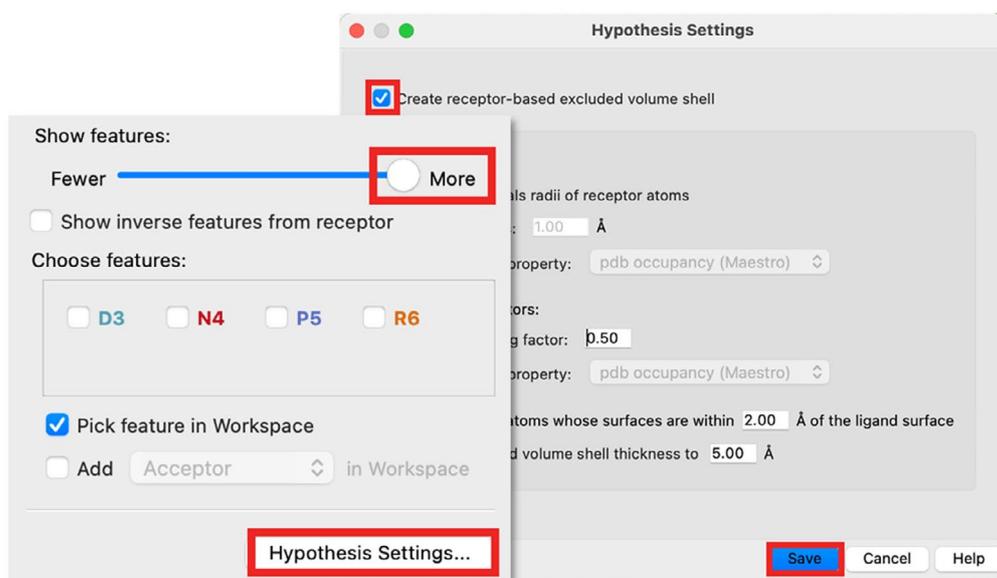


Рис. 3.23. Настройки сайтов гипотезы и создание исключенных объемов

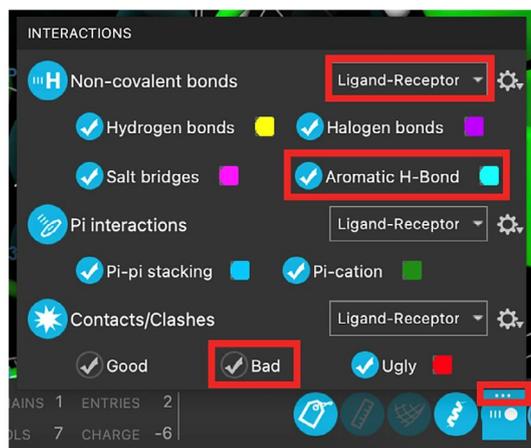


Рис. 3.24. Настройки отображения взаимодействий в Workspace

В панели **Structure Hierarchy**, нажмите на значок поиска и введите «TRP 311». Будет выполнен поиск среди всех остатков записи. Щелкните ЛКМ по строке TRP311 среди результатов поиска, тем самым выбрав эту аминокислоту. В панели настройки фокусировки рядом с **Fit** нажмите на **Fit view to selected atoms**. Теперь фокус в **Workspace** будет смещен на выбранные атомы (рис. 3.25).

Вернитесь к панели **Develop Pharmacophore Model**, в разделе **Choose features** установите флажок напротив **Add** и выберите **Donor in Workspace**. Появится баннер, предлагающий выбрать атом.

Выберите атом водорода кольца R7 лиганда, участвующего в образовании ароматической водородной связи. Вектор сайта D7 появится в **Workspace**, а сам D7 будет добавлен в раздел **Choose features** панели **Develop Pharmacophore Model** (рис. 3.26).

Скорректируйте направление вектора. Щелкните ПКМ по D7 и в открывшемся окне редактирования сайта (**Edit Feature D7**) поставьте галочку напротив пункта **Manually reposition feature**. С зажатой Shift (или Alt) с помощью колесика мыши (или перетаскивания с зажатой ЛКМ) измените направление D7 так, чтобы оно совпадало с направлением ароматической связи, отображенной в **Workspace**. Нажмите ОК (рис. 3.27).

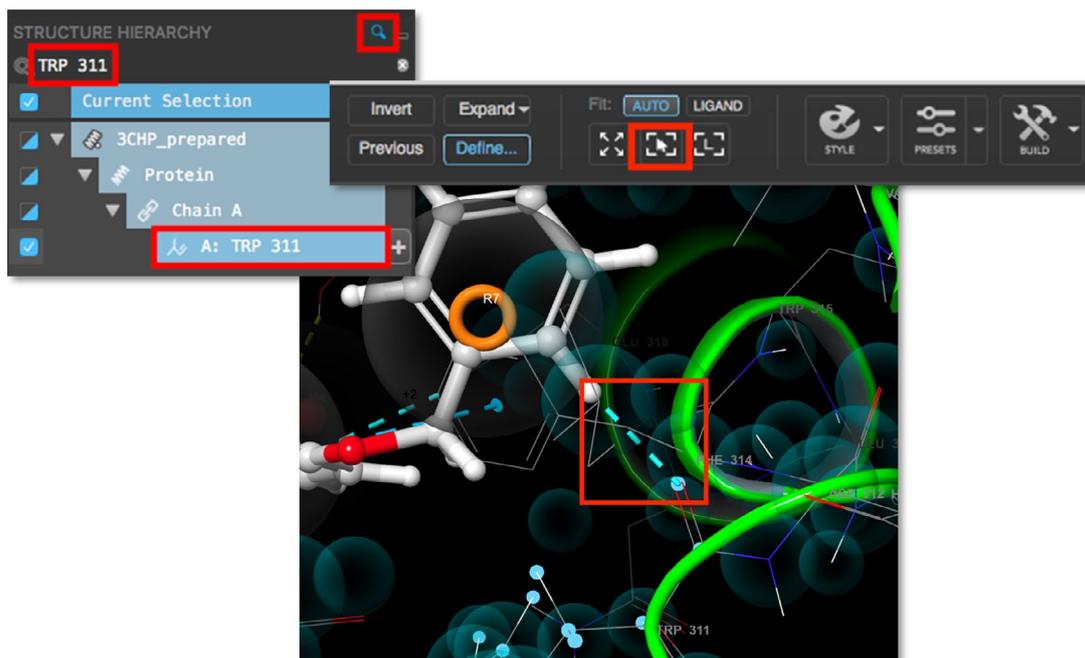


Рис. 3.25. Фокусировка на остатке TRP311 в Workspace

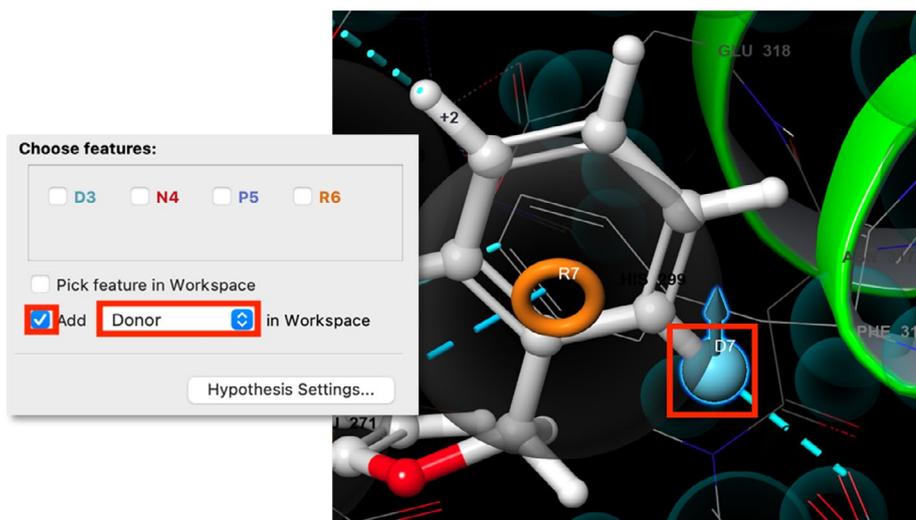


Рис. 3.26. Ручное добавление фармакофорного сайта в гипотезу

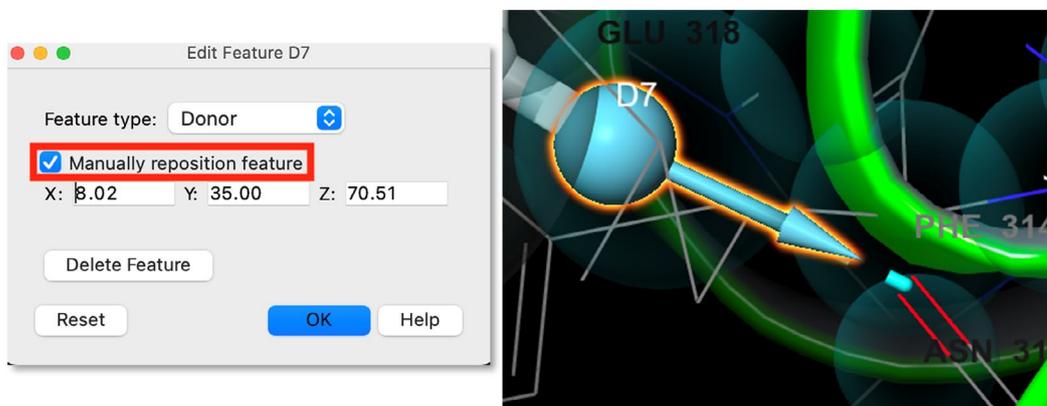


Рис. 3.27. Корректировка направления векторного компонента сайта

В окне **Develop Pharmacophore Model** в разделе **Choose features** выберите оставшиеся пять сайтов, назовите гипотезу «**lkha4_manual**» и нажмите **Create**. Сайт **D3** может быть не обнаружен. Если вы столкнулись с такой ситуацией, его добавление не требуется. Расчет займет 3–6 минут (рис. 3.28).

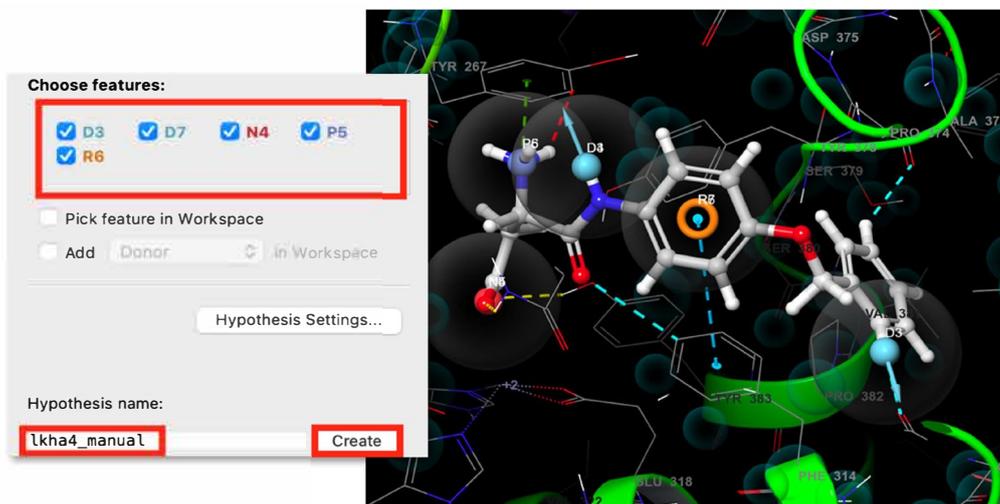


Рис. 3.28. Выбор сайтов и создание гипотезы

Для просмотра свойств сайтов в **Workspace** в **Entry List** напротив записи гипотезы **lkha4_manual** нажмите на значок «Н», затем **Show > Feature Properties** (Гипотеза должна быть включена в **Workspace**) (рис. 3.29).

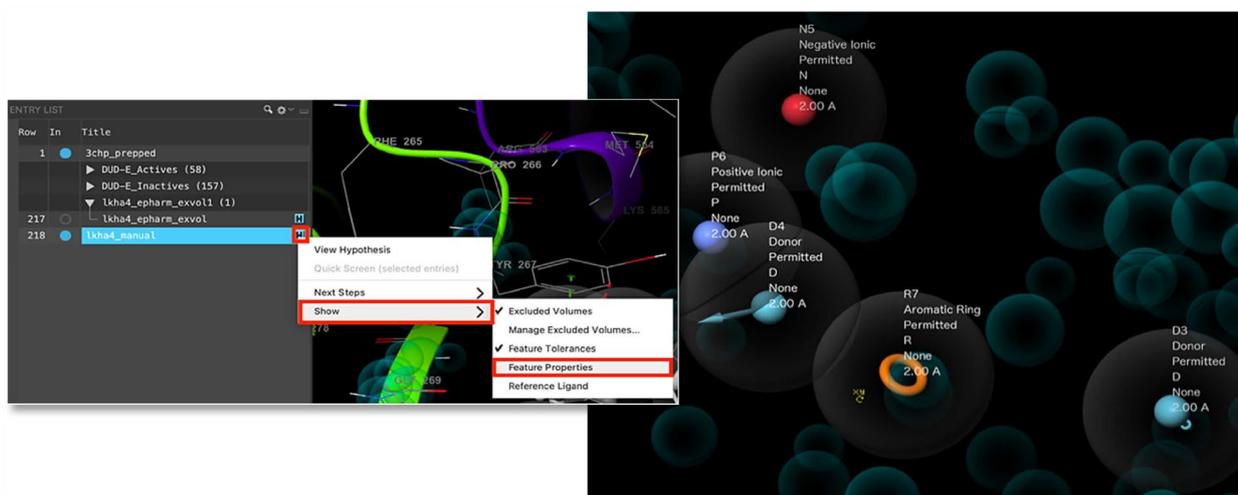


Рис. 3.29. Отображение свойств сайтов

Исключенные объемы также могут быть добавлены вручную. Хотя их генерация в автоматическом режиме происходит чаще всего правильно, иногда может возникнуть потребность в дополнительном создании такого рода сайтов. Добавим исключенный объем вокруг **HIP139**. Для этого также выберите атомы остатка и сфокусируйтесь на них в **Workspace**. Напротив **lkha4_manual** нажмите на «Н» и выберите **Manage Excluded Volumes**. В открывшемся окне **Manage Excluded Volumes** нажмите **Selected Atoms**. Объемы будут созданы вокруг всех атомов остатка. Для удаления или изменения параметров каждого сайта необходимо воспользоваться таблицей ниже. Выберите последнюю строчку и нажмите **Delete**. Нажмите **Close** (рис. 3.30).

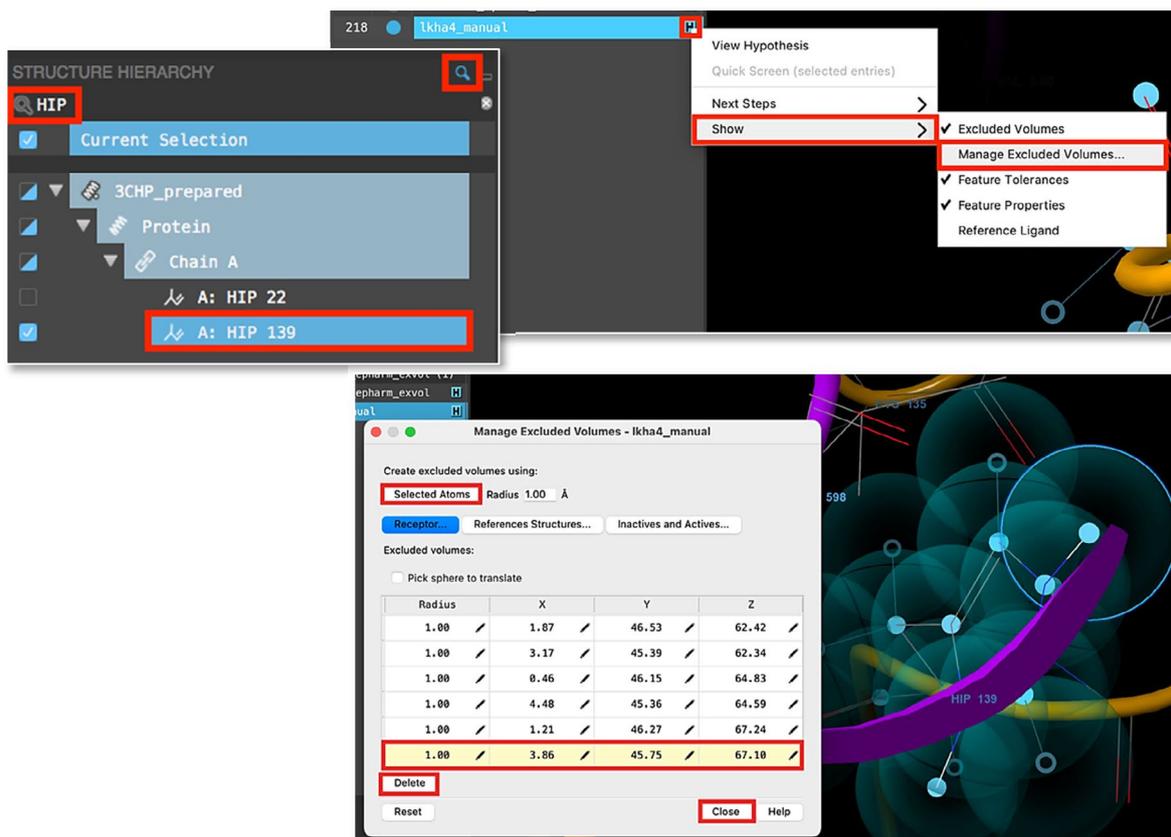


Рис. 3.30. Создание исключенных объемов в ручном режиме

Объедините гипотезы. Это действие позволяет комплексно оценить, какие сайты и свойства важны для получения самой точной модели и избавиться от избыточности. Включите обе гипотезы в **Workspace**, затем нажмите на значок «Н» рядом с каждой из них и выберите **Show > Reference Ligand**. В **Workspace** должны отображаться обе гипотезы и референсные лиганды для них. Вернитесь в окно **Develop Pharmacophore Model** и в разделе **Create pharmacophore hypothesis using** выберите **Merged hypotheses (Workspace)**. В разделе **Choose features** для **from lkha4_manual** выберите сайты **D3, N5, P6, и XVols**, а для **lkha4_epharm_exvol** выберите **R6 and R7**. Назовите работу «**lkha4_merge**» и нажмите **Create** (рис. 3.31).

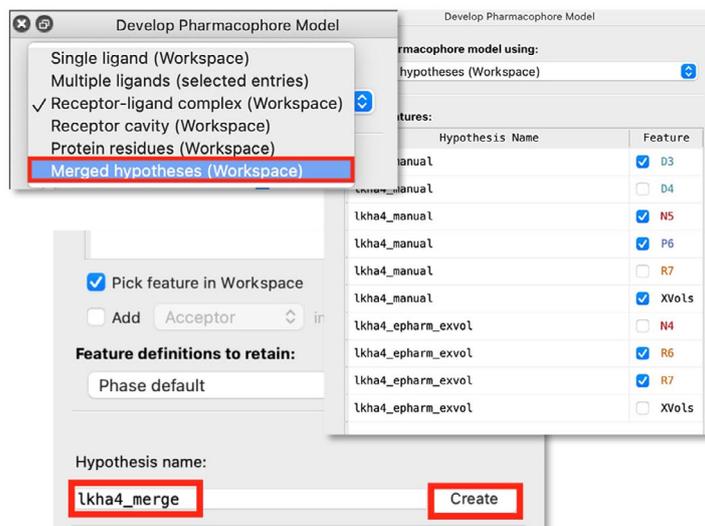


Рис. 3.31. Создание объединенной гипотезы

3.3.4. Фармакофорный скрининг

Импортируйте файлы **DUD-E_Actives.mae** и **DUDE_Inactives.mae**, содержащие лиганды, которые будут использованы при скрининге. Выберите в **Entry List** одноименные группы и включите гипотезу **lkha4_merge** в **Workspace**.

Следуя путем **Tasks > Browse > Ligand-Based Virtual Screening > Ligand Screening**, откройте модуль **Phase Ligand Screening**, в разделе **Ligands to screen** выберите **Project Table (selected entries)**. Нажмите **Add Hypothesis** и выберите **Workspace**. Гипотеза **lkha4_merge** будет добавлена в таблицу. В столбце **Matches** выберите «**3 of 5**». Откройте настройки скрининга. Перейдите на вкладку **Groups**. В **Workspace** с помощью ЛКМ выберите **D1** и **R5**, затем нажмите **Create Group**. Теперь лиганд, который не покажет соответствие хотя бы одному из сгруппированных сайтов, будет отклонен в ходе скрининга, даже если совпадут все остальные сайты. Нажмите **Save** (рис. 3.32).

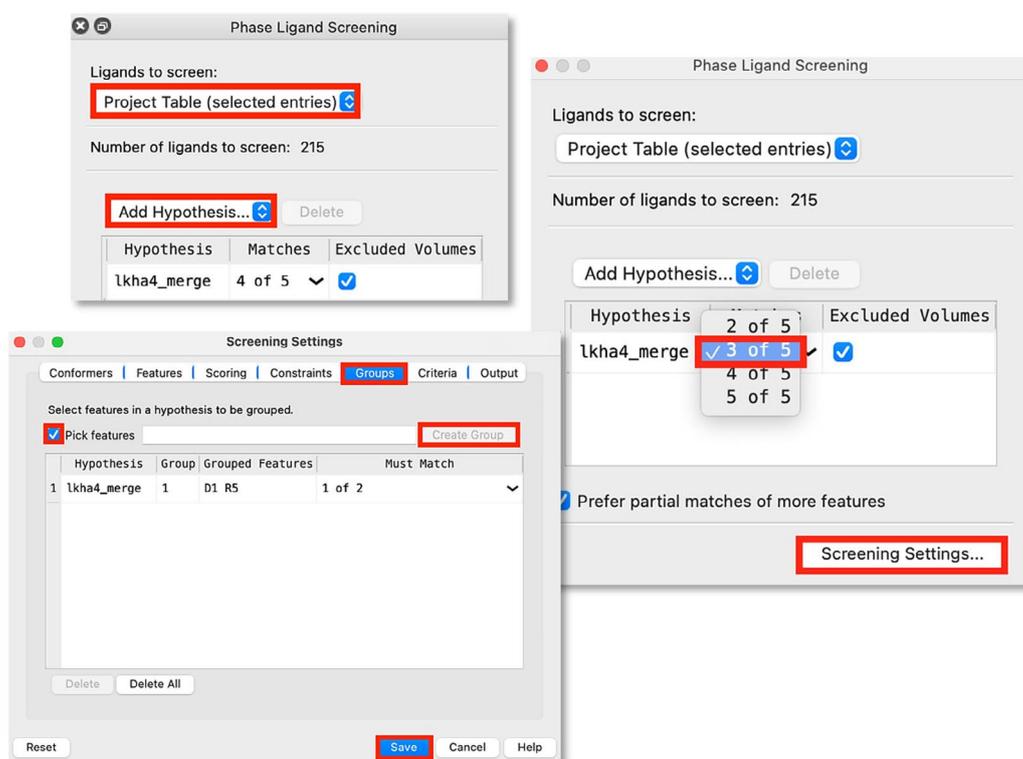


Рис. 3.32. Настройки фармакофорного скрининга

Измените имя задания на «**lkha4_screen**» и нажмите **Run**. Расчет может занять до 10 минут.

Визуализируйте и проанализируйте результаты так же, как вы делали это в случае лиганд-ориентированного фармакофорного моделирования.

Отразите в отчете:

- 1) 2D-фармакофорную карту для **lkha4_merge**;
- 2) результаты скрининга в виде таблицы, в которой отразите значения **Phase-SreenScore**, а также **Mathced Ligand Site**;
- 3) рассчитанные значения чувствительности (**SEN**) и специфичности (**SPC**) для гипотезы **lkha4_merge**. Укажите количества ложноположительных (**FP**), истинно-положительных (**TP**), ложно- и истинно-отрицательных (**FN** и **TN**) результатов по формулам

$$SEN = \frac{TP}{TP + FN},$$

$$SPC = \frac{TN}{TN + FP},$$

где *TP* (англ. True positive), *TN* (англ. True negative), *FP* (англ. False positive), *FN* (англ. False negative) – число соединений, классифицированных как истинно-положительные, истинно-отрицательные, ложноположительные и ложноотрицательные соответственно.

Вопросы к защите

1. Сформулируйте определение понятия «фармакофор».
2. Сформулируйте определение понятия «фармакофорный сайт».
3. Какие виды фармакофорных сайтов вы знаете?
4. Какие атрибуты характеризуют фармакофорные сайты?
5. Что такое исключенные объемы и чем полезны признаки такого типа при выполнении фармакофорного моделирования?
6. Что такое фармакофорный ключ?
7. Какой смысл несут в себе термины $w_cD\Delta E$ и $w_{act}A$ в составе оценочной функции PhaseHypoScore?
8. Когда фармакофорное моделирование может применяться в процессе открытия лекарств?
9. Перечислите ограничения методов фармакофорного моделирования.

Рекомендуемая литература

1. Glossary of terms used in medicinal chemistry : (IUPAC Recommendations 1998) / C. G. Wermuth, C. R. Ganellin, P. Lindberg, L. A. Mitscher // Pure and Applied Chemistry. – 19987. – Vol. 70, № 5. – P. 1129–1143.
2. Preclinical characterization of pirtobrutinib, a highly selective, noncovalent (reversible) BTK inhibitor / E. B. Gomez, K. Ebata, H. S. Randeria [et al.] // Blood. – 2023. – Vol. 142, № 1. – P. 62–72.
3. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results / S. L. Dixon, A. M. Smondyrev, E. H. Knoll [et al.] // Journal of Computer-Aided Molecular Design. – 2006. – Vol. 20, № 10-11. – P. 647–671.
4. Dixon, S. L. PHASE: a Novel Approach to Pharmacophore Modeling and 3D Database Searching / S. L. Dixon, A. M. Smondyrev, S. N. Rao // Chemical Biology & Drug Design. – 2006. – Vol. 67, № 5. – P. 370–372.
5. Schrödinger Release 2023-4: Phase. – New York : Schrödinger, 2023.
6. Pharmacophores and Pharmacophore Searches / Edited by T. Langer, R. D. Hoffmann. – Weinheim : Wiley-VCH, 2006. – XX, 375 p. – ISBN 978-3-527-31250-4.
7. Comprehensive Medicinal Chemistry III / Eds.: S. Chackalamannil [et al.]. – 3rd Ed. – Amsterdam : Elsevier, 2017. – [4536] p. – ISBN 978-0-12-803200-8.
8. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes / R. A. Friesner, R. B. Murphy,

- M. P. Repasky [et al.] // *Journal of Medicinal Chemistry*. – 2006. – Vol. 49, № 21. – P. 6177–6196.
9. Discovery of Leukotriene A4 Hydrolase Inhibitors Using Metabolomics Biased Fragment Crystallography / D. R. Davies, B. Mamat, O. T. Magnusson [et al.] // *Journal of Medicinal Chemistry*. – 2009. – Vol. 52, № 15. – P. 4694–4715.
10. The development of novel LTA4H modulators to selectively target LTB4 generation / C. M. Low, S. Akthar, D. F. Patel [et al.] // *Scientific Reports*. – 2017. – Vol. 7. – Article number 44449.



Лабораторная работа 4

МОЛЕКУЛЯРНЫЙ ДОКИНГ

Цель работы: освоение основных подходов к проведению молекулярного докинга с жестким рецептором для нужд дизайна лекарств.

Вам предстоит:

- 1) подготовить структуры белка и рецептора для проведения молекулярного докинга;
- 2) выполнить расчет сетки потенциалов (грид-бокса) для активного сайта рецептора, а также провести докинг нативного лиганда;
- 3) осуществить процедуру стыковки с использованием различных протоколов точности и сравнить результаты.

Теоретическое введение

Согласно определению, данному в Глоссарии русскоязычных терминов в медицинской химии: «**докинг, молекулярный докинг (docking, molecular docking)** – методы молекулярного моделирования, обеспечивающие быстрый поиск и оценку возможности, места и способа связывания (оптимального соответствия) между лигандом и сайтом связывания биомишени» [1].

Большинство программных решений реализуют стыковку с жестким рецептором, в ходе которой геометрия аминокислотных остатков белка остается неизменной. С точки зрения конформационной подвижности лиганда можно выделить гибкий (flexible) и жесткий (rigid) докинг. В первом случае малой молекуле «позволяется» дополнительная конформационная подвижность, достигаемая путем изменения валентных и торсионных углов, а во втором – нет, и разрешены только вращательные и поступательные движения для всей структуры целиком.

Ранжирование поз происходит с помощью скоринговых функций – математических выражений, позволяющих количественно оценить *эффективность* взаимодействий, *стабильность/выгоду* ориентации лиганда в сайте [2, 3]. *Здесь и далее оценочные функции обозначаются как ΔG , однако вы должны понимать, что получаемые значения не соответствуют истинным свободным энергиям связывания, а являются лишь их оценками.*

Выделяют несколько типов скоринговых функций.

С помощью функций, основанных на силовых полях, энергию связывания вычисляют путем суммарной оценки ван-дер-ваальсовых и электростатических взаимодействий между всеми парами атомов «белок – лиганд», обычно учитывая энтальпийный эффект взаимодействия малой молекулы с мишенью, но пренебрегая энтропийным, хотя в ряде случаев энергия десольватации может учитываться с моделированием растворителя в явном или неявном виде [4, 5]:

$$\Delta G = E_{vdw} + E_{elec},$$

где E_{vdw} и E_{elec} – значения суммарных попарных вкладов ван-дер-ваальсовых и электростатических взаимодействий соответственно.

Эмпирические скоринговые функции приблизительно оценивают энергию связывания комплекса путем суммирования вкладов (термов) важных для связывания взаимодействий, таких как водородные связи, гидрофобные и электростатические контакты и т. д. Обычно уравнение эмпирической функции получается с помощью регрессионного анализа на обучающем наборе комплексов «белок – лиганд» с известным сродством для получения значений и оптимизации коэффициентов энергетических термов [6]:

$$\Delta G = w_0 + w_1 \Delta G_{vdw} + w_2 \Delta G_{HBond} + w_3 \Delta G_{Phob} + w_4 \Delta G_{rot},$$

где w_n – веса энергетических термов, а ΔG_{vdw} , ΔG_{HBond} , ΔG_{Phob} , ΔG_{rot} – вклады ван-дер-ваальсовых, водородных связей, гидрофобных взаимодействий и торсионной энтропии лиганда, соответственно.

Knowledge-based (*англ.* KB) функции основаны на статистическом анализе взаимодействующих пар атомов из доступных в PDB комплексов «белок – лиганд». Эти данные о парных контактах преобразуются в псевдопотенциалы, также известные как

потенциалы средней силы, которые описывают предпочтительную геометрию белок-лигандных взаимодействий [7]:

$$\Delta G = \sum_{i=1}^L \sum_{j=1}^R -k_B T \ln \left(\frac{\rho_{ij}(r)}{\rho_{ij}^*} \right),$$

где k_B и T – константа Больцмана и температура соответственно, а $\rho_{ij}(r)$ и ρ_{ij}^* – плотность парных контактов между атомами i и j на расстоянии r в текущем и референсном комплексах соответственно.

Большая часть программ требует предварительной генерации грид-бокса, ограничивающего область стыковки, внутри которого также заранее рассчитываются атомные потенциалы взаимодействия с помощью различных зондов. Единоразово рассчитанная сетка рецептора будет затем использована для оценки конформаций всех лигандов.

Рабочий процесс алгоритма Glide [8–11], реализованный в ПО Schrödinger, включает шаги, описанные ниже. Как уже говорилось ранее, на первом этапе генерируется грид-бокс, в ходе чего рассчитываются несколько типов молекулярных полей рецептора. Среди остатков, содержащих тиольные или гидроксильные группы, пользователем могут быть указаны те, которым будет позволено дополнительное вращение при выполнении стыковки.

Первоначальный конформационный поиск позволяет исчерпывающе исследовать конформационное пространство лиганда. Конформации представляются в виде жесткого ядра (*англ.* core) и прикрепленных к нему вращаемых групп (*англ.* rotatable groups).

Полученные конформации затем проходят серию иерархических фильтров и оценок, представленных ниже (рис. 4.1).

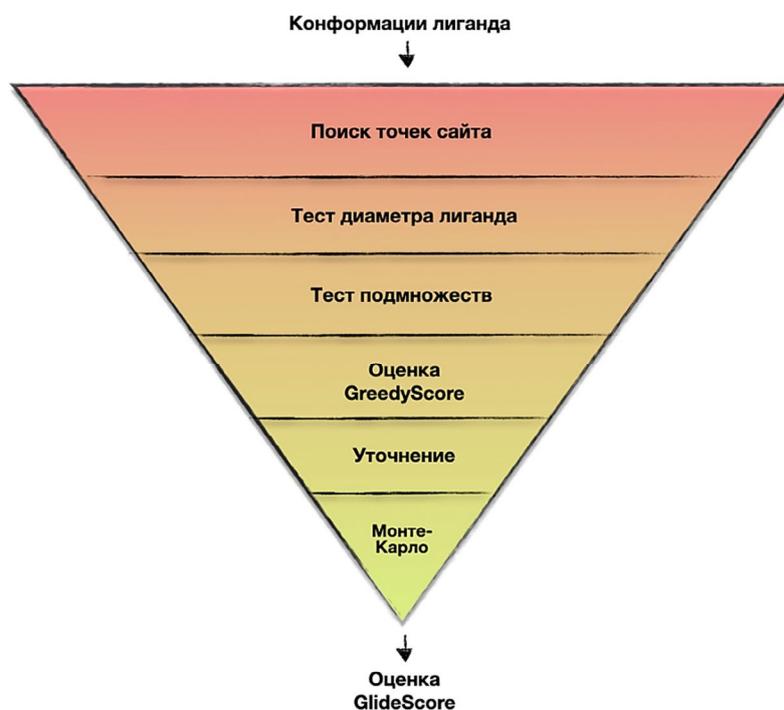


Рис. 4.1. «Воронка» Glide [8]

Для конформаций ядра выполняется поиск возможных положений и ориентаций. Поиск начинается с выбора точек сайта, которые может занимать ядро, определяемых

на сетке с шагом 2 Å. Чтобы сделать этот выбор, предварительно вычисленные расстояния от точки сайта до поверхности рецептора сравниваются с расстояниями от центра лиганда до поверхности лиганда. Для гибкой стыковки центр лиганда определяется как середина двух наиболее удаленных друг от друга атомов. Линия, проходящая через эти атомы, называется диаметром лиганда. Glide позиционирует центр лиганда в точке сайта, если имеется достаточно хорошее совпадение расстояний.

На втором этапе исследуется размещение атомов, лежащих на заданном расстоянии от оси диаметра лиганда, для заранее заданного выбора возможных ориентаций диаметра (тест диаметра лиганда). Поза пропускается при наличии стерических столкновений с рецептором.

Далее лиганд поворачивается вокруг диаметра и рассчитывается подмножество атомов, способных образовывать водородные связи с рецептором (тест подмножеств). Если эта оценка удовлетворительна, оцениваются все взаимодействия с рецептором с помощью ограниченной версии основной функции скоринга GreedyScore. Здесь же реализуется этап уточнения, в ходе которого лигандам позволяет жестко двигаться в пределах 1 Å и быть повторно оцененными.

Небольшое количество уточненных поз (обычно не более 400) минимизируется с помощью силового поля OPLSP [12]. Наконец, несколько поз с наименьшей энергией, полученных таким образом, подвергаются процедуре поиска конформаций Монте-Карло и повторной минимизации. Финальный скоринг происходит с использованием функции Glide GScore:

$$\begin{aligned} \Delta G = & C_{lipo-lipo} \sum f(r_{lr}) + C_{HBond} \sum g(\Delta r) h(\Delta \alpha) + C_{coul} E_{coul} + \\ & + C_{vdW} E_{vdW} + C_{polar-phob} V_{polar-phob} + C_{metal-ion} \sum f(r_{lm}) + \\ & + solv.term + C_{rota} E_{rota}, \end{aligned}$$

где первый член (*lipo-lipo*) включает зависимую от расстояния оценку гидрофобных взаимодействий всех пар атомов лиганд-рецептор (*lr*), которые определяются как липофильные; второй (*Hbond*) кодирует энергетический терм водородных связей. Функции *f*, *g* и *h* дают полную оценку расстояниям Δ*r* или углам Δα, когда те принимают значения в пределах определенных порогов. Основной вклад в GScore вносят энергии кулоновского и ван-дер-ваальсового взаимодействий (третий и четвертый члены). Пятый член оценивает вклад полярных, но не имеющих водородных связей атомов в гидрофобных областях. Термин *metal-ion* кодирует энергию взаимодействий лиганд-металл (*lm*); слагаемое *solv.term* неявно учитывает эффекты растворителя и десольватации лиганда, а последний терм назначает «штраф» за большое количество вращающихся связей [8].

Кроме того, алгоритм Glide позволяет проводить стыковку в трех режимах точности:

- HTVS (*англ.* High-throughput Virtual Screening) – предназначен для быстрого скрининга очень большого количества лигандов. HTVS имеет ограниченную конформационную выборку, а расширенные настройки для него недоступны [13];
- SP (*англ.* Standard Precision) – более точный протокол, реализующий исчерпывающий поиск ориентаций лиганда в сайте, в том числе в режиме гибкой стыковки. Протокол SP установлен по умолчанию при открытии панели Glide в Maestro [13];

- **XP** (*англ.* Extra Precision) – мощная и избирательная процедура, выполнение которой занимает больше времени, чем докинг, выполняемый на остальных протоколах. XP предназначен для повторной оценки поз лигандов, имеющих высокий балл по результатам HTVS/SP и отсеивания ложноположительных результатов. Рекомендуется сначала применить SP, затем взять от 10 до 30 % окончательных лучших поз и пристыковать их повторно с помощью XP [10].

При запуске стыковки в режиме повышенной точности пользователь может получить дополнительную информацию о значениях следующих энергетических терминов скоринговой функции XP GScore (табл. 4.1):

Таблица 4.1

Дескрипторы XP [13]

Дескриптор	Описание
GScore	Полный GlideScore. Сумма всех дескрипторов
LipophilicEvdW	Терм липофильности, определяемый по степени комплементарности гидрофобных поверхностей лиганда и рецептора
PhobEn	Награда за размещение гидрофобных частей лиганда в локальных гидрофобных областях сайта
PhobEnHB	Награда за водородные связи, образованные в локальных гидрофобных областях сайта
PhobEnPairHB	Награда за коррелированные (парные) водородные связи, образованные в локальных гидрофобных областях сайта
HBond	Дескриптор, кодирующий вклад водородных связей
Electro	Дескриптор, кодирующий вклад электростатических взаимодействий
π -Cat	Награда за π -катионные взаимодействия
ClBr	Награда за атомы Cl/Br, находящиеся в гидрофобном окружении, которые взаимодействуют с Asp или Glu
LowMW	Награда для лигандов, обладающих небольшим молекулярным весом
Penalties	Суммарный штраф за невыгодную ориентацию полярных групп лиганда, десольватацию и внутримолекулярные контакты лиганда
HBPenal	Штраф лигандам с большим числом гидрофобных контактов и низкими показателями XP HBond
ExposPenal	Штраф за группы лиганда, ориентирующиеся в область растворителя
RotPenal	Штраф за вращающиеся связи

Экспериментальная часть

4.1. Создание проекта

В своем личном каталоге создайте папку LW4. Дважды щелкните по ярлыку Maestro на рабочем столе. На верхней панели перейдите на вкладку **File > Change**

Working Directory. В открывшемся окне укажите путь к рабочей директории LW4 в своей личной папке.

Затем перейдите **File > Save project as**, задайте проекту имя LW4 в графе **File name**, укажите путь к рабочей директории в графе **Location**.

4.2. Подготовка к молекулярному докингу

4.2.1. Подготовка структур

С помощью модуля **Protein Preparation Workflow** выполните загрузку и подготовку комплекса фактора свертывания Ха (PDB ID: 1FJS [14]) с ингибитором ZK-807834 (CI-1031) в соответствии со стандартным протоколом, описанным в лабораторной работе 1. Дважды нажмите на панель **Presets** для настройки удобной визуализации структур в **Workspace**. Выполните сплитирование (**Ligands, Waters, Other**). Подготовьте нативный лиганд с помощью модуля **LigPrep**. Используйте стандартный протокол ($\text{pH}=7.0\pm 2$), имя задания: «**ligprep_1fjs**».

4.2.2. Генерация сетки потенциалов

Выделите запись подготовленного комплекса в **Entry List**, затем откройте модуль генерации сетки потенциалов: **Tasks > Browse > Receptor-Based Virtual Screening > Receptor Grid Generation**. В открывшемся окне на вкладке **Receptor** определите область для стыковки. Для этого в разделе **Define Receptor** поставьте галочки напротив пунктов **Pick to identify the ligand (Molecule)** и **Show markers**. Затем в **Workspace** щелкните ЛКМ по любому атому лиганда (рис. 4.2).

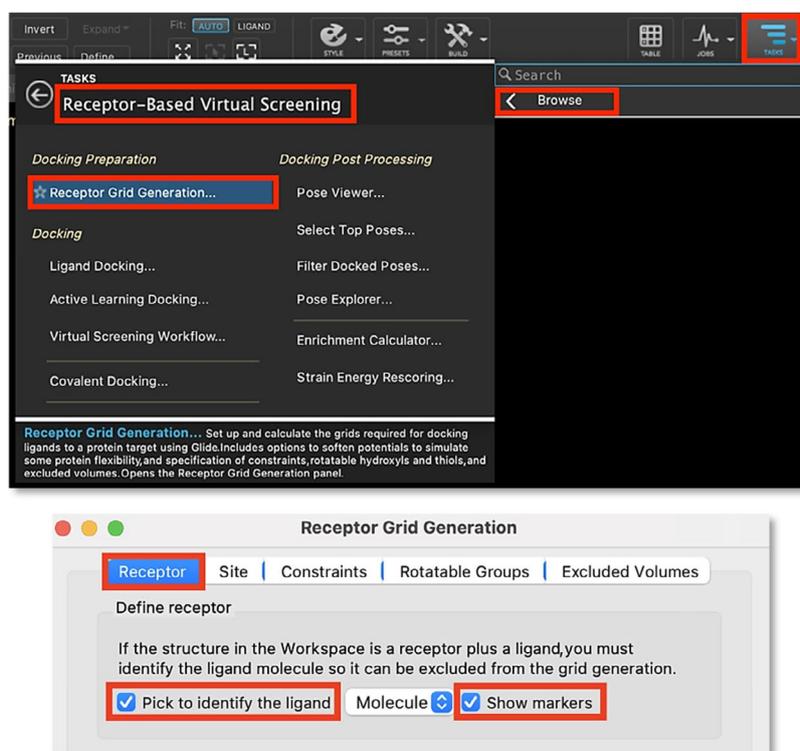


Рис. 4.2. Открытие модуля подготовки грид-бокса, определение лиганда и области стыковки

Важно: определить область стыковки можно не только по координатам лиганда! Точки сайтов, рассчитанные **SiteMap** [14], и координаты остатков также могут быть использованы для этого. Однако гораздо важнее, что в случае построения сетки потенциалов на основе белково-лигандного комплекса малая молекула не должна учитываться при расчете. Указание лиганда позволяет вывести его из уравнения при генерации грид-бокса [13].

После определения лиганда его атомы будут подсвечены, а в **Workspace** отобразится область докинга в виде куба с ребрами пурпурного цвета. Центр куба соответствует центру масс молекулы ингибитора. Перейдите на вкладку **Site**. Проверьте, чтобы напротив пункта **Centroid of Workspace ligand (selected in the Receptor tab)** стояла галочка. Откройте расширенные настройки. Установите длину ребра как «**10 Å**» (рис. 4.3). При этом в **Workspace** появится зеленый бокс меньшего размера, чем пурпурный. Этот бокс определяет область, в которой должен находиться центр масс стыкуемой молекулы, чтобы преодолеть начальные стадии докинга.

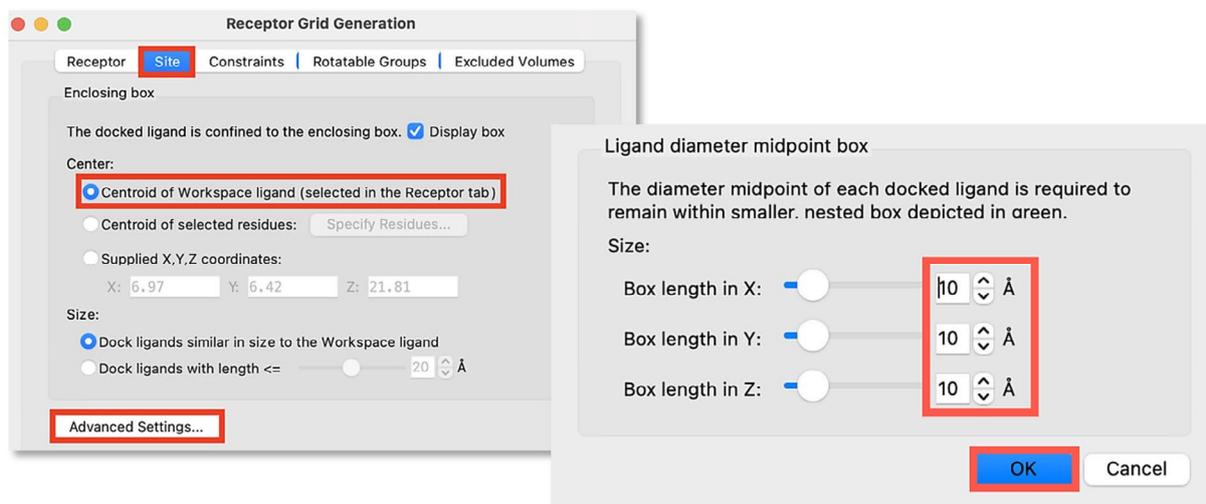


Рис. 4.3. Настройки области стыковки

Заметка: по мнению Адлера и коллег [15], солевой мостик, образующийся между ингибитором и ASP189, наблюдаемый в кристаллической структуре комплекса, способствует эффективности лиганда. Установим ограничение именно для этой конкретной водородной связи.

Установим ограничения на образование водородных связей в ходе стыковки. Для удобства настоятельно советуем в панели **Quick Select** нажать на **три точки**, затем **Residues** в **Interactions > Inverse**. Таким образом, в структуре комплекса выделятся все остатки, которые не участвуют во взаимодействиях малой молекулы с рецептором. Скройте их.

Перейдите на вкладку **Constraints > H-Bond/Metal**. В **Workspace** выберите атом кислорода остатка **ASP189** (рис. 4.4).

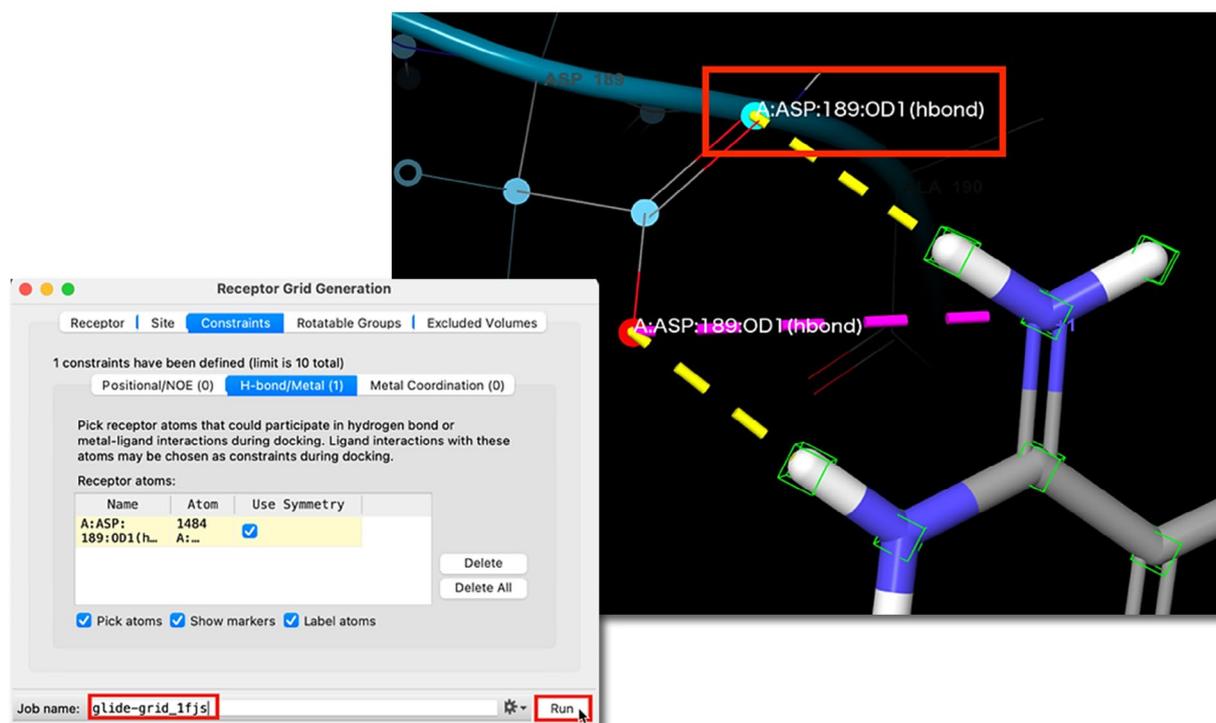


Рис. 4.4. Создание ограничений

Запись будет добавлена в таблицу на вкладке, а сама карбоксильная группа подсвечена. Назовите задание «**glide-grid_1fjs**» и нажмите **Run**. Расчет займет 5–6 минут.

4.3. Докинг нативного лиганда

Очистите **Workspace**. Выделите в **Entry List** заголовок нативного лиганда в группе **split**. Откройте модуль молекулярного докинга: **Tasks > Browse > Receptor-Based Virtual Screening > Ligand Docking**. В разделе **Receptor grid** нажмите **Browse** и выберите файл грид-бокса **glide-grid_1fjs.zip**. В открывшемся окне нажмите **OK** (рис. 4.5).

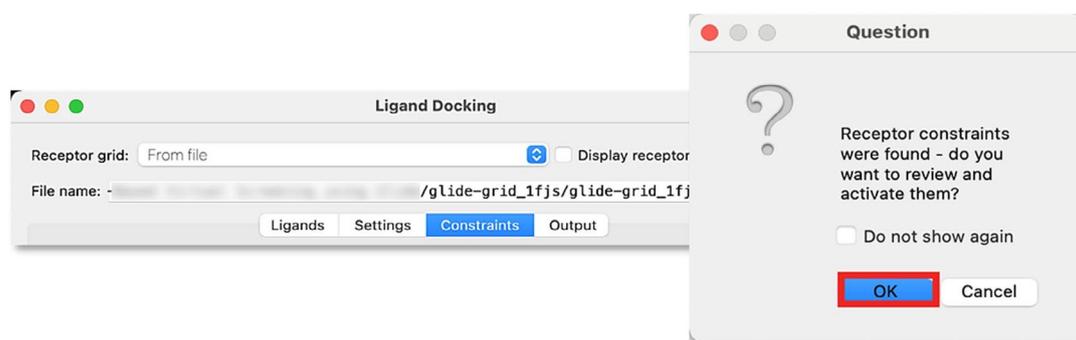


Рис. 4.5. Указание файла сетки потенциалов

Вы окажетесь на вкладке **Constraints**. Поставьте галочку в столбце **Use** напротив **ASP 189**. Теперь те позы, в которых лиганд не образует взаимодействия с этим остатком, будут отброшены. Затем переместитесь на вкладку **Settings**, в разделе **Precision** выберите **SP (standart precision)**, а в разделе **Ligand sampling – flexible**. **Glide** выполнит докинг в режиме стандартной точности с подвижным лигандом. На вкладке **Output**

поставьте галочку рядом с пунктом **Compute RMSD to input ligand geometries**. Расчет RMSD поможет нам количественно оценить, насколько точно **Glide** сможет воспроизвести кристаллическую конформацию лиганда (рис. 4.6).

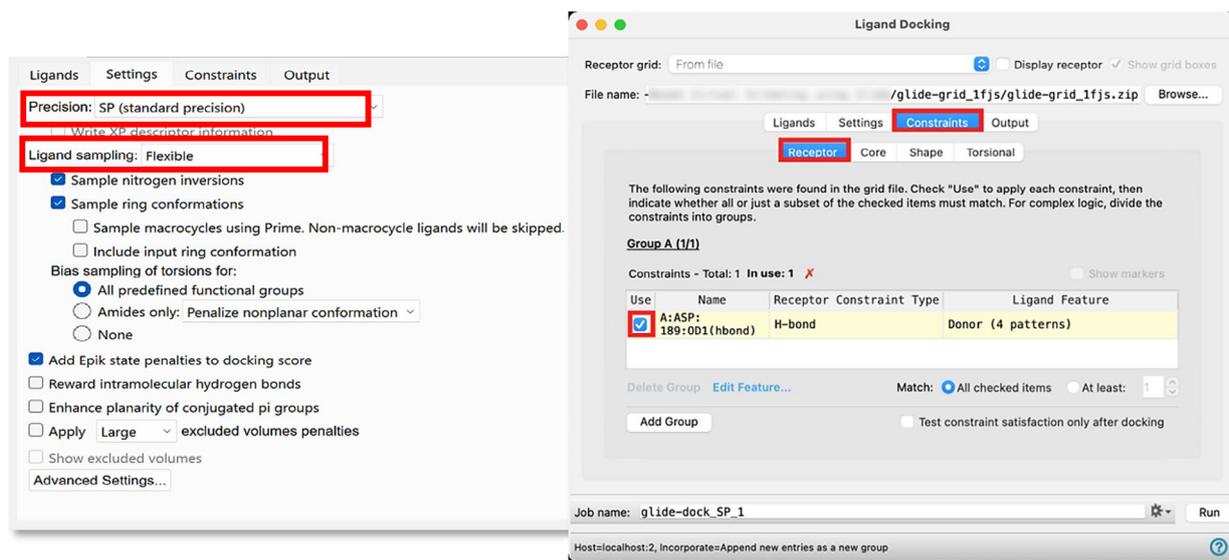


Рис. 4.6. Указание ограничений, режима точности и подвижности лиганда

На вкладке **Ligands** выберите **Files** и укажите файл **ligprep: -out.maegz** в разделе **Use ligands from**. Измените имя задания на «**glide_1FJS_cognate_Hbond**» и нажмите **Run**. Расчет на одном процессоре займет 5–10 минут (рис. 4.7).

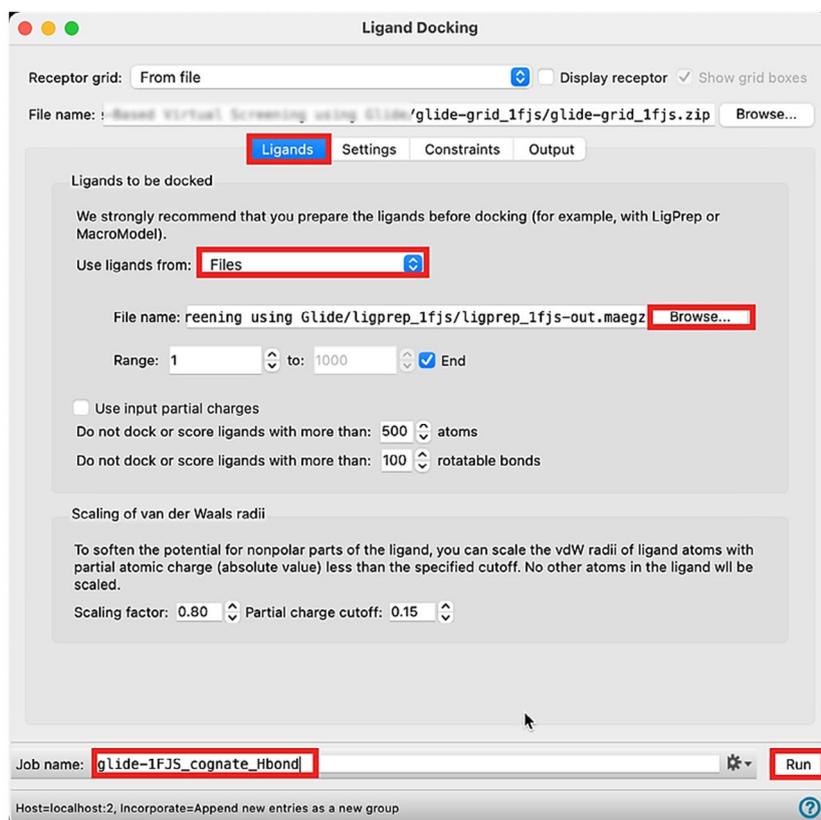


Рис. 4.7. Запуск расчета стыковки

Если позы найдены, то результаты расчета автоматически будут импортированы в **Entry List**. Первая запись лиганда в группе соответствует комплексу с самым лучшим значением скоринга. Нажмите на иконку рядом с заголовком группы, затем — **View poses**, в открывшейся панели **Pose Viewer** выберите **Set Up Poses**. Осмотрите сгенерированные позы.

Отразите в отчете: снимок **Workspace** с наложением лучшей позы среди результатов докинга и конформации кристаллического лиганда. Также приведите значения **Docking Score**, **Emodel**, **Glide GScore** и **RMSD**.

4.4. Докинг виртуальной библиотеки с использованием Glide

Импортируйте файл **50ligs_epik.maegz**, выберите всю группу в **Entry List**. Осуществите молекулярный докинг **Glide** с помощью модуля **Ligand Docking**. В разделе **Use ligands from** на вкладке **Ligands** выберите **Project Table (selected entries)**. Запустите два задания **Glide**: одно — в режиме **flexible, Glide XP** (укажите **Write XP Descriptor Information** на вкладке **Settings**), второе — в режиме **flexible, Glide SP**. Выберите на вкладке **Constraints** остаток **ASP 189**. Задачи назовите «**glide-dock_50ligs_epik_f_XP**» и «**glide-dock_50ligs_epik_f_SP**» соответственно. Расчет суммарно займет не более часа. Оцените результаты стыковки после импорта.

Отразите в отчете:

- 1) таблицу, содержащую информацию о скоринге лигандов, для которых были найдены позы; значения скоринговых функций можно найти в **Project Table > Property Tree: XP GScore** и **Glide GScore** соответственно. Отсортируйте результаты по возрастанию значения функции (табл. 4.2).

Таблица 4.2

Результаты молекулярного докинга

XP точность		SP точность	
Лиганд	XP GScore, ккал/моль	Лиганд	Glide GScore, ккал/моль

- 2) значения **точности** (**ACC**, *англ. Accuracy*), **чувствительности** (**SEN**, *англ. Sensitivity*) и **специфичности** (**SPC**, *англ. Specificity*), рассчитанные на основе результатов как XP, так и SP докинга. Считайте активными лиганды, обладающие значениями скоринга ≤ 7.000 ккал/моль:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN},$$

$$SEN = \frac{TP}{TP + FN},$$

$$SPC = \frac{TN}{TN + FP},$$

где *TP* (англ. True positive), *TN* (англ. True negative), *FP* (англ. False positive), *FN* (англ. False negative) – число соединений, классифицированных как истинно-положительные, истинно-отрицательные, ложноположительные и ложноотрицательные, соответственно.

По вашему мнению, позволяет ли использование алгоритма **Glide XP** отсеивать ложноположительные отклонения?

Вопросы к защите

1. Дайте определение понятию «молекулярный докинг».
2. Чем отличаются жесткий и гибкий (с точки зрения лиганда) подходы к стыковке?
3. Дайте определение понятию «скоринговая функция».
4. Какие виды скоринговых функций вы знаете?
5. Зачем перед проведением стыковки выполнять расчет сетки потенциалов (грид-бокса)?
6. Какие требования предъявляются к структурам лигандов и белка для выполнения докинга?
7. Перечислите основные виды алгоритмов стыковки.
8. Зачем выполняется процедура нативного докинга?
9. Перечислите основные достоинства методов молекулярной стыковки в контексте открытия лекарств.
10. Перечислите основные ограничения методов молекулярной стыковки в контексте открытия лекарств.
11. Какие взаимодействия вносят наибольший вклад в энергию взаимодействия лиганда с белком?

Рекомендуемая литература

1. Глоссарий русскоязычных терминов в медицинской химии / О. Н. Зефирова, К. В. Балакин, М. Ю. Красавин [и др.] // Известия Академии наук. Серия химическая. – 2019. – № 12. – С. 2381–2395.
2. Molecular modeling : Basic Principles and Applications / H.-D. Holtje, W. Sippl, D. Rognan, G. Folkers. – 3rd, Revised and Expanded Ed. – Weinheim : Wiley-VCH, 2008. – X, 310 p. – ISBN 978-3-527-31568-0.
3. Docking Screens for Drug Discovery / Edited by W. F. de Azevedo. – New York : Humana Press, 2019. – XVII, 286 p. – (Methods in Molecular Biology ; Vol. 2053). – ISBN 978-1-4939-9751-0.
4. *Kramer, B.* Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking / B. Kramer, M. Rarey, T. Lengauer // PROTEINS: Structure, Function, and Bioinformatics. – 1999. – Vol. 37, № 2. – P. 228–241.
5. *Li, J.* An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking / J. Li, A. Fu, Le Zhang // Interdisciplinary Sciences: Computational Life Sciences. – 2019. – Vol. 11, № 2. – P. 320–328.
6. *Guedes, I. A.* Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges / I. A. Guedes, F. S. S. Pereira, L. E. Dardenne // Frontiers in Pharmacology. Sec. Experimental Pharmacology and Drug Discovery. – 2018. – Vol. 9. – Article number 01089.

7. *Gohlke, H.* Knowledge-based scoring function to predict protein-ligand interactions / H. Gohlke, M. Hendlich, G. Klebe // *Journal of Molecular Biology*. – 2000. – Vol. 295, № 2. – P. 337–356.
8. *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy* / R. A. Friesner, J. L. Banks, R. B. Murphy [et al.] // *Journal of Medicinal Chemistry*. – 2004. – Vol. 47, № 7. – P. 1739–1749.
9. *Schrödinger Release 2023-4: Glide*. – New York : Schrödinger, 2023.
10. *Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes* / R. A. Friesner, R. B. Murphy, M. P. Repasky [et al.] // *Journal of Medicinal Chemistry*. – 2006. – Vol. 49, № 21. – P. 6177–6196.
11. *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening* / T. A. Halgren, R. B. Murphy, R. A. Friesner [et al.] // *Journal of Medicinal Chemistry*. – 2004. – Vol. 47, № 7. – P. 1750–1759.
12. *Jorgensen, W. L.* Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids / W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives // *Journal of the American Chemical Society*. – 1996. – Vol. 118, № 45. – P. 11225–11236.
13. *Glide 6.7 : User Manual ; Glide : version 6.4 ; Prime version 3.7*. – New York : Schrödinger, 2015. – 15 p.
14. *Halgren, T. A.* Identifying and Characterizing Binding Sites and Assessing Druggability // *Journal of Chemical Informatics and Modeling*. – 2009. – Vol. 49, № 2. – P. 377–389.



Лабораторная работа 5

ИДЕНТИФИКАЦИЯ САЙТА СВЯЗЫВАНИЯ

Цель работы: освоение основных подходов к поиску, аннотированию и оценке потенциальных сайтов связывания в структуре белка.

Вам предстоит:

- 1) освоить инструмент оценки качества белковой структуры Protein Reliability Report;
- 2) произвести поиск потенциальных сайтов связывания малых молекул на поверхности киназного домена Vcr-ABL с помощью алгоритма SiteMap;
- 3) с помощью все того же модуля выполнить углубленную оценку одиночного сайта связывания;
- 4) проанализировать различия и сходства между сайтами связывания в апо- и холо-структурах киназных доменов Vcr-ABL.

Теоретическое введение

Идентификация потенциальных мест связывания лигандов является нужной и полезной процедурой в процессе структурно-ориентированного открытия лекарств как в случае, когда сайт связывания неизвестен, так и в случае, когда стоит задача дизайна аллостерического модулятора мишени.

Алгоритм SiteMap [1–3] позволяет провести быстрый, но исчерпывающий поиск по всей поверхности белка-мишени с целью идентификации, оценки и ранжирования всех потенциальных областей, в которых может связаться малая молекула.

На первом этапе SiteMap выполняет лишь обнаружение сайтов. Алгоритм поиска начинается с построения сетки с шагом в 1 Å вокруг всего белка или внутри области, выбранной пользователем. Каждая точка сетки затем классифицируется как расположенная «внутри» или «снаружи» рецептора путем сравнения квадратов расстояния до близлежащих атомов белка (по умолчанию 2.5 Å для классификации точки как расположенной «внутри»). «Внешние» точки все равно исследуются для определения тех из них, что находятся в хорошем контакте с рецептором. Точки сетки, соответствующие обоим критериям, добавляются в список точек-кандидатов. Следующий шаг объединяет точки в группы. Чтобы точка сайта рассматривалась как член группы, некоторое минимально допустимое количество точек-кандидатов (три по умолчанию) должно находиться в пределах заданного расстояния от нее (1.76 Å по умолчанию). Точки сайта, не попавшие ни в одну группу, отбрасываются. На последнем этапе группы точек сайта объединяются, когда длина разрыва между ними, измеряемая расстоянием наибольшей близости, возникает в области растворителя и меньше или равна указанному расстоянию (6.5 Å по умолчанию).

Создание полей молекулярных взаимодействий осуществляется с помощью зондов, оценивающих зависимость от расстояния электростатические и ван-дер-ваальсовы взаимодействия. Зонд, грубо моделирующий молекулу воды, представляет собой сферу радиусом 1.6 Å и глубиной потенциальной ямы 0.13 ккал/моль.

Сайты визуализируются на поверхности рецептора так, как показано на рисунке ниже (рис. 5.1).

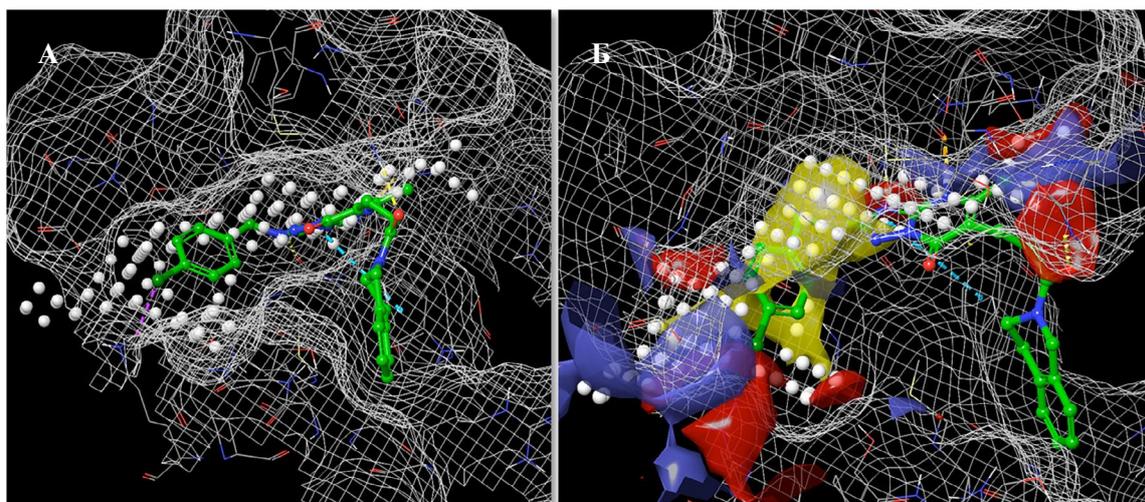


Рис. 5.1. Визуализация сайтов SiteMap: А – сайт связывания АТФ в каталитической субъединице P3Kα с ингибитором (PDB ID: 5UBR [4]), показаны точки сайтов; Б – тот же сайт, но показаны гидрофобные области (желтые), карты доноров (синий) и акцепторов (красный) водородных связей

Кроме того, SiteMap позволяет оценить сайты с помощью двух параметров: SiteScore и DScore:

$$\text{SiteScore} = 0.0733\sqrt{n} + 0.6688e - 0.20p,$$

$$\text{DScore} = 0.094\sqrt{n} + 0.60e - 0.324p,$$

где n – число точек сайтов, e – enclosure score, показатель «открытости» сайта для растворителя, p – hydrophilic score, усредненный потенциал гидрофильности, определяемый во время зондирования рецептора.

По сути, DScore является просто модифицированной версией SiteScore, параметризованной на большем количестве комплексов [2, 3].

Заинтересованный читатель также может ознакомиться с работами [5, 6], в которых SiteMap используется по прямому назначению.

Экспериментальная часть

5.1. Создание проекта

В своем личном каталоге создайте папку LW5. Дважды щелкните по ярлыку Maestro на рабочем столе. На верхней панели перейдите на вкладку **File > Change Working Directory**. В открывшемся окне укажите путь к рабочей директории LW5 в своей личной папке.

Затем перейдите **File > Save project as**, задайте проекту имя LW5 в графе **File name**, укажите путь к рабочей директории в графе **Location**.

5.2. Подготовка к поиску сайтов связывания

5.2.1. Подготовка структур

Выполните загрузку комплекса киназного домена Abl с иматинибом (PDB ID: 2HYU [7]). Структура является тетрамером, удалите все цепи вместе с содержащимися в них лигандами, кроме цепи A.

Важно: при загрузке структуры в окне **Get PDB** поставьте галочку у пункта **Diffraction data** напротив **Include**.

Выполним проверку «надежности» структуры с помощью модуля **Protein Reliability Report: Tasks > Browse > Structure Analysis > Protein Reliability Report**. В разделе **Analyze** выберите **Region around ligand**, нажмите **Pick Ligand** и выберите лиганд в **Workspace**. По умолчанию анализ проводится в области, удаленной от лиганда на 10 Å. Установите **Shell** как «30 Å». В разделе **Use diffraction data** нажмите **Browse** и укажите путь к файлу **2HYU.cv**, содержащему данные о значениях F_{obs} и прогнозируемых ошибках (Sigma). Измените имя задания на «2HYU_reliability_raw» и нажмите **Run**. Расчет займет не больше 5 минут. Новые записи будут добавлены в **Entry List** (рис. 5.2).

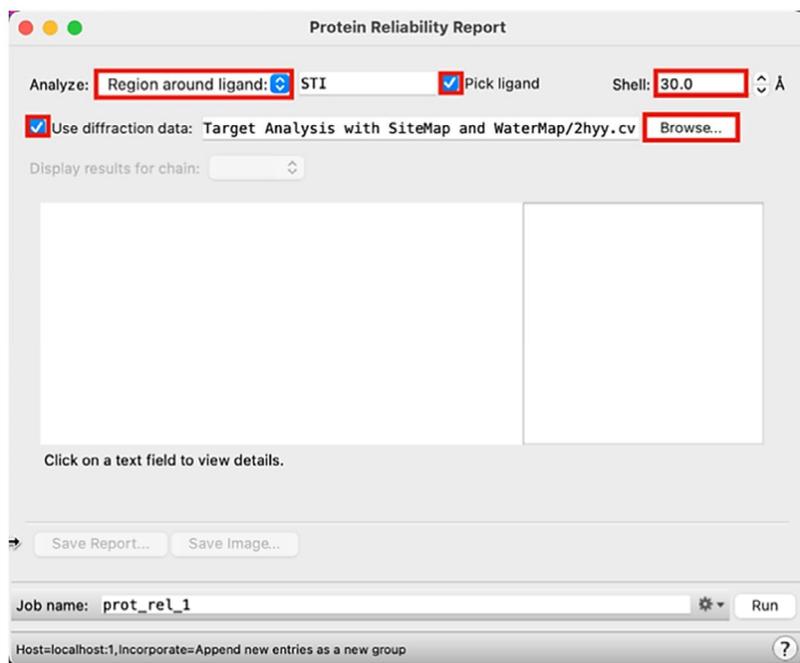


Рис. 5.2. Запуск задания Protein Reliability Report

В появившемся баннере выберите **Reliability Report** и дождитесь открытия окна, затем в разделе **Display results for chain** выберите цепь **A**. Оцените проблемы, присутствующие в структуре (рис. 5.3).



Рис. 5.3. Окно Protein Reliability Report

Теперь выполните подготовку белка с помощью **Protein Preparation Workflow**. Поставьте галочку напротив пункта **Fill in missing side chains using Prime** в разделе **Preprocess**. Удалите кристаллографические артефакты, если они есть, **но не воду**. Выполните **HBond Assignment** и минимизацию (уберите все галочки в разделе **Delete Waters**).

Вернитесь к панели **Protein Reliability Report**. Включите подготовленный комплекс в **Workspace**, укажите лиганд, путь к файлу с дифракционными данными и установите **Shell** так, как делали это ранее.

Измените имя работы на «**2HYY_reliability_prepared**» и нажмите **Run**. Расчет займет не более 5 минут (рис. 5.4).

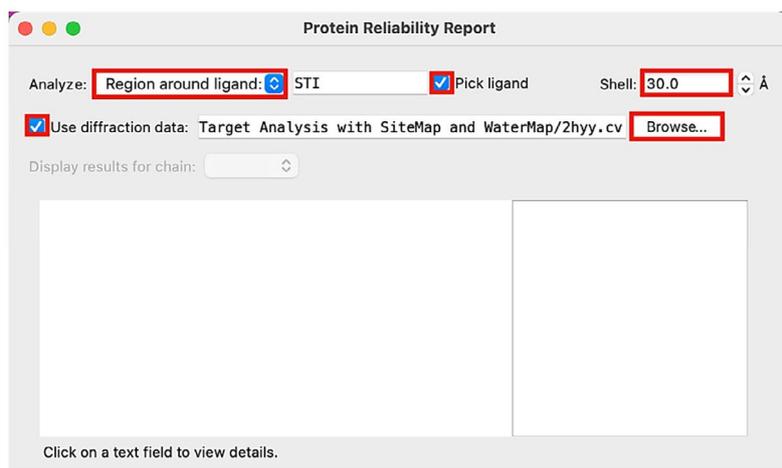


Рис. 5.4. Запуск задания Protein Reliability Report

Вновь оцените результаты с помощью диаграммы **Protein Reliability Report**. Обратите внимание на то, что большая часть проблем была устранена, однако значение **Binding Site RSCC** (*англ.* Real Space Correlation Coefficient)¹ все еще не находится в пределах нормы. Нажав на соответствующий кружок, в правой панели вы сможете увидеть проблемные остатки.

Отразите в отчете: снимки экранов, на которых содержатся диаграммы **Protein Reliability Report** до и после подготовки.

5.3. Анализ SiteMap

Загрузите файл рецептора в апо-форме **BCRABL.mae**, затем переименуйте запись белка как «**2HYY - minimized_apo**» и включите ее в **Workspace**.

Откройте модуль **SiteMap: Tasks > Browse > Structure Analysis > Binding Site Detection**. В разделе **Task** выберите **Identify top-ranked potential receptor binding site**. Остальные настройки оставьте без изменений. Назовите задание как «**sitemap-2HYY-minimized-apo**» и нажмите **Run**. Расчет займет около 10 минут (рис. 5.5).

В случае успешного завершения расчета в **Entry List** будет добавлена новая группа, каждая запись в которой соответствует идентифицированным сайтам и содержит точки, данные о поверхностях и областях. Сайты автоматически ранжируются от лучшего к худшему. Зафиксируйте в **Workspace** структуру комплекса и включите запись **sitemap_2HYY-minimized-apo_site_1** (рис. 5.6).

¹ RSCC — это мера сходства между картой электронной плотности, рассчитанной непосредственно из структурной модели, и картой, рассчитанной на основе экспериментальных данных.

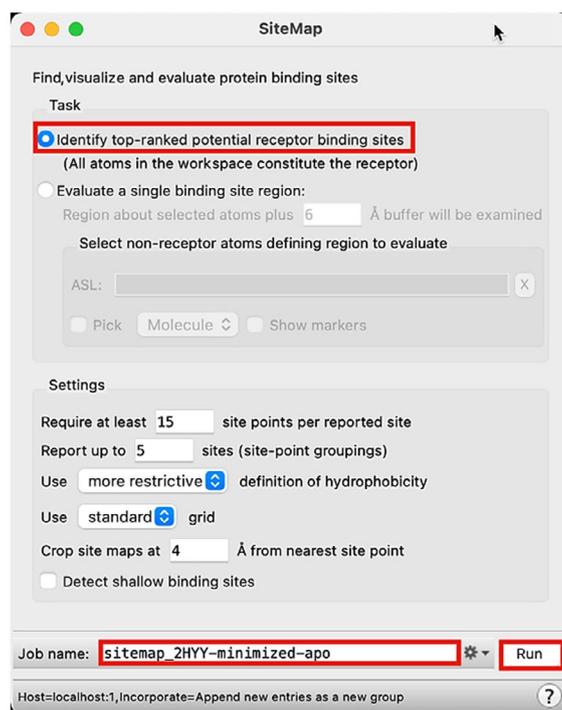


Рис. 5.5. Запуск задания Site Map

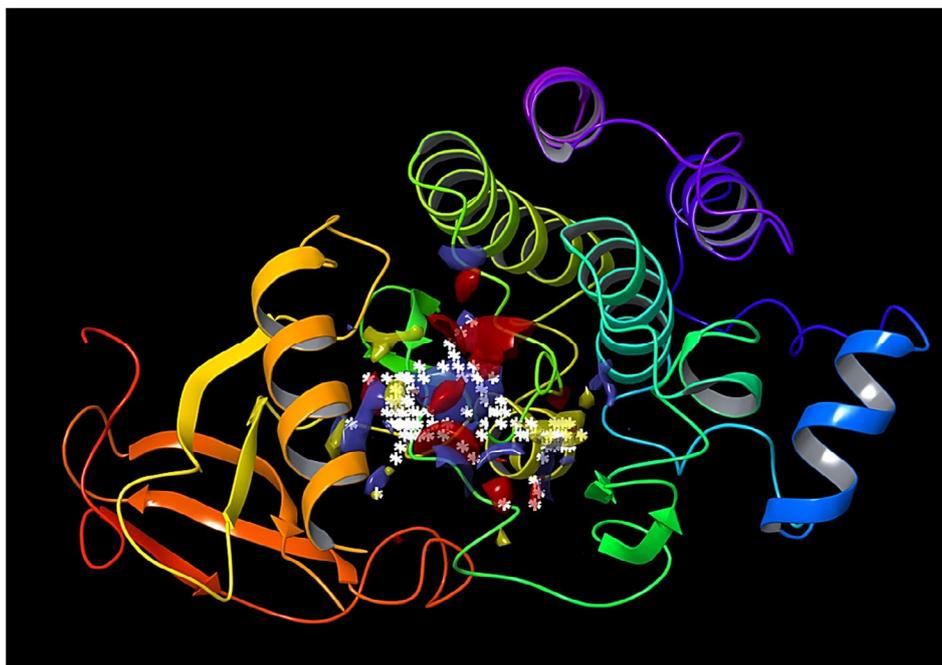


Рис. 5.6. Наложение site1 на структуру подготовленного комплекса

Отразите в отчете:

- 1) наложение всех найденных сайтов на структуру рецептора. Для этого необходимо будет прикрепить в отчет несколько снимков **Workspace**. Выделите запись комплекса, с помощью панели **Style** настройте отображение **Ribbons**, а также молекулярной поверхности (**Surface**). Скройте все атомы белка. Включите все найденные **SiteMap** сайты вместе со структурой белка. Подпишите сайты на рисунке;

- 2) таблицу, содержащую информацию о **SiteScore**, **DScore**, **Site Volume** для каждого из сайтов. Удастся ли **SiteMap** ранжировать на первое место сайт связывания иматиниба? Для этого сравните структуру подготовленного вами комплекса с результатами анализа SiteMap для апо-формы (возможно потребуется выровнять белки).

Применим режим исследования конкретного сайта. Выберите в **Entry List** структуру подготовленного комплекса. В панели **SiteMap** выберите **Evaluate a single binding site region**. Нажмите в **Workspace** на любой из атомов лиганда для определения области сайта. Измените название работы на «**sitemap_2HYY-minimized-holo**», нажмите **Run**. Расчет займет 3–5 минут.

Отразите в отчете: таблицу, содержащую результаты анализа **SiteMap** для апо- и холо-структур, а также наложение карт сайтов. В чем сходства и различия сайтов в двух структурах (табл. 5.1)?

Таблица 5.1

Пример таблицы для сравнения результатов SiteMap

sitemap_2HYY-minimized-holo			sitemap_2HYY-minimized-apo		
Site Vol, Å ³	SiteScore	DScore	Site Vol, Å ³	SiteScore	DScore

Вопросы к защите

1. Дайте определение понятию «сайт связывания».
2. Перечислите основные положения трех моделей, описывающих лиганд-белковые взаимодействия.
3. Опишите основные положения теории аллостерической модуляции.
4. Можно ли использовать SiteMap для обнаружения аллостерических сайтов связывания?
5. Какие зонды могут использоваться для построения полей молекулярного взаимодействия?
6. Какую оценку — SiteScore или DScore — стоит в первую очередь использовать для оценки сайтов?
7. Какие типы полей молекулярного взаимодействия рассчитывает SiteMap?
8. На каком этапе разработки лекарства может использоваться SiteMap?

Рекомендуемая литература

1. Schrödinger Release 2023-4: SiteMap. — New York : Schrödinger, 2023.
2. *Halgren, T. A.* Identifying and Characterizing Binding Sites and Assessing Druggability // Journal of Chemical Informatics and Modeling. — 2009. — Vol. 49, № 2. — P. 377–389.
3. *Halgren, T.* New Method for Fast and Accurate Binding-site Identification and Analysis // Chemical Biology & Drug Design. — 2007. — Vol. 69, № 2. — P. 146–148.
4. Discovery of 7-(3-(piperazin-1-yl)phenyl)pyrrolo[2,1-f][1,2,4]triazin-4-amine derivatives as highly potent and selective PI₃Kδ delta inhibitors / L.-Y. Qin, Z. Ruan,

- R. J. Cherney [et al.] // *Bioorganic and Medicinal Chemistry Letters*. – 2017. – Vol. 27, № 4. – P. 855–861.
5. Structure-based assessment and druggability classification of protein–protein interaction sites / L. Alzyoud, R. A. Bryce, M. Al. Sorkhy [et al.] // *Scientific Reports*. – 2022. – Vol. 12. – Article number 7975.
 6. A Computational Approach to Enzyme Design: Predicting ω -Aminotransferase Catalytic Activity Using Docking and MM-GBSA Scoring / S. Sirin, R. Kumar, C. Martinez [et al.] // *Journal of Chemical Informatics and Modeling*. – 2014. – Vol. 54, № 8. – P. 2334–2346.
 7. Structural Biology Contributions to the Discovery of Drugs to Treat Chronic Myelogenous Leukaemia / S. W. Cowan-Jacob, G. Fendrich, A. Floersheimer [et al.] // *From Molecules to Medicines : Structure of Biological Macromolecules and Its Relevance in Combating New Diseases and Bioterrorism* / Edited by J. L. Sussman, P. Spadon. – Dordrecht, 2009. – P. 37–61. – (NATO Science for Peace and Security. Series A. Chemistry and Biology).



Лабораторная работа 6

СОЗДАНИЕ И ФИЛЬТРАЦИЯ ВИРТУАЛЬНЫХ БИБЛИОТЕК СОЕДИНЕНИЙ

Цель работы: освоение основных подходов к созданию и подготовке виртуальных библиотек соединений.

Вам предстоит:

- 1) создать виртуальную библиотеку с помощью инструмента перечисления заместителей R-Group Enumeration;
- 2) выполнить расчет физико-химических свойств, а также молекулярный докинг соединений набора с использованием алгоритмов QikProp и Glide соответственно;
- 3) осуществить аннотирование структур и свойств соединений, а затем провести фильтрацию библиотеки, применив инструмент Ligand Filtering.

Теоретическое введение

Начать теоретическое введение к данной работе следует с определений двух часто используемых в медицинской химии терминов – «виртуальная библиотека» и «химическое пространство».

Виртуальной библиотекой называют набор любых химических соединений в виде электронной базы. Структуры в ней могут быть представлены как в виде линейной нотации (.smi), так и в 2D- (.sdf) и 3D-форме (например, .maegz для файлов Maestro, .xuz и т. д.). Такие наборы впоследствии могут быть использованы в ходе виртуального скрининга, анализа QSAR, моделирования фармакофоров и т. д. [1, 2].

Сколько же соединений может быть в такой библиотеке? На момент написания данного лабораторного практикума электронная база данных Химической реферативной службы (*англ.* CAS, Chemical Abstract Service) представляет собой библиотеку, содержащую примерно 204 миллиона известных органических веществ, сплавов, координационных соединений, минералов, смесей, полимеров и солей [3]. Увеличение скорости компьютерных вычислений и емкости памяти уже сейчас позволяет исследователям создавать и подвергать скринингу виртуальные наборы, **содержащие сотни миллионов структур**. Хотя эти цифры кажутся большими, они составляют лишь небольшую часть от тех органических соединений, которые потенциально могут быть синтезированы [4]. Например, Реймонд сообщил, что путем систематического вычислительного перечисления синтетически доступных и стабильных молекул, содержащих до 17 атомов С, S, N, O и галогенов, его группе удалось создать базу GDB-17 (*англ.* Generated database), включающую 166.4 миллиарда структур [5, 6]. Общее же число молекул, удовлетворяющих правилам Липински [7], оценивается в $\sim 10^{60}$ молекул [8]. Это больше, чем теоретически подсчитанное число звезд во Вселенной, и вряд ли человечеству когда-нибудь удастся исследовать каждую из этих структур, даже ограничившись лишь методами молекулярного моделирования. Говоря об этом гипотетическом наборе соединений, обычно употребляют выражение «химическое пространство лекарственныхподобных соединений».

В самом широком смысле под термином «химическое пространство» понимают «дискретный набор возможных молекул химических соединений» [9]. Однако более точное определение будет звучать иначе: «химическое пространство – это множество химических объектов, между которыми определены отношения сходства» [10]. Два объекта, находящиеся рядом друг с другом в химическом пространстве, будут *похожи* (по физико-химическим свойствам, структуре или проявляемой активности – в зависимости от подхода к оценке подобия).

Некоторые большие коллекции органических соединений лекарственныхподобных молекул представлены в табл. 6.1. Часть из них является коммерческими библиотеками – в случае необходимости исследователи могут приобрести те структуры, которые представляют для них интерес.

Таблица 6.1

Перечень некоторых баз данных

База данных	Примечание	Размер	Ссылка
PubChem	Синтезированные молекулы из большого числа различных источников. Открытый доступ	$116 \cdot 10^6$	11
ChemSpider	База данных Королевского химического общества. Открытый доступ	$129 \cdot 10^6$	12

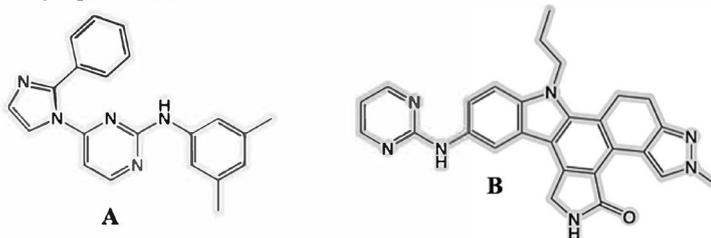
База данных	Примечание	Размер	Ссылка
ZINC	Открытое для скачивания химическое пространство соединений, готовых к стыковке. Большая часть молекул коммерчески доступна	$1.4 \cdot 10^9$	13
BindingDB	Открытая база данных, содержащая данные об аффинности/активности биологически активных соединений	$2.8 \cdot 10^6$	14
ChEMBL	Открытая база данных известных соединений с аннотированием их мишеней, биологической активности, физико-химических свойств и т. д.	$2.4 \cdot 10^6$	15
DrugBank	Аннотированная база данных одобренных лекарств с клиническими/ФК/ФД и др. данными. Открытый доступ	$16.6 \cdot 10^3$	16
ChemDiv	Каталог коммерческих библиотек, часть которых содержит аннотированные скрининговые данные; база доступна для скачивания	$1.6 \cdot 10^6$	17
Enamine (Real Space)	Коммерчески доступное химическое пространство. Соединения в нем генерируются <i>on-the-fly</i> во время поиска, но доступны синтетически по требованию. Поиск осуществляется через ПО BioSolveIT; Enamine предоставляет доступ также к другим коммерчески доступным наборам соединений	$38 \cdot 10^9$	18

Оценка сходства/несходства двух объектов химического пространства / библиотеки – субъективная мера хотя бы потому, что она зависит от типа молекулярного представления: будь то физико-химические свойства, отпечатки пальцев, графы, фармакофоры или молекулярные поверхности (рис. 6.1).

Подобие физико-химических свойств

	ММ, Да	LogP	Число вращающихся связей	Число Аг-колец	Число «тяжелых» атомов
A	341.4	5.23	4	4	26
B	463.5	4.43	4	5	33

2D-молекулярное подобие



3D-молекулярное подобие

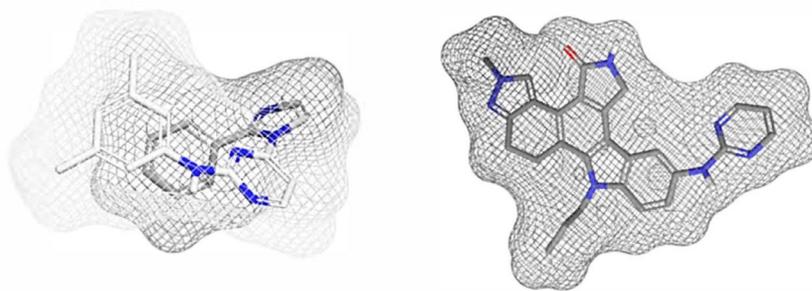
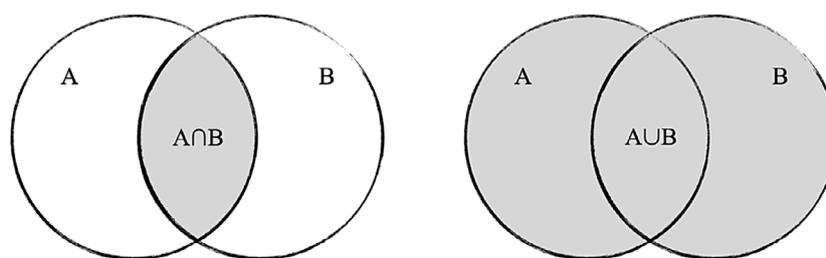


Рис. 6.1. Различные варианты представления молекул и, соответственно, подходов к оценке подобия

Существует большое число различных метрик подобия. Самой известной из них, вероятнее всего, является коэффициент Танимото [19], отражающий отношение числа одинаковых элементов двух множеств к общему числу их элементов.

В случае графов, например, числителем в выражении для коэффициента является максимальная общая подструктура, а в случае фингерпринтов – одинаковые элементы двух векторов. С точки зрения теории множеств, определение коэффициента Танимото представлено на рис. 6.2. Последнее выражение указывает на то, что данный индекс подобия является симметричной метрикой [20, 21].



$$S_{Tan} = \frac{n(A \cap B)}{n(A) + n(B) - n(A \cap B)} = \frac{n(A \cap B)}{n(A \cup B)}$$

$A \cap B$

Пересечение множеств

$A \cup B$

Объединение множеств

$$S_{Tan}(A, B) = S_{Tan}(B, A)$$

Рис. 6.2. Графическое представление и математическое выражение коэффициента подобия Танимото

Некоторые другие метрики подобия/несходства, а также математические выражения, соответствующие векторным представлениям химических объектов, представлены в табл. 6.2.

Таблица 6.2

Метрики сходства/несходства

Метрика	Математическое выражение*	Ссылка
S_{Dice} Коэффициент Дайса	$S_{Dice} = \frac{2a}{2a + b + c}$	22
S_{Cos} Коэффициент косинусов	$S_{Cos} = \frac{a}{\sqrt{(a + b)(a + c)}}$	23
D_{Euc} Эвклидово расстояние Метрика несходства	$D_{Euc}(a, b) = \sqrt{\sum_i^n (b_i - a_i)^2}$	24
D_{Ham} Расстояние Хэмминга Метрика несходства	$D_{Ham}(a, b) = a + b - 2c$	25

* a и b – число элементов (битов) бинарных векторов A и B ; c – число общих элементов (битов) векторов; b_i и a_i – значения i -го дескриптора в вещественных векторах B и A соответственно.

Разумеется, это не вся номенклатура метрик подобия. С некоторыми другими из них, в частности с асимметричными индексами, вы можете познакомиться в соответствующей литературе [26].

Таким образом, немаловажным свойством, характеризующим библиотеки, является их разнообразие, т. е. мера структурного сходства/несходства ее компонентов между собой. Структурно-разнообразные наборы, охватывающие широкие области химического пространства, называются *диверсифицированными библиотеками*, и обычно используются при первичном скрининге с целью идентификации хитов. Библиотеки, содержащие похожие соединения, например имеющие один и тот же скаффолд или нацеленные на связывание с одной и той же мишенью, являются *сфокусированными* [10].

Любое соединение в библиотеке может быть описано комбинацией таких элементов, как скаффолд (*англ.* template, scaffold – остов, основа, ядро) – общий для соединений набора структурный мотив, служащий фундаментом для крепления всех остальных фрагментов; строительный блок (*англ.* building block) – переменная часть будущей молекулы, а также линкер, соединяющий их вместе (может отсутствовать). Точки крепления на скаффолде, линкерах и строительных блоках, как правило, маркируются (удобно представлять в виде структур Маркуша) (рис. 6.3).

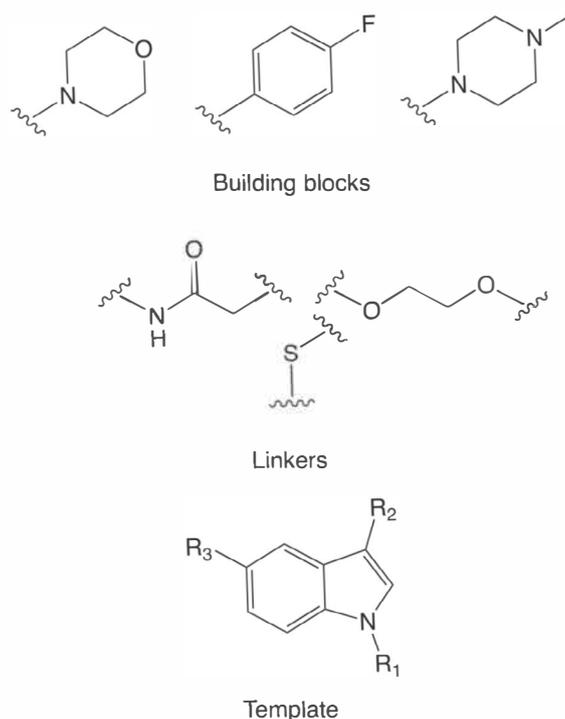


Рис. 6.3. Типы фрагментов, из которых могут состоять библиотеки соединений

Подходы к конструированию виртуальных библиотек многочисленны и имеются в том числе методы, позволяющие создать набор структур на основе сайта биомишени [27], но точно можно сказать, что библиотека не создается бездумным перечислением всех возможных комбинаций описанных выше фрагментов. В частности, могут отбрасываться соединения, которые не соответствуют правилам Липински; также могут отбираться только те молекулы, которые смогут обеспечить заданный уровень разнообразия. В случае если соединение не обладает желаемым значением того или иного физико-химического свойства, можно прибегнуть к биоизостерной замене [10].

Помимо перечисления и отбора комбинаций фрагментов на основе маркировки точек конъюгации также можно применять подход реакционных трансформаций (*англ.* reaction-based enumeration), в ходе которого целевые молекулы библиотеки представляют собой возможные продукты определенных реакций, для которых определяются наборы возможных реагентов, содержащие переменные фрагменты [28].

На рис. 6.4 представлен пример составления сфокусированной библиотеки.

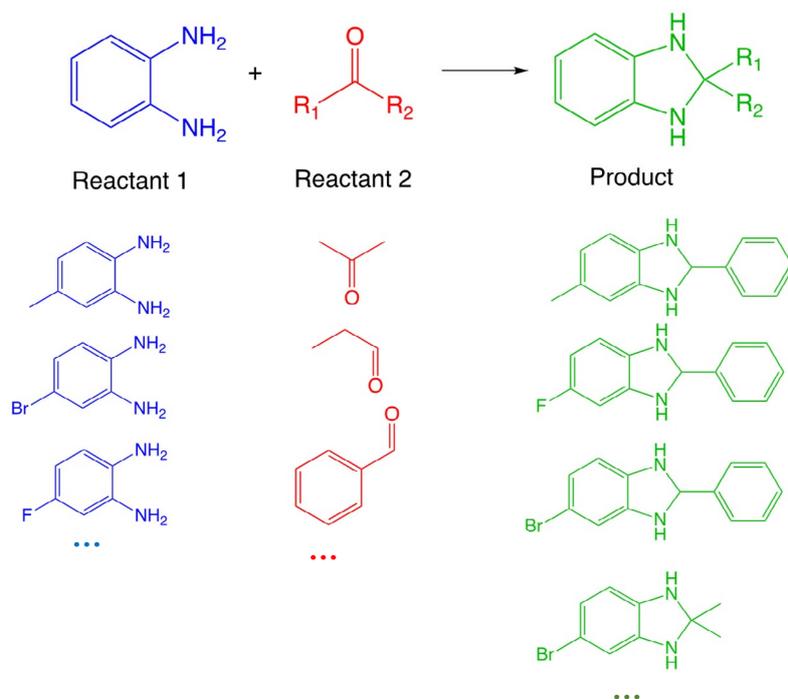


Рис. 6.4. Принцип составления библиотек на основе reaction-based enumeration

Созданный набор должен быть подвергнут фильтрации для исключения соединений, не удовлетворяющих определенным правилам (например, правилам Липински), нежелательных структур, *структурных алертов* (*англ.* structural alerts), а также тех соединений, которые, как предполагается, не будут обладать должной активностью. Для решения последней задачи часто используют методы молекулярного докинга, анализа комплементарности формы между лигандом и сайтом мишени или фармакофорные подходы [29].

Под *структурными алертами* понимаются фрагменты и мотивы соединений библиотеки, которые могут обуславливать нежелательную активность, токсичность или же приводить к неверной интерпретации результатов биологических анализов (PAINS, *англ.* Pan-assay interference compounds [30]). Удобно проводить отсеивание таких нежелательных лигандов на основе поиска подструктур с помощью нотаций SMARTS. Некоторые примеры *структурных алертов* представлены на рис. 6.5.

На финальном этапе из полученного массива структур осуществляется отбор соединений в библиотеку. При этом важно соблюсти требуемый уровень структурного разнообразия, даже для сфокусированных наборов. Существует довольно много подходов к тому, как отбирать соединения — с использованием кластерного анализа, метрик разнообразия или разделения химического пространства.

В первом случае между всеми соединениями (1) рассчитываются значения попарных расстояний, (2) алгоритм кластеризации разбивает структуры на определенные группы (кластеры), а (3) отбор нескольких соединений из каждого кластера в библиотеку завершает процесс (рис. 6.6).

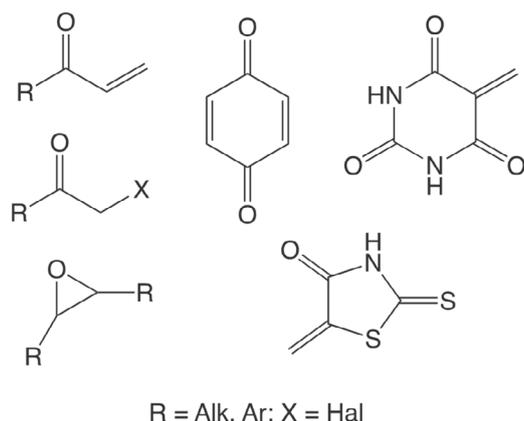


Рис. 6.5. Структурные алерты и PAINS

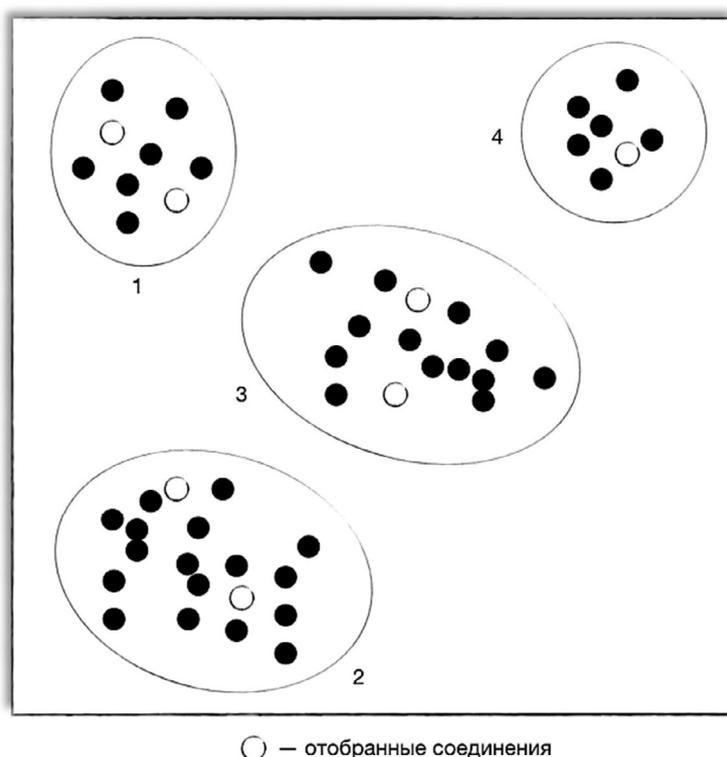


Рис. 6.6. Некоторое химическое пространство, соединения в котором разбиты на 4 кластера

Как можно понять, методы, использующие кластеризацию, подразумевают первоначальное разбиение исходного набора на подгруппы. Алгоритмы, основанные на разнообразии (DBCS, *англ.* Dissimilarity-based compound selection), не включают этот шаг и производят непосредственный отбор соединений. Типичный рабочий процесс в этом случае заключается в (1) выборе первой структуры и включении ее в библиотеку, (2) расчете расстояния (или другой метрики различия) между выбранной структурой и соединениями исходного набора, (3) отборе соединения, наиболее непохожего на первое [31].

Двумя самыми распространенными методами отбора являются MaxSum (выбирает соединение i с максимальной суммой расстояний D до всех уже отобранных m соединений) и MaxMin (выбирает структуру i с максимальным расстоянием до ближайшего соседа среди m соединений) [32] (рис. 6.7).

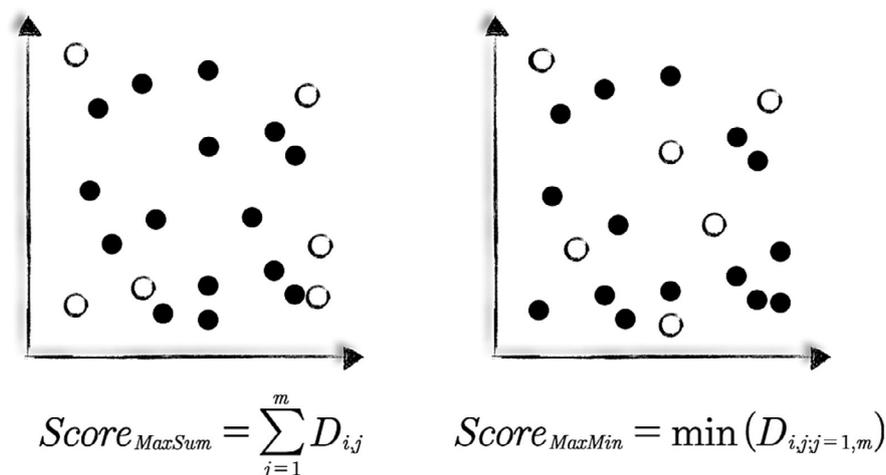


Рис. 6.7. Отбор соединений с помощью подходов DBCS

Таким образом, MaxMin направлен на максимизацию минимального расстояния между любой парой различных элементов библиотеки [33], обеспечивая равномерный отбор молекул из химического пространства, в отличие от MaxSum, в случае использования которого в библиотеку неизбежно будут попадать похожие соединения. С еще одним подходом – алгоритмом исключения сферы (*англ.* sphere exclusion algorithm), вы можете ознакомиться в [10, 32].

Основой подходов разбиения (*англ.* cell-based или partitioning-based methods) является разделение химического пространства² на несколько областей (ячеек, *англ.* cell) и отбор из каждой одного или нескольких соединений. Ось каждого из свойств делится на серию диапазонов значений, чтобы дать серию «корзин» (*англ.* bins). Стоит отметить, что «корзины» не должны быть равны по размеру, и что некоторые свойства могут (по крайней мере, теоретически) охватывать бесконечный диапазон (рис. 6.8).

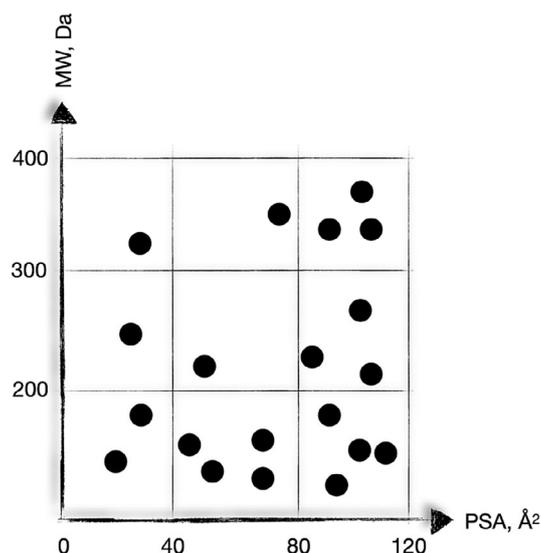


Рис. 6.8. Разбиение химического пространства на 9 ячеек путем разделения каждого свойства на три диапазона значений (на три «корзины»)

² В данном случае речь идет обычно о низкоммерном пространстве дескрипторов.

Очевидно, что с точки зрения влияния повышения диверсифицированности создаваемой библиотеки **необходимо стремиться** не к равномерному заполнению ячеек, а к **заполнению наибольшего числа ячеек**. Оценить равномерность заполнения можно с помощью такого параметра, как «энтропия» S :

$$S = -\sum_{i=1}^{N_{cells}} \frac{N_i}{N_{sel}} \ln\left(\frac{N_i}{N_{sel}}\right),$$

где N_i – количество молекул в ячейке i , N_{sel} – количество отобранных молекул, а N_{cells} – общее количество ячеек в химическом пространстве [32].

Вопрос оптимизации отбора молекул из химического пространства выходит за рамки теоретического введения к данной работе, однако вы можете обратиться к материалам лекций, а также к [2, 10, 32] для более детального ознакомления.

В этой работе вам будет предложено самостоятельно создать библиотеки с помощью перечисления строительных блоков, а затем провести фильтрацию и обогащение.

Экспериментальная часть

6.1. Создание проекта

В своем личном каталоге создайте папку LW6. Дважды щелкните по ярлыку Maestro на рабочем столе. На верхней панели перейдите на вкладку **File > Change Working Directory**. В открывшемся окне укажите путь к рабочей директории LW6 в своей личной папке.

Затем перейдите **File > Save project as**, задайте проекту имя LW6 в графе **File name**, укажите путь к рабочей директории в графе **Location**.

6.2. Создание библиотеки с помощью R-Group Enumeration

6.2.1. Создание и подготовка библиотеки

Импортируйте файл **Hit.maegz** и включите структуру в **Workspace**. Включив метки атомов (**Atom Labels>Element+Atom Number**), выделите атом водорода **H21** в пуриновом ядре. Выбранный атом будет являться центром прикрепления всех заместителей (рис. 6.9).

Откройте модуль создания библиотек **R-Group Enumeration: Tasks > Browse > Enumeration and Ideation > R-Group Enumeration**.

Заметка: R-Group Enumeration позволяет сгенерировать скрининговый набор путем систематического перечисления продуктов замещения выбранных пользователем атомов в целевом соединении на различные фрагменты. В этой работе мы будем использовать встроенную в Schrödinger библиотеку **Diverse_R-groups**, однако вы всегда вольны создать собственную базу заместителей с помощью **R-group Creator**.

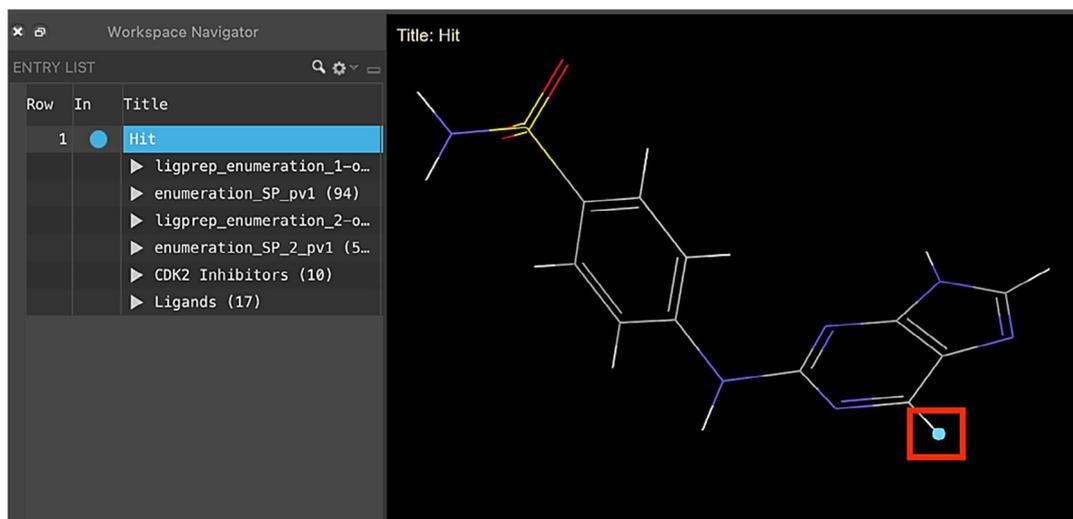


Рис. 6.9. Выделение точки крепления заместителей

В открывшемся окне выберите **Diverse R-groups (43)** в столбце **R-group Library**. Убедитесь, что отмеченный атом **H21** отображается в качестве **R1** (позиция 1). Измените название задачи на «**R-Group Enumeration**» и нажмите **Run**. Расчет обычно занимает не более 3 минут. Закройте панель **Custom R-Group Enumeration** (рис. 6.10).

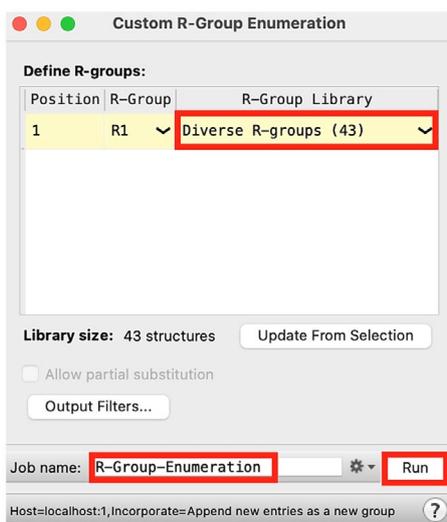


Рис. 6.10. Запуск задания

По окончании расчета выделите всю группу в **Workspace** и выполните подготовку лигандов с помощью **LigPrep** (установите $\text{pH}=7.0\pm 2$ и поставьте галочку напротив пункта **Retain specified chiralities**). Назовите работу «**ligprep_enumeration**». Расчет займет не более 6 минут.

6.2.2. Докинг библиотеки

Загрузите и подготовьте структуру комплекса CDK2/СусА (PDB ID: 5NEV [34]) с помощью **Protein Preparation Workflow** в соответствии со стандартным протоколом, описанном в лабораторной работе 1. Удалите лишние ионы, гетероатомы и молекулы, осуществите минимизацию и удаление всей воды. Обратите внимание, что данная

кристаллическая структура представлена в виде димера. Удалите цепи C и D. Суммарно расчет займет примерно 15 минут.

Рассчитайте сетку потенциалов, определив область стыковки по кристаллическому лиганду и не устанавливая ограничений на взаимодействия с остатками сайта. Назовите задание «**glide-grid_enumeration**». Расчет займет 5–6 минут.

Выполните стыковку подготовленных лигандов библиотеки с помощью **Ligand Docking**. В качестве файла грида выберите полученную вами ранее сетку **glide-grid_enumeration.zip**, а **SP flexible** – в качестве протокола точности. Назовите задание как «**enumeration_SP**». Расчет займет 10–15 минут.

6.2.3. Предсказание физико-химических свойств

Откройте модуль **QikProp: Tasks > Browse > ADME and Molecular Properties > Ligand-Based ADME/Tox Prediction**. Выделите группу **enumeration_SP_pv1** в **Entry List**. В разделе выбора источника структур для предсказания (**Use structures from**) выберите **Project Table (selected entries)**, поставьте галочку напротив пункта **Fast Mode**, назовите задание «**qikprop_enumeration_1**» и нажмите **Run**. Расчет займет несколько секунд (рис. 6.11).

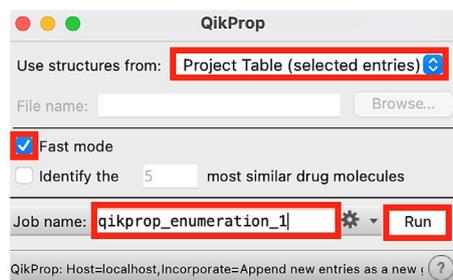


Рис. 6.11. Запуск расчета свойств QikProp

Заметка: **QikProp** позволяет рассчитать более 20 различных физико-химических/биологических дескрипторов, включая LogP , проницаемость Caco-2 и $\log(\text{IC}_{50})$ блокировки HERG K^+ -каналов. Программа использует алгоритм, имитирующий моделирование статистической механики Монте-Карло, и регрессионный анализ для предсказания значений свойств. **QikProp** был разработан William L. Jorgensen и коллегами [35–37].

Отразите в отчете:

- 1) аннотирование свойств полученной библиотеки. Для этого получите графики распределения основных физико-химических свойств с помощью модуля **Advanced Project Table: Task > Browse > Advanced Project Table Plotting**. В качестве типа графика выберите **Histogram**. Выделите группу, содержащую соединения библиотеки в **Entry List**. В открывшемся окне справа в разделе **Property** выберите **QLogPo/w**. Настройте цвет и ширину столбцов по своему усмотрению. Полученный график сохраните. Повторите процедуру также для таких свойств, как **PSA**, **acceptHB**, **donorHB**, **molMW**. Включите графики в отчет. Прокомментируйте, все ли соединения библиотеки являются частью химического пространства, определенного правилами Липински (рис. 6.12);

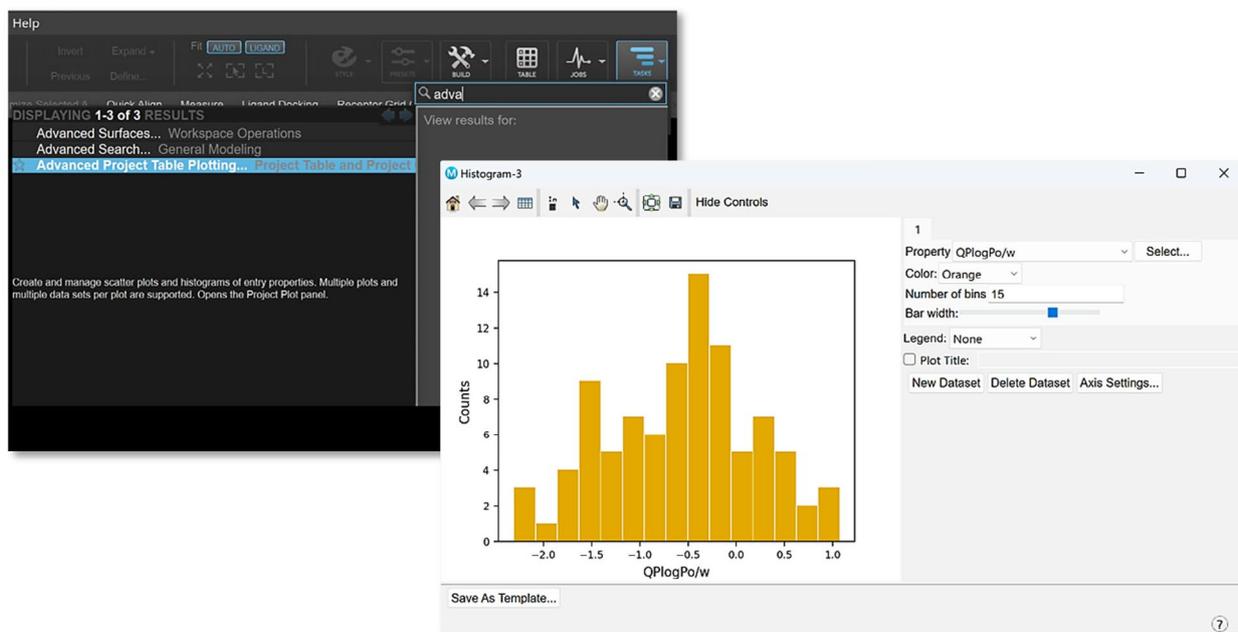


Рис. 6.12. Построение графиков распределения физико-химических свойств

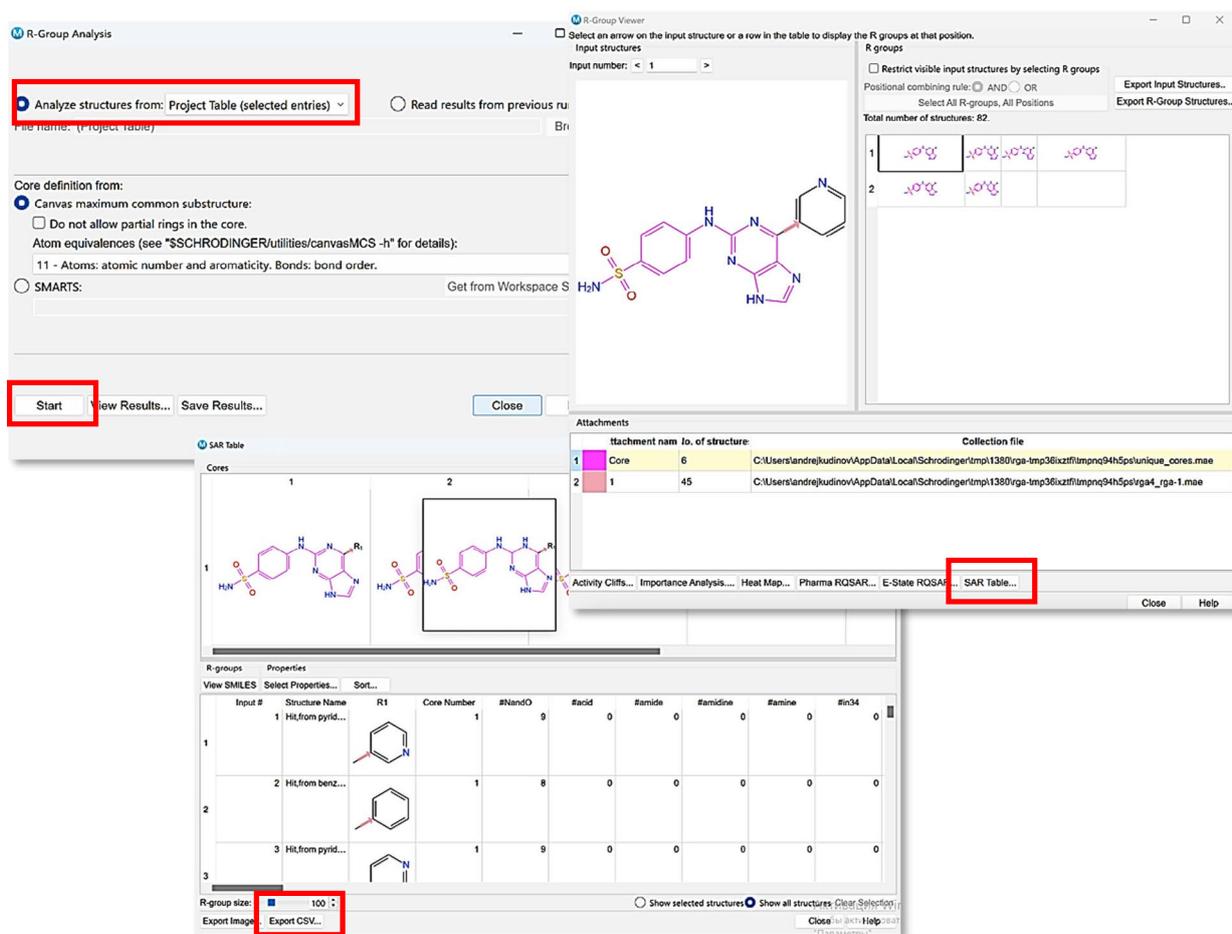


Рис. 6.13. Экспорт библиотеки в формате .csv

- 2) структуры Маркуша для соединений библиотеки. Для этого воспользуйтесь модулем **R-Group Analysis: Task > Browse > R-Group Analysis**. Выделите группу, содержащую соединения библиотеки в **Entry List**. В разделе **Analyze structure from** выберите **Project Table (selected entries)**. Оставьте настройки **Core definitions from** по умолчанию и нажмите **Start**. Расчет займет 3–4 минуты. Окно, содержащее результаты анализа, откроется автоматически. В нем, выбрав **Core 1** в качестве скаффолда, нажмите на кнопку **SAR Table** на нижней панели. Затем – **Export CSV**. Сохраните файл электронной таблицы, в которой будут содержаться нотации SMILES для всех заместителей. Используйте их для отрисовки структур в ChemDraw (рис. 6.13).

Как можно увидеть, библиотека требует фильтрации. Откройте модуль **Ligand Filtering: Tasks > Browse > Ligand Preparation and Library Design > Ligand Filtering**. Выберите в **Entry List** группу **qikprop_enumeration_1_out**, содержащую результаты расчета свойств. В разделе **Use structures from** укажите **Project Table (selected entries)**. В таблице **Available properties**, найдите и выберите свойство **glide_gscore(Impact)**. Затем в разделе **Property** установите значение критерия фильтрации **glide_gscore < -8.5** и нажмите **Add**. Повторите процедуры для свойств **RuleOfFive** и **PSA (RuleOfFive < 1 и PSA < 140)**. В разделе **Filtering definitions and criteria** должны отображаться следующие критерии фильтрации:

```
# Filter criteria
#
r_i_glide_gscore <-8.5
i_qp_RuleOfFive <1
r_qp_PSA <140
```

Проверьте, чтобы были выбраны пункты **All criteria are satisfied** напротив **Ligand matches if** и **Retain** напротив **Matching ligands**. Нажмите на значок шестеренки рядом с кнопкой **Run** (рис. 6.14).

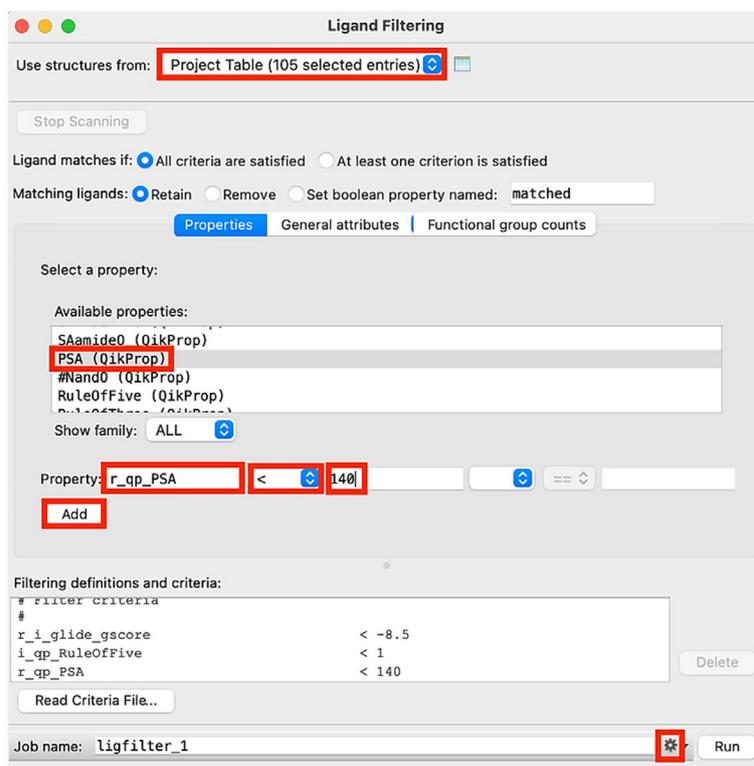


Рис. 6.14. Фильтрация лигандов

В открывшемся окне **Job Settings** выберите **Append new entries as a new group** в разделе **Incorporate**. Смените название задания на «**ligfilter_enumeration_1**». Нажмите **Run**. Расчет займет несколько секунд.

В результате будут отфильтрованы и импортированы только те структуры, которые показывают значение скоринга Glide SP лучше, чем -8.5 ккал/моль, соответствуют правилам Липински, а также имеют площадь полярной поверхности меньше 140 \AA^2 (рис. 6.15).

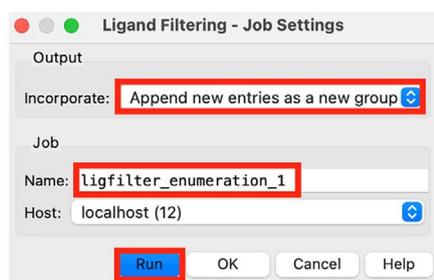


Рис. 6.15. Расширенные настройки задания

Отразите в отчете: аннотирование свойств библиотеки после фильтрации. Для этого получите графики распределения основных физико-химических свойств с помощью модуля **Advanced Project Table**.

Вопросы к защите

1. Дайте определение понятию «виртуальная библиотека».
2. Дайте определение понятию «токсофор».
3. Почему PAINS являются нежелательными структурами? Приведите несколько примеров PAINS, поясните механизм, опосредующий возникновение ложноположительных результатов.
4. Каким образом можно обеспечить должный уровень структурного разнообразия создаваемой библиотеки?
5. Какие подходы к созданию виртуальных библиотек существуют?
6. Какие требования предъявляются к строительным блокам, используемым для построения библиотек?
7. К какому типу фильтров относится фильтр, отсеивающий соединения, содержащие PAINS-фрагменты?
8. Как можно повысить уровень обогащения виртуальной библиотеки потенциально активными соединениями?
9. В чем заключаются достоинства метода реакционных трансформаций в контексте создания виртуальных библиотек?
10. Для чего в ходе создания виртуальных библиотек может применяться прием биоизостерной замены?

Рекомендуемая литература

1. Molecular Modeling: Basic Principles and Applications / H.-D. Holtje, W. Sippl, D. Rognan, G. Folkers. — 3rd, Revised and Expanded Ed. — Weinheim : Wiley-VCH, 2008. — X, 310 p. — ISBN 978-3-527-31568-0.
2. Chemical Library Design / Edited by J. Z. Zhou. — New York : Humana Press, 2016. — X, 362 p. — (Methods in Molecular Biology ; Vol. 685). — ISBN 978-1-4939-6154-2.

3. CAS. Chemical Abstract Service : A division of the American Chemical Society : [сайт]. – URL: www.cas.org/cas-data/cas-registry (дата обращения: 18.01.2025).
4. *Walters, W. P.* Virtual Chemical Libraries // *Journal of Medicinal Chemistry*. – 2019. – Vol. 62, № 3. – P. 1116–1124.
5. *Reymond, J.-L.* The Chemical Space Project // *Accounts of Chemical Research*. – 2015. – Vol. 48, № 3. – P. 722–730.
6. Reymond Research Group : [сайт] / University of Bern. – URL: gdb.unibe.ch/downloads/ (дата обращения: 18.01.2024).
7. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings / C. A. Lipinsky, F. Lombardo, B. W. Dominy, P. J. Feeney // *Advanced Drug Delivery Reviews*. – 1997. – Vol. 23, № 1–3. – P. 3–25.
8. *Bohacek, R. S.* The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective / R. S. Bohacek, C. McMartin, W. C. Guida // *Medicinal Research Reviews*. – 1996. – Vol. 16, № 1. – P. 3–50.
9. Глоссарий русскоязычных терминов в медицинской химии / О. Н. Зефирова, К. В. Балакин, М. Ю. Красавин [и др.] // *Известия Академии наук. Серия химическая*. – 2019. – № 12. – С. 2381–2395.
10. *Маджидов, Т. И.* Введение в хемоинформатику. Учебное пособие. Часть 6. Химическое пространство и виртуальный скрининг / Т. И. Маджидов, И. И. Баскин, А. А. Варнек ; Казанский федеральный университет. – Казань : Издательство Казанского университета, 2019. – 238 с. – ISBN 978-5-00130-174-5.
11. PubChem : [open chemistry database] / National Institutes of Health // National Library of Medicine : The National Center for Biotechnology Information : [сайт]. – URL: pubchem.ncbi.nlm.nih.gov (дата обращения: 18.01.2025).
12. ChemSpider : [сайт] / Royal Society of Chemistry. – URL: www.chemspider.com (дата обращения: 18.01.2025).
13. ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery / J. J. Irwin, K. G. Tang, J. Young [et al.] // *Journal of Chemical Information and Modeling*. – 2020. – Vol. 60, № 12. – P. 6065–6073.
14. BindingDB : public molecular recognition database. – URL: www.bindingdb.org/rwd/bind/index.jsp (дата обращения: 18.01.2025).
15. ChEMBL : database of bioactive molecules with drug-like properties. – URL: www.ebi.ac.uk/chembl/ (дата обращения: 18.01.2025).
16. DrugBank : clinical development intelligence platform. – URL: go.drugbank.com (дата обращения: 18.01.2025).
17. ChemDiv : [сайт]. – URL: chemdiv.com (дата обращения: 18.01.2025).
18. Enamine REALSpace // Enamine : [сайт]. – URL: enamine.net/compound-collections/real-compounds/real-space-navigator (дата обращения: 18.01.2024).
19. *Rogers, D. J.* A Computer Program for Classifying Plants / D. J. Rogers, T. T. Tanimoto // *Science*. – 1960. – Vol. 132, № 3434. – P. 1115–1118.
20. *Chemoinformatics and Computational Chemical Biology* / Edited by J. Bajorath. – New York : Humana Press, 2011. – X, 588 p. – (Methods in Molecular Biology ; Vol. 672). – ISBN 978-1-60761-838-6.
21. *Molecular Similarity in Medicinal Chemistry* / G. Maggiora, M. Vogt, D. Stumpfe, J. Bajorath // *Journal of Medicinal Chemistry*. – 2014. – Vol. 57, № 8. – P. 3186–3204.
22. *Sørensen, T.* A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons. – København : Ejnar Munksgaard, 1948. – 34 p. – (Biologiske Skrifter / Det Kongelige Danske Videnskabernes Selskabs ; Bd. 5, № 4).

23. *Ochiai, A.* Zoogeographical Studies on the Soleoid Fishes found in Japan and its Neighboring Regions-I // *Nippon Suisan Gakkaishi*. – 1957. – Vol. 22, № 9. – P. 522–525.
24. *Deza, M. M.* Encyclopedia of Distances / M. M. Deza, E. Deza. – 4th Ed. – Berlin [et al.] : Springer, 2016. – XXII, 756 p. – ISBN 978-3-662-52843-3.
25. *Hamming, R. W.* Error Detecting and Error Correcting Codes // *Bell System Technical Journal*. – 1950. – Vol. 29, № 2. – P. 147–160.
26. *Маджидов, Т. И.* Введение в хемоинформатику. Учебное пособие. Часть 2. Химические базы данных / Т. И. Маджидов, И. И. Баскин, А. А. Варнек ; Казанский федеральный университет. – Казань : Издательство Казанского университета, 2015. – 185 с. – ISBN 978-5-00019-429-4.
27. Structure-based library design and the discovery of a potent and selective mast cell β -tryptase inhibitor as an oral therapeutic agent / G. Liang, S. Aldous, G. Merri-man [et al.] // *Bioorganic & Medicinal Chemistry Letters*. – 2012. – Vol. 22, № 2. – P. 1049–1054.
28. *Saldivar-Gonzalez, F. I.* Chemoinformatics-based enumeration of chemical libraries: a tutorial / F. I. Saldivar-Gonzalez, C. S. Hueta-Gonzalez, J. L. Medina-Franco // *Journal of Cheminformatics*. – 2020. – № 12. – Article number 64.
29. Structural Alert/Reactive Metabolite Concept as Applied in Medicinal Chemistry to Mitigate the Risk of Idiosyncratic Drug Toxicity: A Perspective Based on the Critical Examination of Trends in the Top 200 Drugs Marketed in the United States / A. F. Stepan, D. P. Walker, J. Bauman [et al.] // *Chemical Research in Toxicology*. – 2011. – Vol. 24, № 9. – P. 1345–1410.
30. *Baell, J. B.* New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays / J. B. Baell, G. A. Holloway // *Journal of Medicinal Chemistry*. – 2010. – Vol. 53, № 7. – P. 2719–2740.
31. *Kennard, R. W.* Computer Aided Design of Experiments / R. W. Kennard, L. A. Stone // *Technometrics*. – 1969. – Vol. 11, № 1. – P. 137–148.
32. *Leach, A. R.* An Introduction to Chemoinformatics / A. R. Leach, V. J. Gillet. – Revised Ed. – Dordrecht : Springer, 2007. – XV, 255 p. – ISBN 978-1-4020-6290-2.
33. Fair Max–Min Diversity Maximization in Streaming and Sliding-Window Models / Y. Wang, F. Fabbri, M. Mathioudakis, J. Li // *Entropy*. – 2023. – Vol. 25, № 7. – Article number e25071066.
34. Cyclin-Dependent Kinase (CDK) Inhibitors: Structure–Activity Relationships and Insights into the CDK-2 Selectivity of 6-Substituted 2-Arylamino-purines / C. R. Coxon, E. Anscombe, S. J. Harnor [et al.] // *Journal of Medicinal Chemistry*. – 2017. – Vol. 60, № 5. – P. 1746–1767.
35. *Jorgensen, W. L.* Prediction of drug solubility from structure / W. L. Jorgensen, E. M. Duffy // *Advanced Drug Delivery Reviews*. – 2002. – Vol. 54, № 3. – P. 355–366.
36. *Jorgensen, W. L.* The Many Roles of Computation in Drug Discovery // *Science*. – 2004. – Vol. 303, № 5665. – P. 1813–1818.
37. *Jorgensen, W. L.* Prediction of Drug Solubility from Monte Carlo Simulations / W. L. Jorgensen, E. M. Duffy // *Bioorganic and Medicinal Chemistry Letters*. – 2000. – Vol. 10, № 11. – P. 1155–1158.
38. Schrödinger Release 2023-4: QikProp. – New York : Schrödinger, 2023.



Лабораторная работа 7

ОПТИМИЗАЦИЯ ВЗАИМОДЕЙСТВИЙ ЛИГАНДА С БЕЛКОМ

Цель работы: ознакомление с подходами к модификации соединений-хитов и генерации идей для SAR-оптимизации

Вам предстоит:

- 1) подготовить структуру комплекса ацетил-КоА карбоксилазы 2 с лигандом;
- 2) с помощью инструмента Ligand Designer определить направление для модификации структуры кристаллического лиганда;
- 3) сгенерировать библиотеку соединений на основе перечисления R-групп, а затем осуществить ее молекулярный докинг.

Теоретическое введение

Анализ способов связывания лиганда с рецептором играет одну из ключевых ролей при определении путей SAR-оптимизации. Визуально проинспектировав позу малой молекулы в активном сайте, вы можете обнаружить, например, незанятый гидрофобный карман или аминокислотный остаток, с которым можно образовать дополнительное взаимодействие.

Schrödinger в своих пакетах ПО предоставляет Ligand Designer [1] – замечательный инструмент, позволяющий генерировать различные идеи для оптимизации взаимодействий «лиганд – мишень» путем модификаций малой молекулы и сразу же подвергать их проверке.

После анализа позы Ligand Designer позволяет отобразить следующие свойства:

- пространства для роста: свободные области между молекулярными поверхностями лиганда и сайта рецептора;
- области растворителя;
- фармакофорные сайты рецептора, с которыми можно образовать взаимодействия: гидрофобные, ароматические, донорные и акцепторные признаки;
- стабильные и нестабильные молекулы воды (необходимо провести анализ WaterMap [2]) (рис. 7.1).

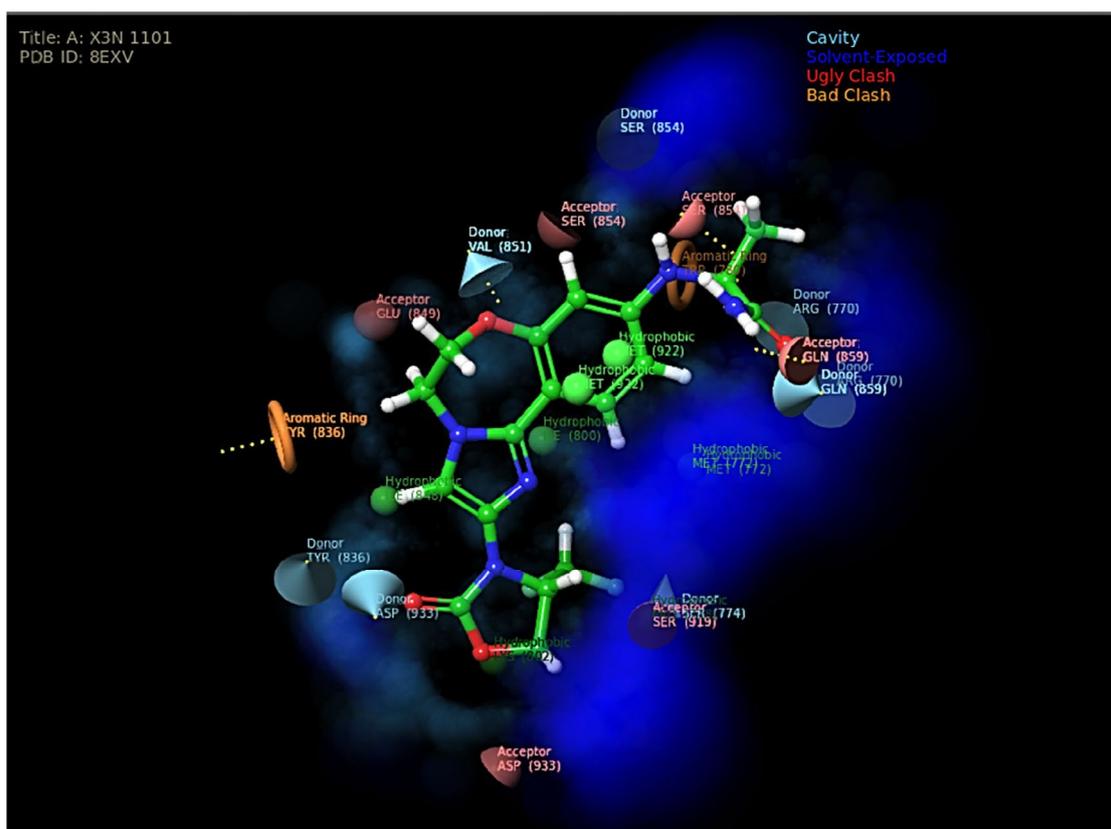


Рис. 7.1. Внешний вид Workspace в режиме анализа Ligand Designer: голубые области – пространства для роста; синие области – области растворителя; желтый пунктир – взаимодействия

Выбор пользователем какого-либо фармакофорного признака или области роста позволяет Ligand Designer проанализировать структуру лиганда и отметить сайты мо-

дификации, которые будут использоваться для автоматизированного создания библиотеки с помощью набора R-групп. Для каждого соединения такой библиотеки затем будут спрогнозированы позы связывания и останутся только те малые молекулы, которые наиболее успешно заполняют свободное пространство или образуют наиболее выгодные целевые взаимодействия.

Кроме того, данный модуль позволяет контролировать ключевые физико-химические свойства создаваемых соединений и проводить многопараметрическую оптимизацию (*англ.* MPO, Multi-Parameter Optimization) (рис. 7.2).

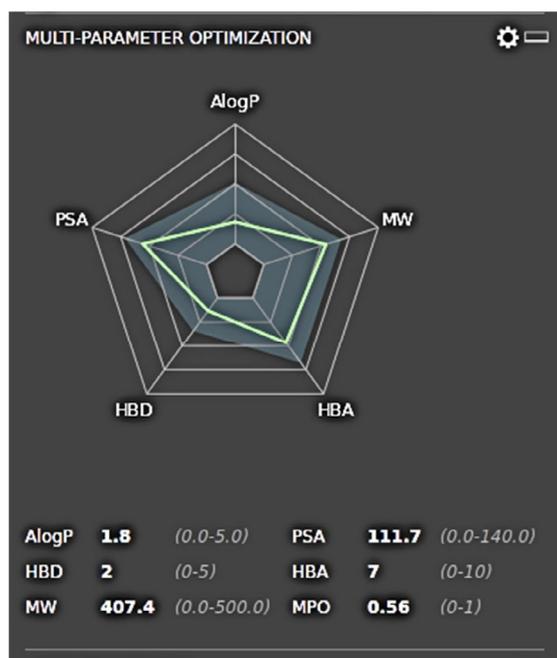


Рис. 7.2. Панель MPO

Как известно, коммерчески успешный, эффективный и безопасный препарат характеризуется некоторым балансом параметров, включая аффинность, ADME, профиль безопасности. Достижение этого баланса, часто противоречивого, является серьезной проблемой. Таким образом, процесс открытия лекарств требует решения задач одновременной оптимизации множества свойств [3]. Подходы MPO, решающие данную проблему, основаны на объединении множества физико-химических, топологических и других параметров малой молекулы в единую оценку MPO Score.

Путь к расчету финальной оценки начинается с нормализации значений каждого из свойств относительно диапазона от 0 до 1 с помощью логистической кривой (также называемой *функцией Ферми* [3] или *функцией желанности*, *англ.* desirability function [4]):

$$D = \frac{1}{1 + e^x},$$

где D и x — нормализованное и исходное значение свойства соответственно.

В простейшем случае логистическая кривая представляет собой сигмоиду, точки перегиба которой задаются двумя порогами (например, a и b), разделяющими значения свойства на три диапазона — **желаемый** (*англ.* good или desirable), в котором $D = 1$, **пороговый** (*англ.* marginal), в котором D монотонно убывает, и **нежелательный**, в котором $D = 0$ (рис. 7.3).

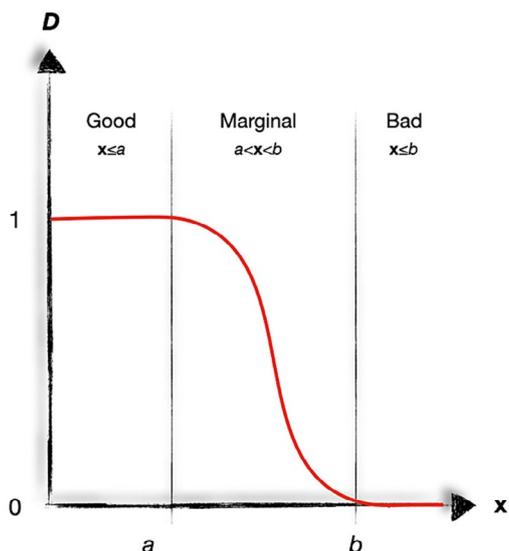


Рис. 7.3. Логистическая кривая с двумя точками перегиба

В варианте кривой, показанной на рис. 7.3, низкие значения свойства $x \leq a$ являются желаемыми, в таком случае показатель степени экспоненты удобнее представить в виде $a(x - b)$. В противном же случае – в виде $a(b - x)$. Если предпочтительным является диапазон значений, лежащий посередине, форма кривой должна задаваться четырьмя порогами (рис. 7.4).

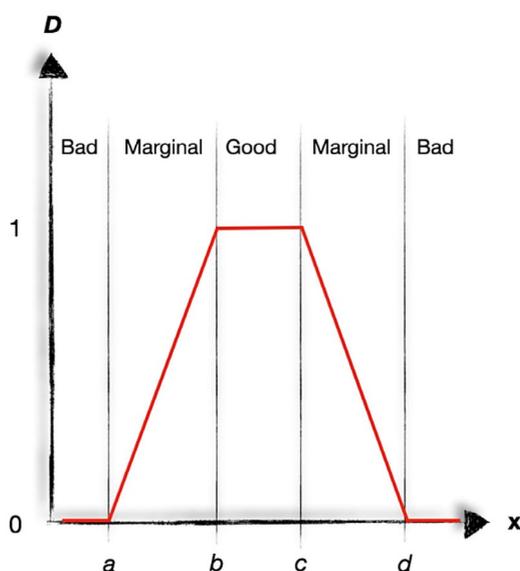


Рис. 7.4. Кривая с четырьмя точками перегиба

Когда свойства нормализованы, каждому из них назначается вес и суммарное значение MPO Score рассчитывается как среднее геометрическое:

$$MPOScore = \exp \left[\frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i \ln D_i \right],$$

где w_i – весовой коэффициент свойства i , D_i – нормализованное значение свойства i .

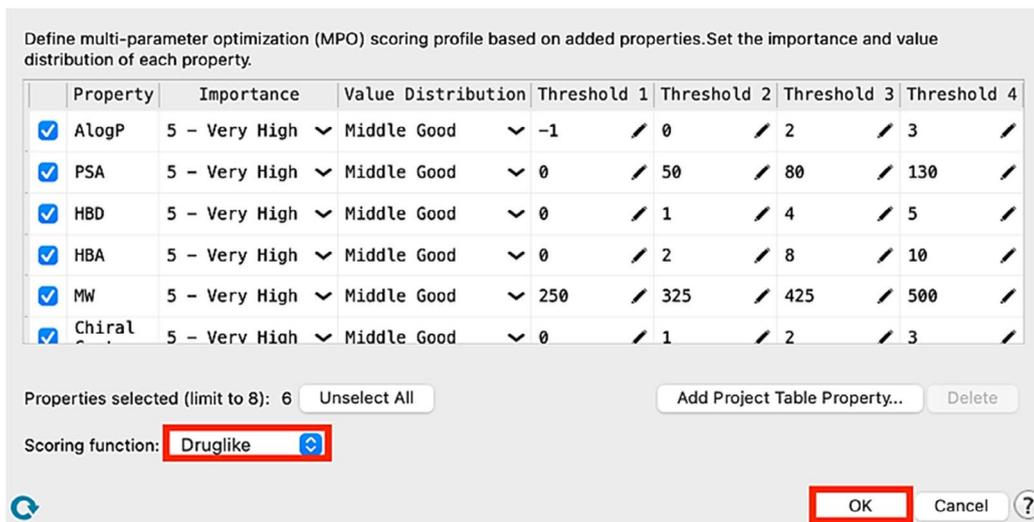


Рис. 7.5. Настройки вкладов и порогов MPO

Подробнее с подходами MPO заинтересованный читатель сможет ознакомиться в [5]. Панель настройки функции MPO Score Ligand Designer позволяет настроить значения порогов, диапазон желаемых значений, а также важность (т. е. вес) в ручном режиме (рис. 7.5) [3].

Экспериментальная часть

7.1. Создание проекта

В своем личном каталоге создайте папку LW7. Дважды щелкните по ярлычку Maestro на рабочем столе. На верхней панели перейдите на вкладку **File > Change Working Directory**. В открывшемся окне укажите путь к рабочей директории LW7 в своей личной папке.

Затем перейдите **File > Save project as**, задайте проекту имя LW7 в графе **File name**, укажите путь к рабочей директории в графе **Location**.

7.2. Оптимизация взаимодействий с использованием Ligand Designer

7.2.1. Подготовка структур белка и лиганда

С помощью модуля подготовки белка импортируйте и подготовьте комплекс ацетил-КоА карбоксилазы 2 (PDB ID: 5KKN [6]) в соответствии со стандартным протоколом. Удалите цепь В вместе с ее лигандом. Выполните сплитирование подготовленного комплекса (**Into Ligands, Water, Other**). Переименуйте белок и лиганд как «ACC Acceptor» и «ND-022» соответственно.

Подготовьте сетку потенциалов для нативного лиганда (не указывая ограничений на взаимодействия), а затем выполните его стыковку в активный сайт с использованием протокола **flexible Glide XP**.

7.2.2. Анализ Ligand Designer

Включите структуры ACC Acceptor и ND-022 в **Workspace**. Затем откройте модуль **Ligand Designer: Tasks > Browse > Lead Optimization > Ligand Designer**. Нажмите **Analyze Workspace**. Записи, содержащие лиганд и белок, для удобства работы будут автоматически добавлены в новую группу в **Workspace**. Расчет займет ориентировочно две минуты (рис. 7.6).

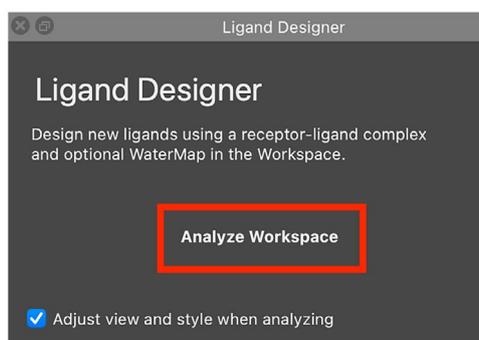


Рис. 7.6. Запуск анализа Ligand Designer

В разделе **Display** выберите **Ligand-Receptor Interactions**, в **Workspace** будут добавлены фармакофорные сайты рецептора, которые потенциально могут образовывать взаимодействия с лигандом (рис. 7.7).

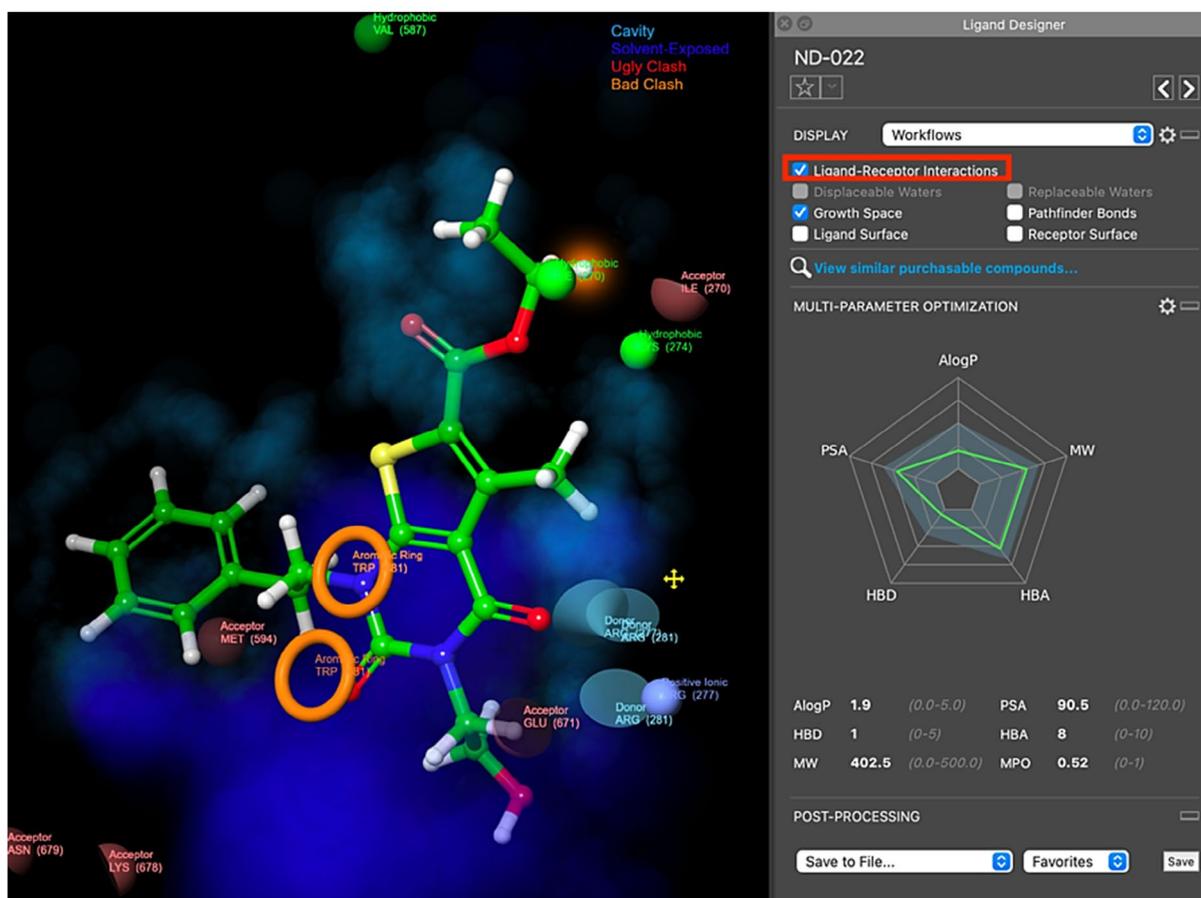


Рис. 7.7. Внешний вид Workspace при запуске Ligand Designer

Выберите в **Workspace** фармакофорный сайт, соответствующий **LYS 678**, нажмите **OK** в появившемся баннере. Можете воспользоваться **Structure Hierarchy** для быстрого поиска остатка. На бензольном кольце лиганда выберите выходной вектор так, как показано на рис. 7.8.

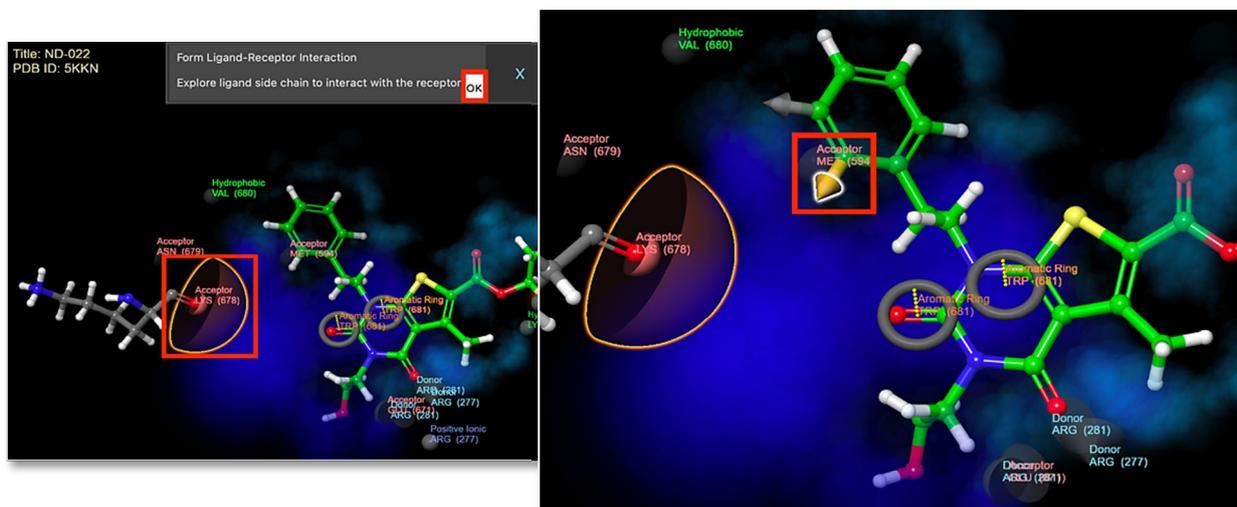


Рис. 7.8. Внешний вид Workspace при запуске Ligand Designer

Нажмите **Enumerate** во всплывшем баннере в верхней части **Workspace**. Расчет займет 5–10 минут. (рис. 7.9).

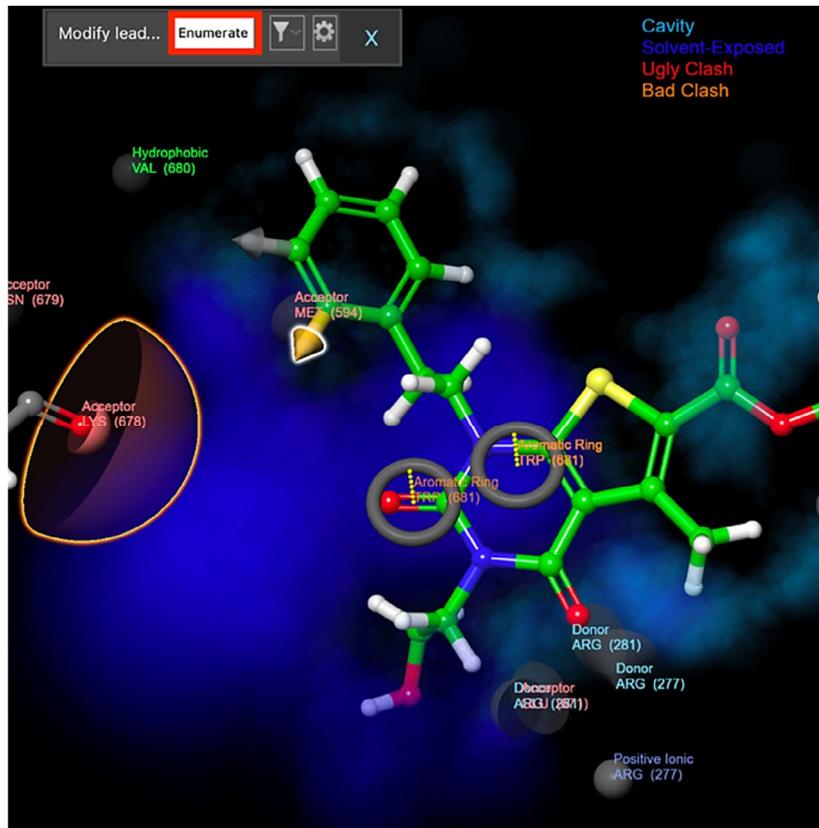


Рис. 7.9. Запуск задания Enumeration в Ligand Designer

Заметка: **Ligand Designer** выполнит достраивание лиганда с помощью библиотеки R-групп, а затем предскажет их ориентацию в сайте с помощью грубой стыковки; структуры, позы для которых найти не удастся, будут отброшены (рис. 7.10). Нажав значок **фильтра** рядом с **Enumerate**, вы сможете отфильтровать результаты по различным свойствам. Щелкнув **ЛКМ** по значку **шестеренки**, вы сможете настроить библиотеку для перечисления. По умолчанию **Ligand Designer** выбирает библиотеку, созданную для выбранного вами типа перечисления. Если вы измените библиотеку и нажмете **Save**, то она будет сохранена в качестве библиотеки по умолчанию (а **шестеренка** будет выделена **синим цветом**, что указывает на то, что используется пользовательская библиотека).

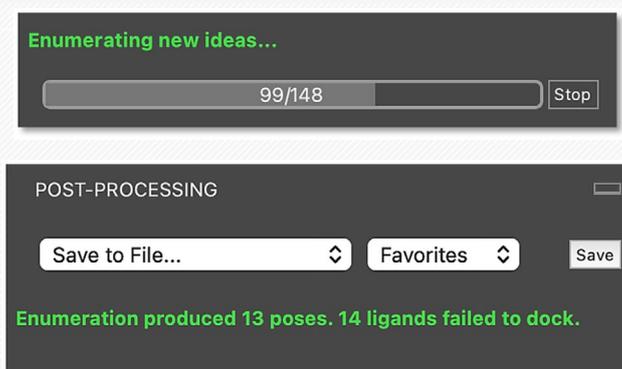


Рис. 7.10. Постпроцессинг Enumeration в Ligand Designer

По завершении расчета в **Entry List** будут добавлены лиганды, прошедшие стыковку. В панели многопараметрической оптимизации (*англ.* Multi-Parameter Optimization, MPO) обратите внимание на радиальную диаграмму распределения свойств. Настроим относительные веса параметров, нажав на значок **шестеренки** в верхней правой части панели. В открывшейся таблице в разделе **Scoring Function** выберите **Druglike**, нажмите **ОК**. Теперь график в панели MPO будет обновлен (рис. 7.11).

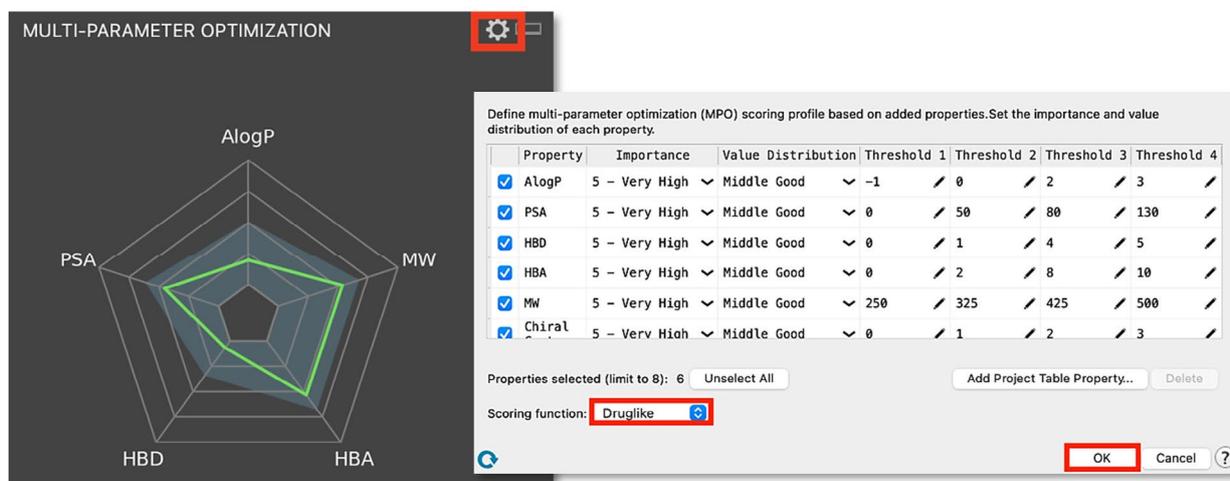


Рис. 7.11. Настройки вкладов MPO

Проведем анализ поз для достроенных лигандов. В **Structure Hierarchy** выберите **LYS 678** и сфокусируйтесь на нем в **Workspace**. При необходимости включите отображение взаимодействий. Находясь в панели **Ligand Designer**, дважды нажмите на клавишу «стрелка вправо» – структура лиганда **ND-022_1** теперь будет показана в **Workspace**. Обратите внимание, что график MPO теперь обновлен (рис. 7.12).

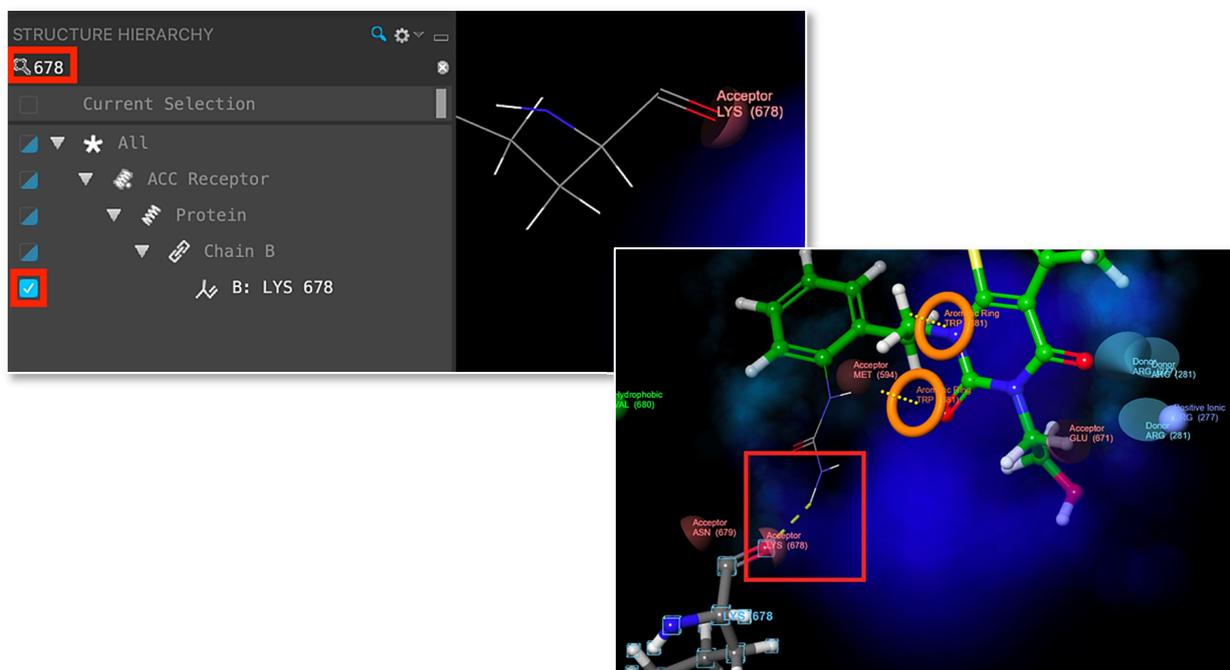


Рис. 7.12. Отображение лигандов для анализа поз

С помощью клавиши «стрелка вправо» вы можете поочередно включать в **Workspace** и другие лиганды, а также отмечать некоторые по своему усмотрению нажатием на значок звездочки (рис. 7.13).

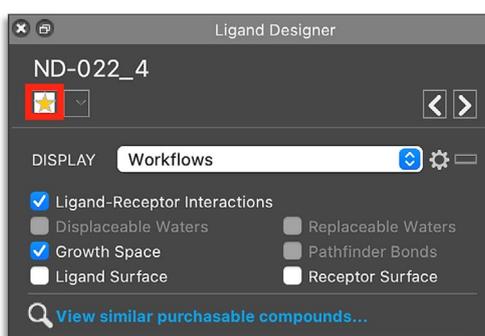


Рис. 7.13. Выделение понравившихся лигандов

Отсортируйте лиганды, выбрав **Sort by MPO** и нажав **Sort** в разделе **Post-processing**. Выделите **ND-022_3** среди отранжированных структур.

Попытаемся улучшить соединение-лидер. В разделе **Display** выберите **2D/3D Editing** и, находясь в режиме редактирования, замените гидроксил на аминогруппу так, как показано на рис. 7.14. Затем нажмите **Predict Pose** и закройте окно редактора.

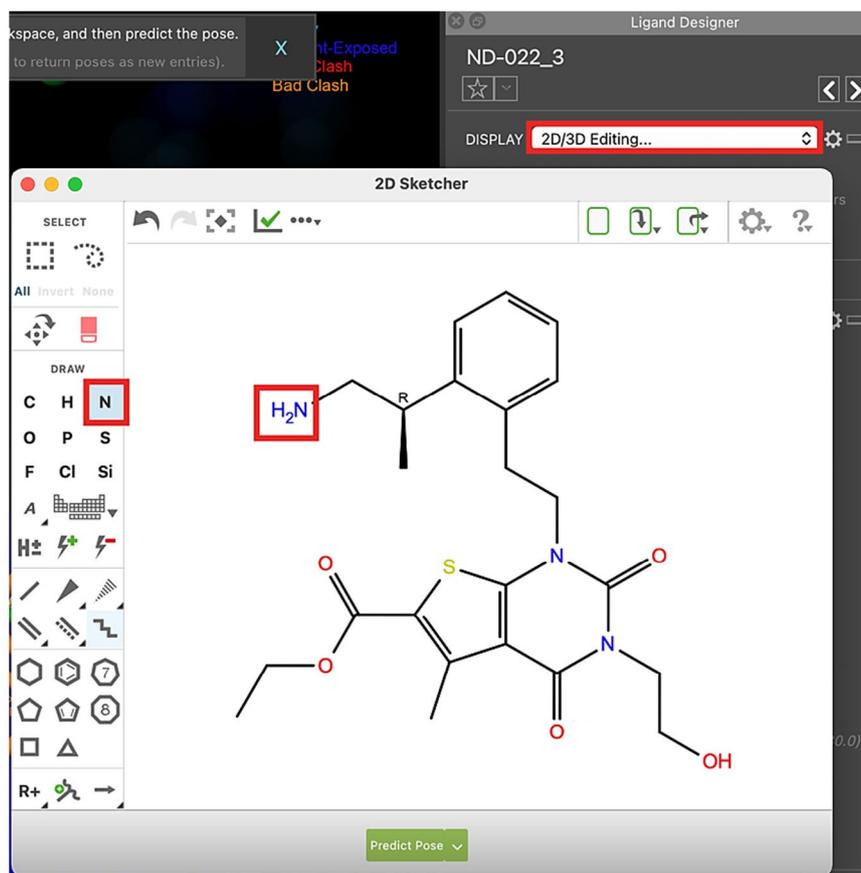


Рис. 7.14. Внесение изменений с помощью 2D-Sketcher

Заметка: вы можете редактировать лиганды **Ligand Designer** либо через **2D-Sketcher**, либо через панель **3D-Builder**. Любые изменения, внесенные вами, будут автоматически отражаться в **Workspace**.

7.3. Пост-процессинг с помощью Glide XP

Выделите все лиганды, для которых **Ligand Designer** нашел позы (в том числе для оптимизированного через **2D-Sketcher**), дублируйте их как несгруппированные (**ПКМ > Duplicate > As ungrouped entries**). Выполните их стыковку в ранее подготовленный вами грид-бокс в режиме **flexible Glide XP**.

Отразите в отчете:

- 1) насколько отличаются позы, предсказанные **Ligand Designer** от поз, предсказанных **Glide XP**? В отчете приведите снимки **Workspace** с наложением поз, где лиганды должны быть выделены разными цветами, а также приведены структуры соединений. Приведите значения RMSD, полученные с помощью модуля **Superimpose Structure**. Сделайте выводы о точности предсказания способа связывания **Ligand Designer**;
- 2) таблицу следующего содержания (табл. 7.1):

Таблица 7.1

Пример таблицы для отчета

Лиганд	XP GScore, ккал/моль	XP GScore ранг	MPO Score	MPO Score ранг
1	2	3	4	5

Столбцы 3 и 5 обозначают позицию лиганда в списке при сортировке по значению соответствующей функции. Изменился ли порядок ранжирования лигандов? Удалось ли добиться скоринга лучше, чем у нативного лиганда?

Вопросы к защите

1. Какие подходы к созданию виртуальных библиотек используются в Ligand Designer?
2. На каком этапе разработки лекарств может использоваться Ligand Designer?
3. Дайте определение понятию «пространство для роста» (*англ.* grow space).
4. Какие физико-химические свойства используются для расчета MPO Score?
5. Какие заместители (в первом приближении) не стоит располагать в областях сайта, доступных для растворителя?
6. Какие фармакофорные признаки используются в Ligand Designer?
7. Для чего используются отсечки (threshold) в MPO скоринге?
8. Какие вычислительные инструменты могут быть использованы для дальнейшей оценки «идей», полученных в результате работы с Ligand Designer?

Рекомендуемая литература

1. Ligand Designer // Schrödinger : [сайт]. – URL: www.schrodinger.com/science-articles/ligand-designer/ (дата обращения: 28.11.2024).
2. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding / R. Abel, T. Young, R. Farid [et al.] // Journal of American Chemical Society. – 2008. – Vol. 130, № 9. – P. 2817–2831.
3. Ligand Designer : User Manual. – New York, 2023.
4. Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties / T. T. Wager, X. Hou, P. R. Verhoest, A. Villalobos // ACS Chemical Neuroscience. – 2010. – Vol. 1, № 6. – P. 435–449.
5. Segall, M. D. Multi-Parameter Optimization: Identifying High Quality Compounds with a Balance of Properties // Current Pharmaceutical Design. – 2012. – Vol. 18, № 9. – P. 1292–1310.
6. Acetyl-CoA carboxylase inhibition by ND-630 reduces hepatic steatosis, improves insulin sensitivity, and modulates dyslipidemia in rats / G. Harriman, J. Greenwood, S. Bhat [et al.] // Proceedings of the National Academy of Sciences of the United States of America. – 2016. – Vol. 113, № 13. – P. 1796–1805.



Лабораторная работа 8

КОВАЛЕНТНЫЙ ДОКИНГ

Цель работы: ознакомление с подходами к стыковке ковалентных лигандов.

Вам предстоит:

- 1) подготовить кристаллическую структуру человеческого катепсина L для задач ковалентной стыковки;
- 2) выполнить докинг нативного лиганда в режиме Pose Prediction с помощью модуля Covalent Docking;
- 3) подготовить сфокусированную библиотеку ковалентных лигандов и провести их стыковку в режиме Virtual Screening с помощью модуля Covalent Docking;
- 4) самостоятельно предложить ряд структурных модификаций для необратимого ингибитора EGFR^{T790M} и проверить свои идеи с помощью ковалентной стыковки.

Теоретическое введение

Ковалентные лиганды являют собой уникальный класс биологически активных молекул, которые связываются с мишенью не только за счет водородных связей, солевых мостиков, гидрофобных взаимодействий и т. д., но и образуют полноценную ковалентную связь с остатками рецептора [1].

Взаимодействие ковалентных лигандов со своими мишенями происходит в два этапа. На первом, как и в случае обратимо связывающихся лекарств, образуется комплекс малой молекулы с рецептором, стабилизированный классическими нековалентными взаимодействиями. На втором реактивный фрагмент лиганда ориентируется относительно реакционно-способной аминокислоты белка и между ними происходит реакция, ведущая к образованию полноценной химической связи.

Лиганд считается связывающимся необратимо, если константа скорости образования ковалентного комплекса k_2 больше константы скорости диссоциации k_{-2} (в противном случае лекарство считается обратимым ковалентным лигандом) (рис. 8.1).

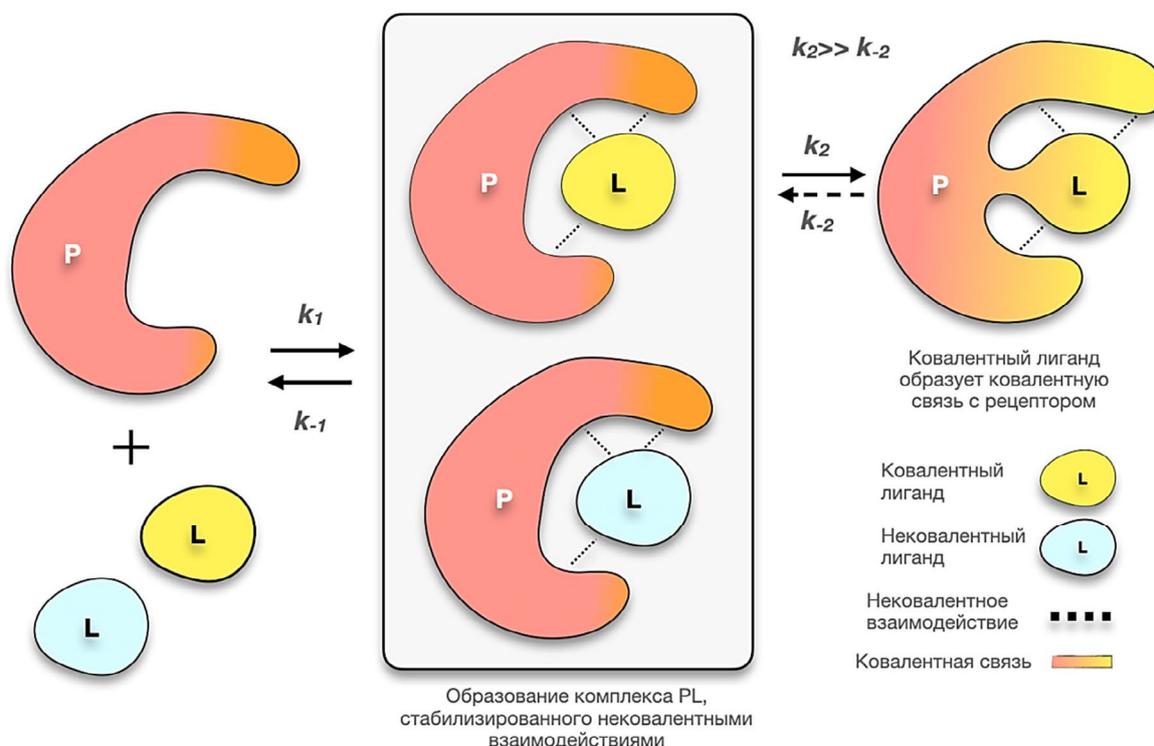


Рис. 8.1. Связывание ковалентных и нековалентных лигандов

Как правило, реактивными фрагментами лигандов являются электрофильные центры, такие как α , β -ненасыщенные карбонильные соединения, акриламиды, эпоксины, нитрилы, цианоакрилаты и др. (рис. 8.2).

Lys, Gln, Ser, Thr, Tyr и особенно Cys представляют пул нуклеофильных аминокислот, на взаимодействие с которыми могут нацеливаться ковалентные лиганды [2, 3].

В качестве примеров ковалентных лекарств можно привести **осимертиниб**, необратимый ингибитор киназной активности рецептора эпидермального фактора роста [4], **CCW-16** – ингибитор E3 убиквитинлигазы RNF4 [5], **ибрутиниб**, ингибитор тирозинкиназы Бруттона [6] (рис. 8.3).

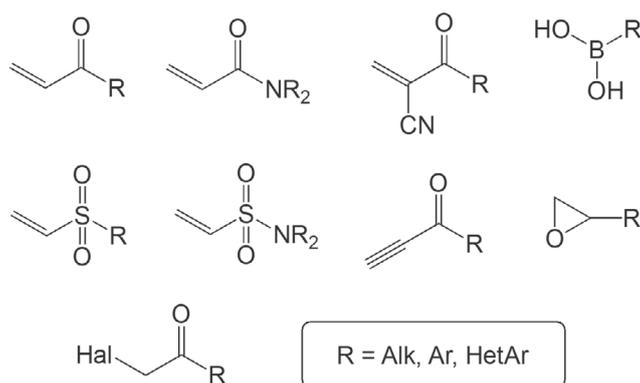


Рис. 8.2. Ковалентно-реактивные группы лигандов

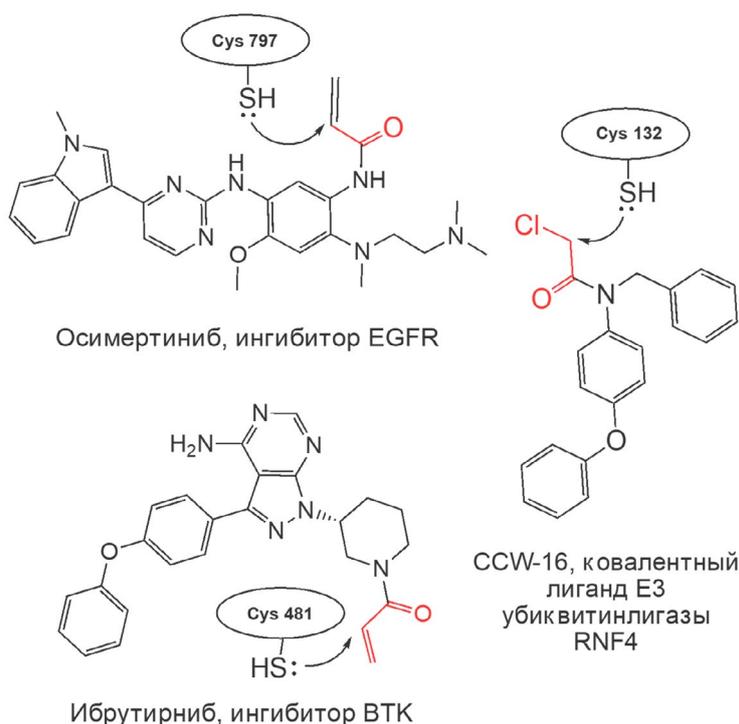


Рис. 8.3. Примеры лекарств, ковалентно связывающихся со своими мишенями

Фокусируясь на молекулах, нацеливающихся на остатки цистеина для необратимого связывания с мишенями, можно отметить, что большая часть боеголовок обеспечивает образование ковалентной связи по реакции сопряженного присоединения по Михаэлю или посредством алифатического нуклеофильного замещения (S_N2). На рис. 8.4 представлен механизм связывания для структур, содержащих акриламид в качестве реактивного фрагмента.

Необратимое ингибирование как терапевтическая стратегия обеспечивает ряд преимуществ перед использованием привычных нековалентных малых молекул. Так, для эффективного неравновесного связывания часто требуются меньшие концентрации биологически активного вещества. Следовательно, более мягкие режимы дозирования могут улучшить РК-профиль препарата. Также стоит отметить, что чувствительность препарата к потенциальным точечным мутациям аминокислот сайта (задействованным в нековалентном связывании) обычно снижается. Более того, появляется возможность связываться в неглубоких карманах рецепторов, что потенциально расширяет пул мишеней, активность которых можно модулировать.

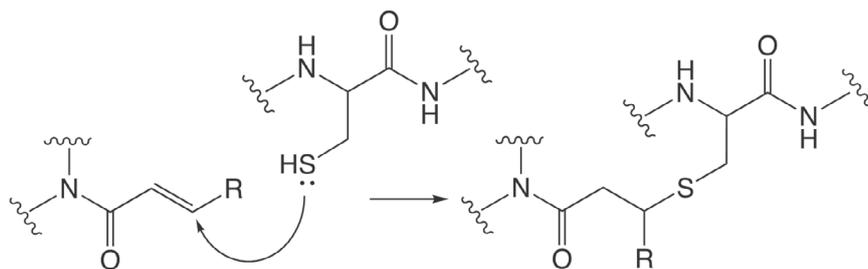


Рис. 8.4. Связывание акриамидной боеголовки с тиольной группой Cys

Однако необратимое связывание не является предпочтительной стратегией в том случае, когда требуется кратковременная инактивация целевого белка.

Также следует с осторожностью относиться к реакционной способности ковалентных лигандов. Гиперреактивные лиганды могут быть вовлечены в нецелевое связывание и стать причиной возникновения токсических эффектов, мутаций или нежелательных реакций иммунной системы [7].

Так или иначе, дизайн ковалентных ингибиторов без сомнения является актуальной и развивающейся областью медицинской химии.

Однако рядовые программы стыковки не подходят для прогнозирования поз связывания таких лигандов из-за отсутствия алгоритмов учета образования и оценки энергетических вкладов образования ковалентных связей. Кроме того, в целом работа с *реакцией* между нуклеофильным остатком белка и электрофильным лигандом требует явного рассмотрения электронного состояния, в первом приближении не может быть описана классическими межмолекулярными потенциалами [8].

Тем не менее существует довольно широкий пул решений проблемы ковалентной стыковки. Например, GOLD определяет атомы в лиганде и белке, которые будут участвовать в образовании ковалентной связи и выполняет первичный нековалентный докинг так, чтобы эти атомы находились рядом [9]. Autodock предлагает подход с использованием «гибкой боковой цепи», в котором ковалентно связанный с остатком лиганд рассматривается как часть рецептора и минимизируется [10].

Следует также отметить, что любое ПО для ковалентной стыковки требует от пользователя четкого понимания *сайта* (т. е. реактивного остатка) и *механизма* реакции между белком и лигандом. Алгоритм CovDock, реализованный в ПО Schrödinger, предлагает следующие типы возможных реакций (табл. 8.1). Реактивные атомы лигандов, равно как и остатки рецептора, задаются с помощью нотаций SMARTS.

Пользователь может добавить необходимые ему реакции, если не обнаружит их среди predefined шаблонов. Кроме того, паттерны фрагментов малых молекул также могут быть изменены [11].

Таблица 8.1

Шаблоны реакций CovDock

Реакция	SMARTS лиганда	Остаток рецептора
Beta Lactam Addition	[O-0X1]=[C]1[C][C][N]1	Cys, Ser
Boronic Acid Addition	[B]([O])[O]	
Conjugate Addition to Alkyne (aryl activated)	[C]#[C]-[c]	

Окончание табл. 8.1

Реакция	SMARTS лиганда	Остаток рецептора
Conjugate Addition to Alkyne (carbonyl activated)	<chem>[C-0X2]#[C-0X2][C-0X3]=[O-0X1]</chem>	Cys, Ser
Conjugate Addition to Alkene (nitrile activated)	<chem>[C,c]=[C,c]-[C,c]#[N,n]</chem>	
Disulfide Formation	<chem>[S;X2;H1]</chem>	–
Epoxyde Opening	<chem>[C;r3][O;r3][C;r3]</chem>	Cys, Ser, His
Imine Condensation	<chem>[C](=[O])-[C]</chem>	Lys
Ion Pair to Covalent Bond Lig Minus	<chem>1,-1]</chem>	–
Ion Pair to Covalent Bond Lig Plus	<chem>1,+1]</chem>	–
Michael Addition	<chem>[C,c]=[C,c]-[C,c,S,s]=[O]</chem>	Cys, Ser
Nucleophilic Addition to a Double Bond	<chem>[C,c]=[O,S]</chem>	
Nucleophilic Addition to a Triple Bond	<chem>[C]#[N]</chem>	Cys, Ser, Lys
Nucleophilic Substitution	<chem>[F,Cl,Br,I]</chem>	Cys, Ser, Lys
Phosphonate Addition	<chem>[P]-[O;H1,-1]</chem>	Cys, Ser

Рабочий процесс модуля CovDock представлен на рис. 8.5 [12, 13].

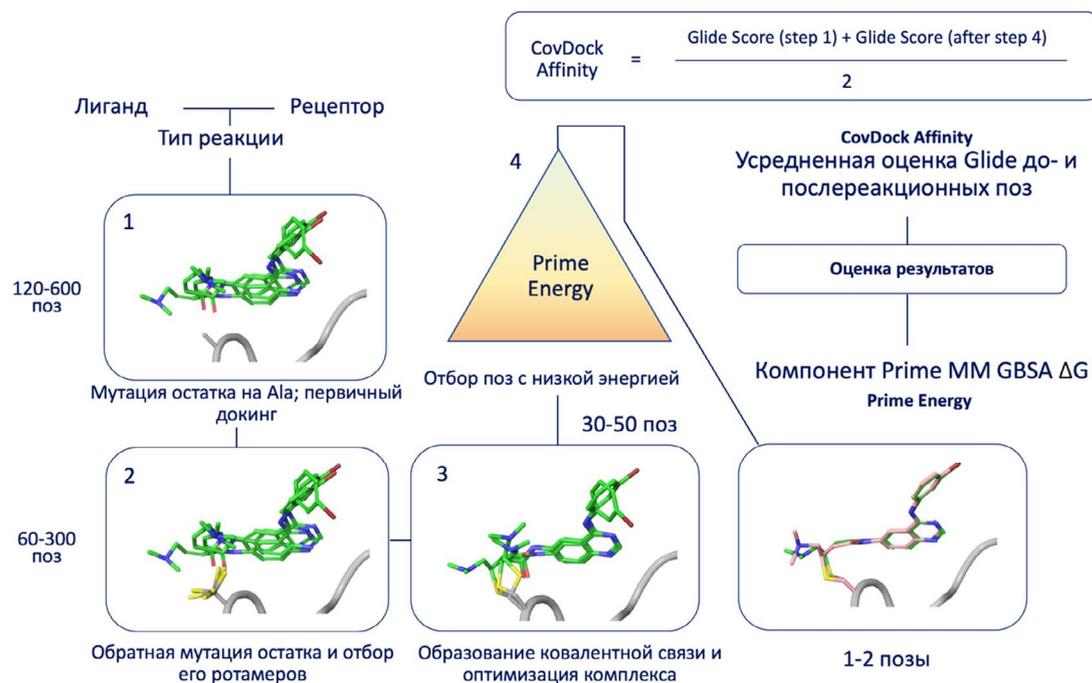


Рис. 8.5. Рабочий процесс CovDock [12]

Когда определена область стыковки, выбран реактивный остаток и тип реакции, выполняется первичный нековалентный докинг Glide. Чтобы избежать столкновений с белком и минимизировать конформационную ошибку, реактивный остаток мутируют

в аланин. Позиционное ограничение в Glide удерживает боеголовку лиганда в пределах 8 Å от C-β атома реакционноспособного остатка. Все положения лигандов с GlideScore ± 2.5 ккал/моль от наименьшего значения выборки сохраняются.

Реактивный остаток мутируют обратно и для него выполняется ограниченный конформационный поиск, направленный на идентификацию набора стабильных ротамеров. Каждый выбранный способ связывания проверяется на предмет того, находятся ли два атома, которые в конечном итоге образуют ковалентную связь, в пределах 5 Å друг от друга. Любая поза, не удовлетворяющая этому критерию, отбрасывается.

Ковалентный комплекс образуется, учитываются все изменения в порядке связи и протонировании лиганда.

Комплексы минимизируются в вакууме для восстановления нормальной длины связей, устранения стерических столкновений и релаксации структуры с помощью модели Prime. Повторно рассчитывается значение GlideScore в режиме *score-in-place*.

Эмпирическая оценочная функция основывается на усредненном GlideScore, полученном на двух этапах моделируемого процесса реакции: при первичном докинге *pre-reactive* видов лиганда и при оценке финальных поз [12].

Подробнее про применение ковалентного докинга в открытии лекарств можно прочитать в работах [14–16].

Экспериментальная часть

8.1. Создание проекта

В своем личном каталоге создайте папку LW8. Дважды щелкните по ярлыку Maestro на рабочем столе. На верхней панели перейдите на вкладку **File > Change Working Directory**. В открывшемся окне укажите путь к рабочей директории LW8 в своей личной папке.

Затем перейдите **File > Save project as**, задайте проекту имя LW8 в графе **File name**, укажите путь к рабочей директории в графе **Location**.

8.2. Ковалентный докинг

8.2.1. Подготовка структур белка и лиганда

С помощью модуля подготовки белка импортируйте из PDB файл комплекса сериновой протеазы – человеческого катепсина L – вместе с лигандом (PDB ID: 5MQY [17]). Импортируйте только цепь A. В разделе **Preprocess** выберите пункты **Fill in missing loop using Prime** и **Cap Termini**. Остальную подготовку выполните в соответствии со стандартным протоколом. Можете выбрать режим **Interactive** (на ваше усмотрение) (рис. 8.6).

В строке избранных заданий откройте **Ligand Interaction Diagram** для того, чтобы лучше познакомиться со способом связывания ковалентного лиганда. Нажатие на значок **шестеренки** в панели **2D Workspace – Ligand Interaction Diagram** откроет доступ к легенде. В структуре должна остаться вода. Не забудьте закрыть окно (рис. 8.7).

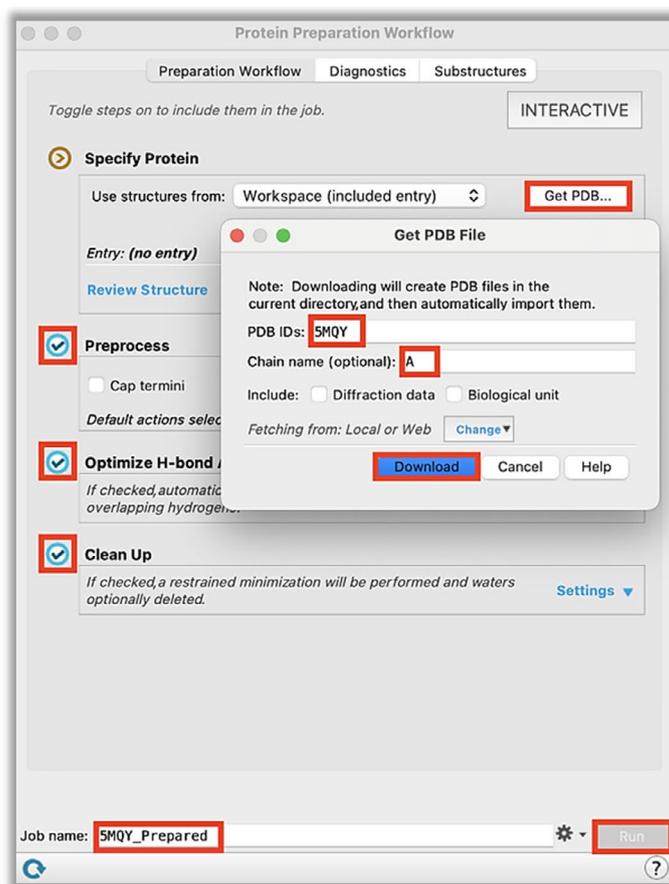


Рис. 8.6. Подготовка структуры комплекса для ковалентной стыковки

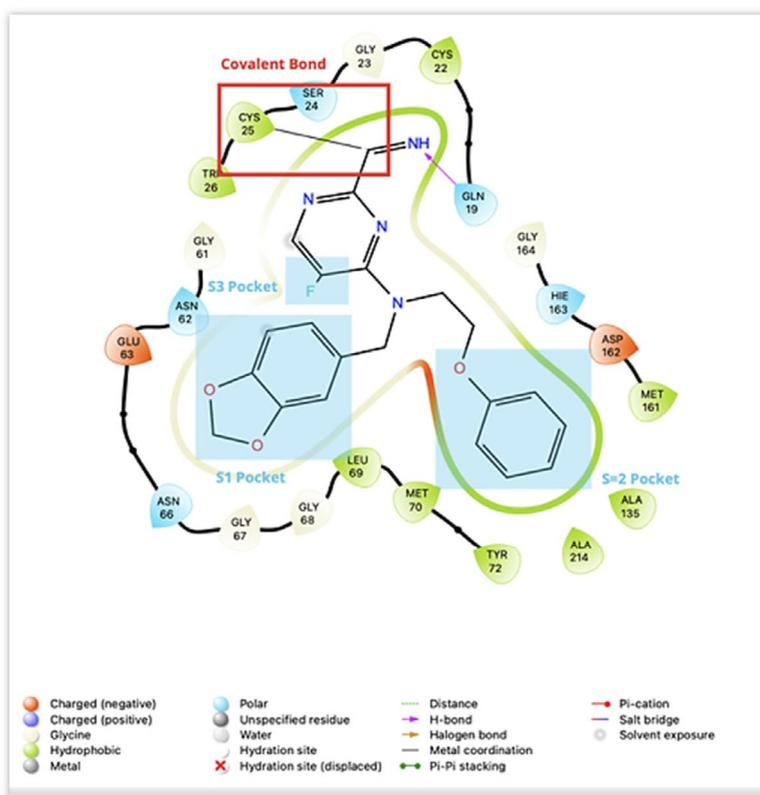


Рис. 8.7. Диаграмма, показывающая способ связывания лиганда в сайте

Отразите в отчете: снимок **Workspace**, в котором покажите области **S1-S3**, ориентируясь на **2D-диаграмму**. Для удобства можете отобразить молекулярную поверхность сайта связывания, для этого в строке избранных заданий щелкните по **Surface (Binding Site)**, в открывшемся окне в разделе **Surface Type** выберите **Molecular surface**, снимите галочку у пункта **Ligand** (чтобы сгенерировать только поверхность рецептора), **Style** и **Color Scheme** можете настроить на свое усмотрение. Нажмите **OK** (рис. 8.8).

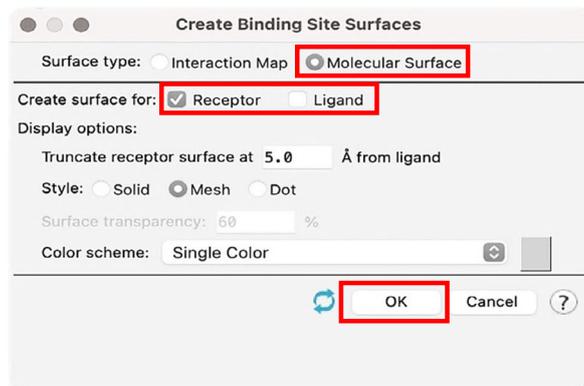


Рис. 8.8. Отображение молекулярной поверхности активного сайта

Импортируйте файл лиганда для докинга **5MQY_Prepared_ligand_modified.maegz** (**File > Import** или **Ctrl+I**). Включите структуру подготовленного комплекса в **Workspace** и выберите запись лиганда в **Entry List**. Лиганд был модифицирован путем замены енаминового фрагмента на нитрил и подготовлен с использованием **LigPrep**.

Перейдите в модуль ковалентного докинга: **Tasks > Browse > Glide > Covalent Docking**. В открывшейся панели **Covalent Docking** выберите **Project Table (selected entry)** рядом с **Use ligands from**. В разделе выбора остатка для стыковки **Reactive Residue** нажмите кнопку **Pick** и затем выберите **CYS 25** в **Workspace**. Аминокислота подсветится в **Workspace** и **A:25** появится в строке рядом с **Reactive Residue**. По аналогии нажмите **Pick** и выделите нативный лиганд в разделе **Centroid of Workspace ligand**. Лиганд будет считан и обозначен как **A:301**. После совершения всех процедур в **Workspace** появится фиолетовый куб (похожий на тот, что создается при генерации сетки потенциалов), ограничивающий область стыковки, а остаток цистеина будет подсвечен (рис. 8.9).

Переместитесь на вкладку **Reaction** и выберите **Nucleophilic Addition to a Triple Bond** в качестве реакции.

Отразите в отчете: механизм образования ковалентной связи между лигандом и остатком CYS 25.

Переместитесь на вкладку **Docking**. В разделе **Docking Mode** выберите высокоточный режим предсказания позы (**Pose Prediction (Thorough)**). Установите количество поз лиганда равным «1» (**Output poses per ligand reaction site**). Измените название задания на «**5MQY_Pose_Prediction**» и нажмите **Run**. Расчет займет ориентировочно 20–30 минут (рис. 8.10).

В случае успешного завершения задания стыковки в **Entry List** будет автоматически импортирована группа **5MQY_Pose_prediction-out**, содержащая запись **5MQY – prepared_ligand modified**, соответствующую комплексу, полученному в результате ковалентного докинга. Включите эту структуру в **Workspace**. Также включите структуру подготовленного комплекса, который не был подвергнут стыковке. Дважды щелкните по панели **Presets** (рис. 8.11).

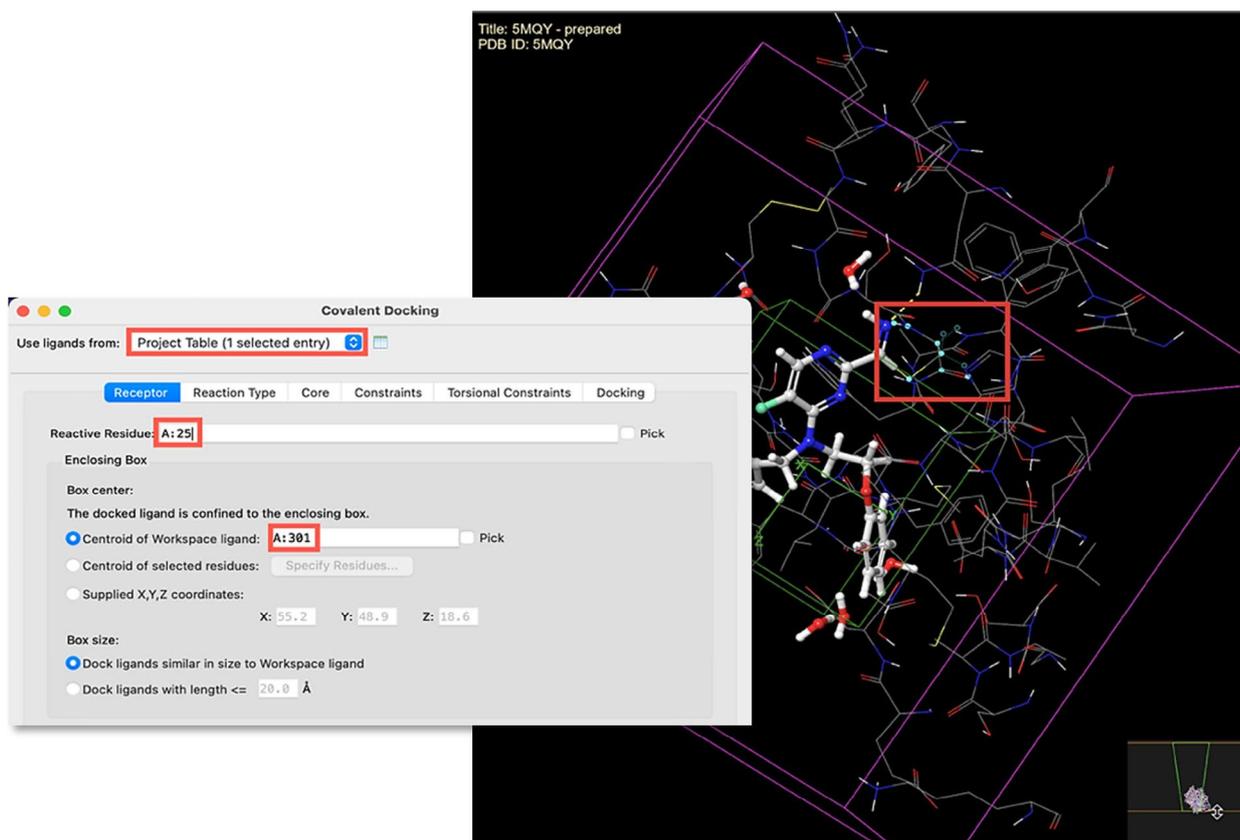


Рис. 8.9. Определение реактивного остатка и лиганда для ковалентной стыковки

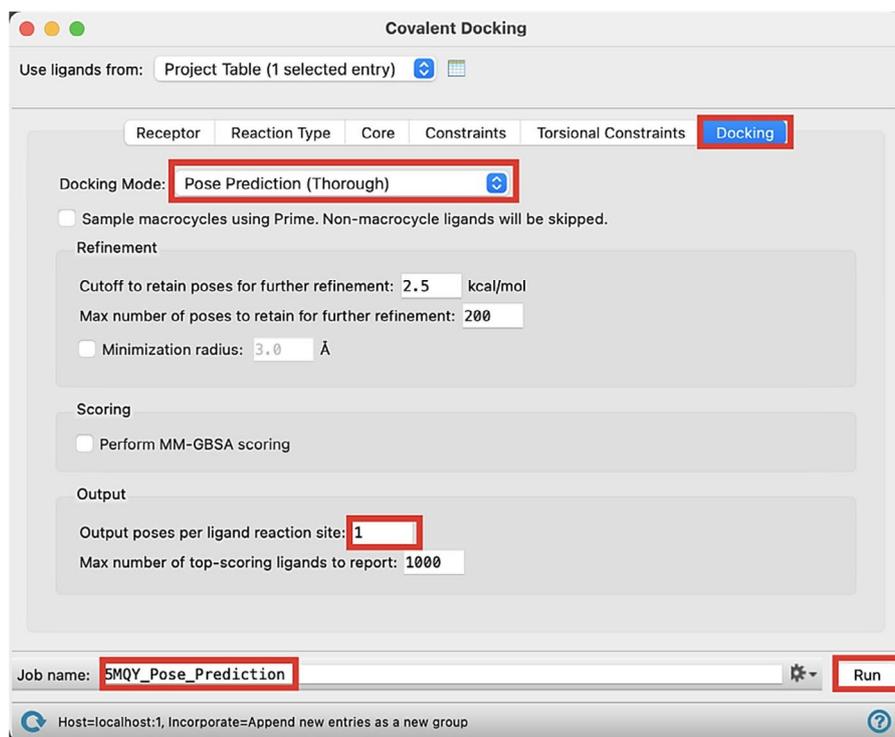


Рис. 8.10. Установка точности стыковки и запуск задания

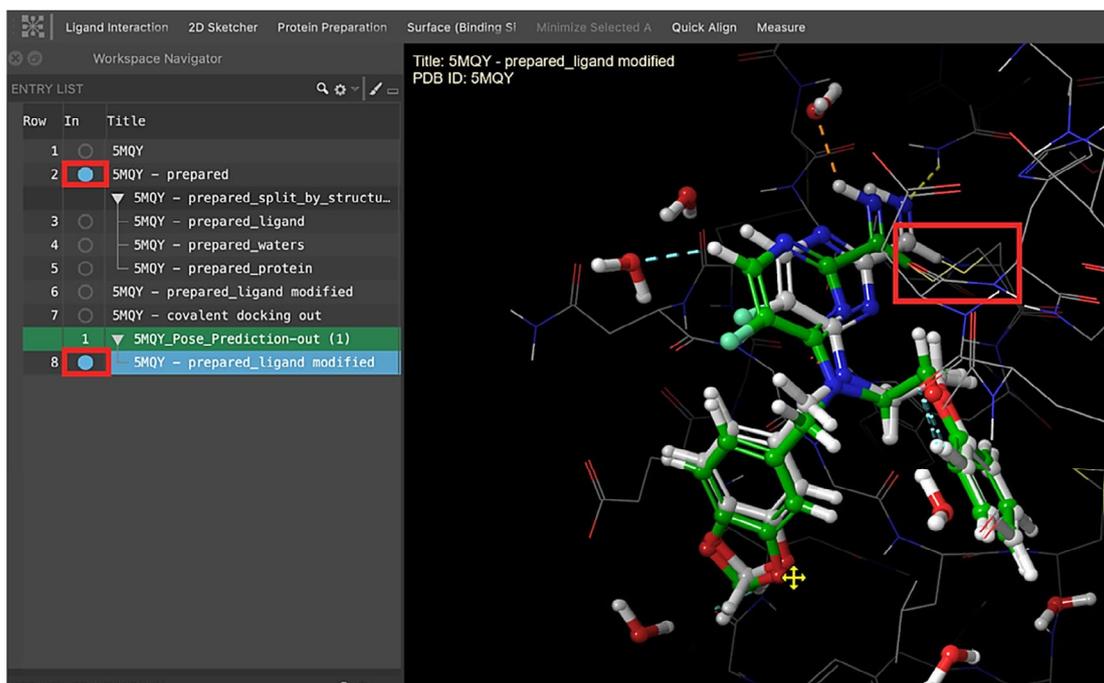


Рис. 8.11. Наложение лигандов

Отразите в отчете: снимок **Workspace**, демонстрирующий наложение кристаллического комплекса и комплекса, полученного после стыковки. Также приведите значения **cdock_score** и **Prime energy**.

Проведем теперь ковалентный докинг в менее точном, но быстром режиме **Virtual Screen**. Для этого первоначально создадим небольшую скрининговую библиотеку, в которой будут перечислены различные заместители, направленные на оптимизацию связывания в области сайта **S2**.

Импортируйте файл **template.scaffold.mae**, включите в **Workspace** и выделите в **Entry List** структуру **Template_Scaffold**. Выделите атом углерода, прикрепленный к аминному азоту и обращенный в область **S2** (рис. 8.12).

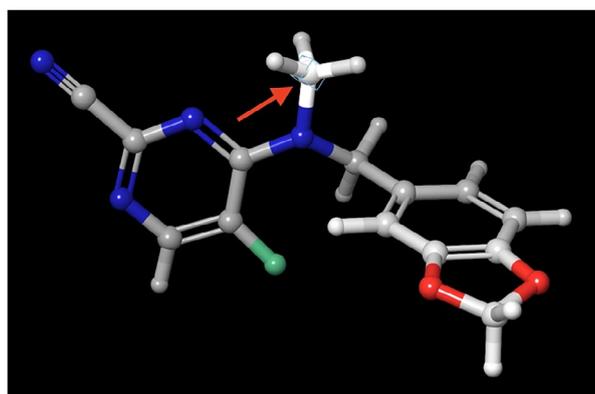


Рис. 8.12. Выделение точки прикрепления заместителей

Откройте модуль создания библиотек: **Tasks > Browse > Custom R-Group**. В столбце **R-Group Library** раскройте выпадающий список, выберите **file** и откройте файл библиотеки **CovDock_Tutorial_Ligands_for_S2_Pocket.mae**. Измените название работы на «**S2_Pocket_Enumeration**» (рис. 8.13).

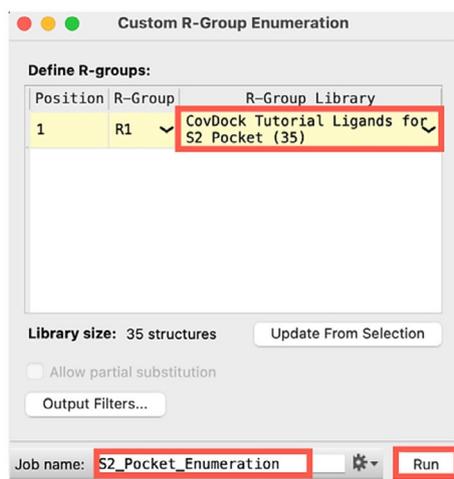


Рис. 8.13. Запуск процесса создания библиотеки

Подготовим лиганды библиотеки. Выделите всю группу **S2_Pocket_Enumeration**. Откройте модуль подготовки малых молекул, выставьте настройки так, как показано на рисунке ниже, измените название работы на «**Ligprep_Enumeration**» и нажмите **Run**. Расчет займет не более 5 минут (рис. 8.14).

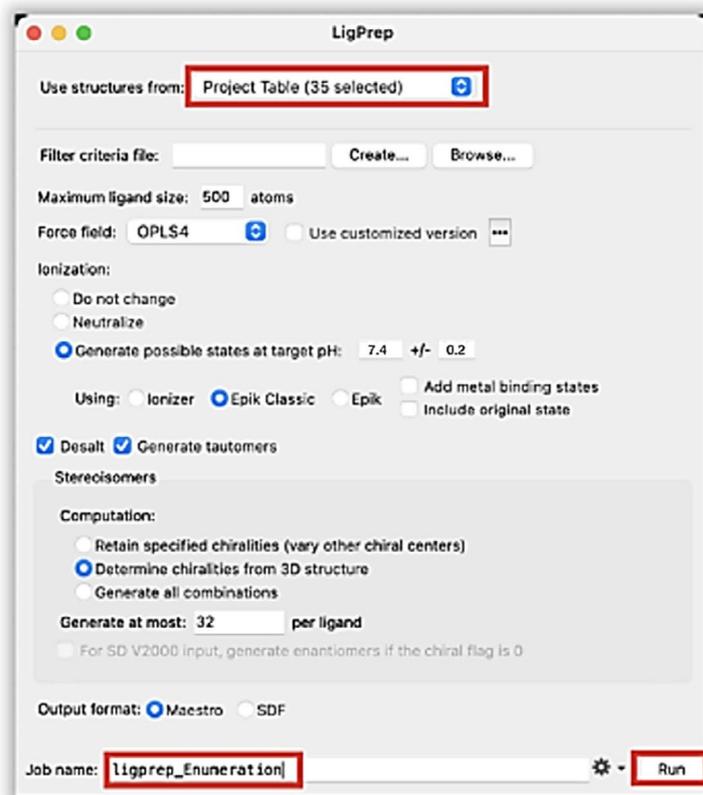


Рис. 8.14. Подготовка лигандов библиотеки

Отразите в отчете: аннотирование полученной библиотеки. Для этого рассчитайте физико-химические свойства с помощью модуля **QikProp**, постройте графики распределения с помощью **Advanced Project Table Plot**. Отражите структуры лигандов (так, как было описано в предыдущих работах).

После завершения расчета приступим к ковалентной стыковке. Выделите всю группу **ligprep_Enumeration_Prepared-out1**, модифицированный лиганд **5MQY_Prepared_ligand modified** в **Entry List**, а структуру подготовленного рецептора включите в **Workspace**. Откройте модуль ковалентной стыковки. В разделе **Use ligands from** выберите **Project Table (selected entry)**.

Все настройки определения рецептора, остатка и типа реакции остаются теми же, что были использованы ранее для докинга в режиме **Pose Prediction**. По умолчанию **Covalent Docking** запоминает их, но при необходимости самостоятельно выставьте так, как было описано ранее.

Переместитесь на вкладку **Docking** и установите в разделе **Docking Mode** режим **Virtual Screening (Fast)**. Измените название работы на «**5MQY_Virtual_Screening_S2_Pocket**». Нажмите **Run**. Расчет займет ориентировочно 50–70 минут. При необходимости получите разрешение преподавателя на использование прегенерированных результатов (рис. 8.15).

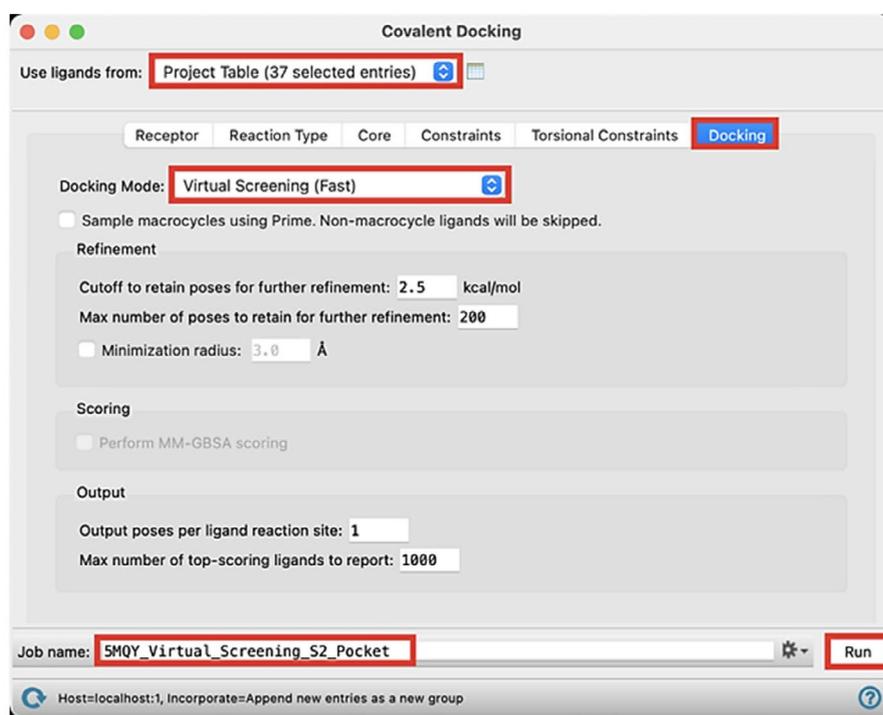


Рис. 8.15. Запуск задания стыковки в режиме Virtual Screening

В случае успешного завершения расчета группа, содержащая комплексы для 38 лигандов, будет импортирована в **Entry List**. Выделите ее и откройте модуль **Pose Viewer: Tasks > Glide > Pose Viewer**, затем нажмите **Set Up Poses**. Визуально осмотрите позы.

Отразите в отчете: таблицу со значениями **cdock_affinity** и **Prime Energy**. Удалось ли лигандам библиотеки показать лучший скоринг по сравнению с нативным? Опираясь на сравнение режимов связывания, ответьте на вопрос, почему удалось / не удалось.

8.2.2. Задание для самостоятельной работы

Вам предстоит выполнить стыковку небольшой библиотеки аналогов афатиниба (рис. 8.16) [18, 19] в активный сайт киназного домена EGFR, несущего мутацию T790M.

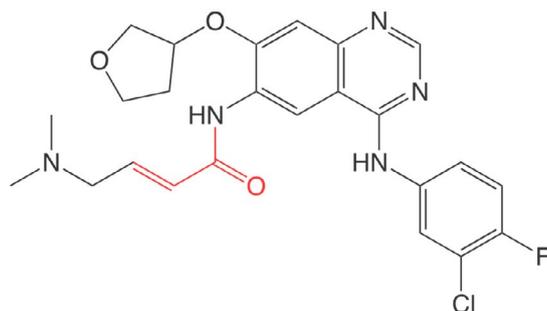


Рис. 8.16. Структура афатиниба, необратимого ингибитора EGFR

Найдите в PDB комплекс мутантного EGFR с афатинибом, импортируйте его в Maestro и подготовьте с помощью **Protein Preparation Workflow** в соответствии со стандартным протоколом. В разделе **Preprocess** выберите пункты **Fill in missing loop using Prime** и **Cap Termini**. Вся структурная вода и посторонние молекулы должны быть удалены.

Дублируйте запись подготовленного комплекса в **Entry List (ПКМ > Duplicate > In place)**, затем выделите атомы, участвующие в образовании ковалентной связи между афатинибом и **CYS 797** в **Workspace**, и удалите ее, используя модуль **Build**. С помощью все того же модуля восстановите структуру тиольной группы CYS 797 (рис. 8.17).

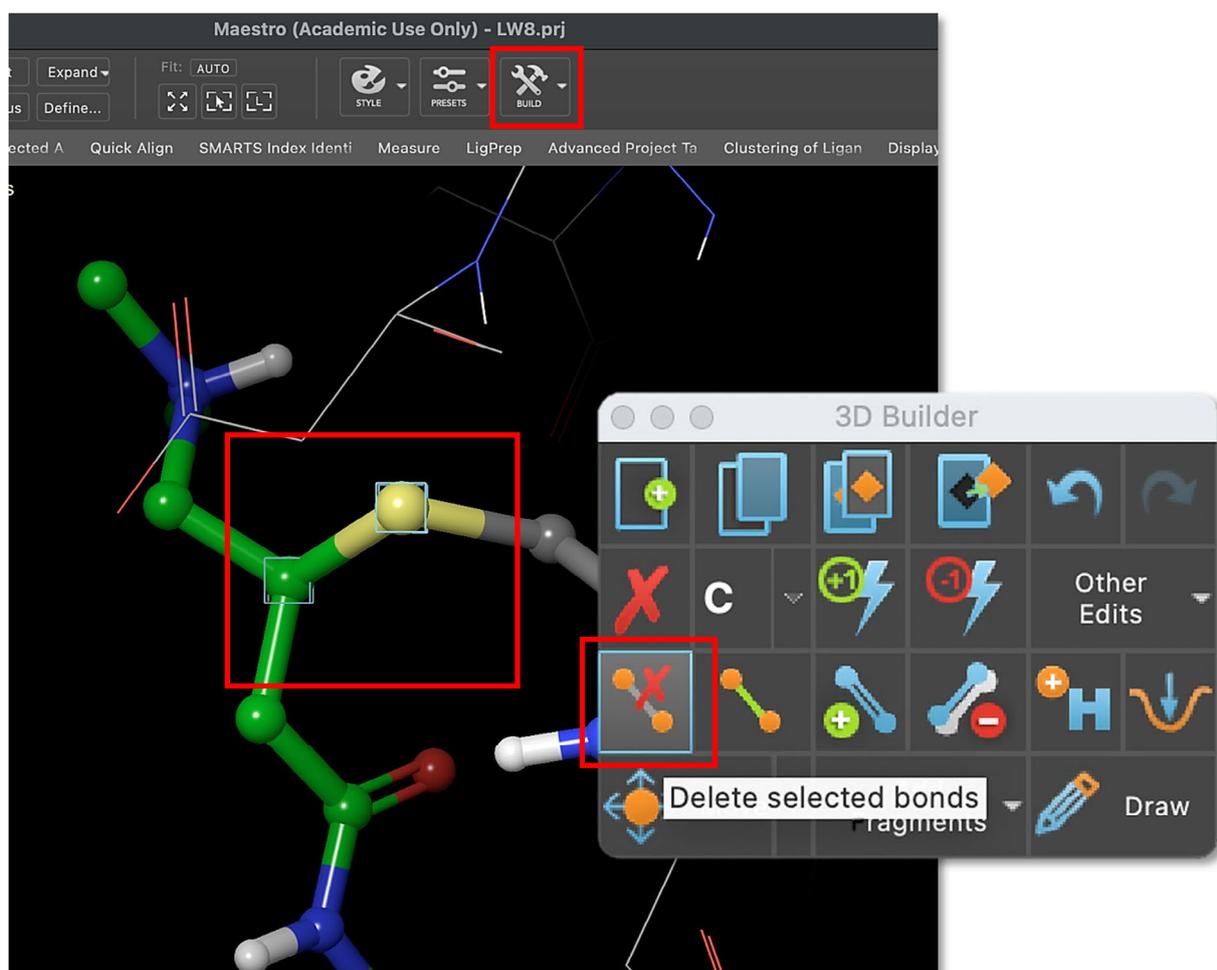


Рис. 8.17. Удаление ковалентной связи между белком и лигандом

Сплитируйте запись дубликата и добавьте двойную связь в молекулу афатиниба, используя **2D-Sketcher**. Выполните подготовку молекулы нативного лиганда в соответствии со стандартным протоколом, используя алгоритм **LigPrep**.

Осуществите нативный докинг подготовленного лиганда с помощью модуля **Covalent Docking**. На вкладке **Receptor** в разделе **Enclosing Box** центрируйте область стыковки на реактивном остатке **CYS (Centroid of selected residues)** и увеличьте длину лигандов, которые могут быть пристыкованы (см. пункт **Dock ligands with length**), до **30 Å**. Запустите задание в режимах **Virtual Screening** и **Pose Prediction** (для данного режима дополнительно выполните расчет энергии связывания с помощью ММ-GBSA).

Отразите в отчете: наложение кристаллического и полученного в результате докинга комплексов. Также приведите значения скоринга.

Сгенерируйте библиотеку структур для докинга с помощью модуля **Custom R-Group Enumeration**. Ваша задача – оптимизировать структуру афатиниба для повышения его растворимости. Самостоятельно выберите фрагмент молекулы для модификации, но учтите, что вносимые вами изменения не должны привести к потере способности образовывать ключевые взаимодействия. Вы можете использовать готовые библиотеки **R-Group** или создать свою. При необходимости, выполните фильтрацию.

Отразите в отчете: аннотирование структур и физико-химических свойств созданных вами соединений. Обязательно укажите рассчитанные **QikProp** значения свойств, характеризующих растворимость.

Осуществите ковалентный докинг соединений в активный сайт EGFR в режиме **Virtual Screening**. Номинируйте несколько соединений (не более 10) в качестве хитов. Свой выбор объясните. Повторно пристыкуйте отобранные лиганды к EGFR в режиме **Pose Prediction** с дополнительным выполнением оценки ММ-GBSA.

Отразите в отчете: таблицу с результатами следующего содержания (табл. 8.2).

Таблица 8.2

Результаты ковалентной стыковки

Код соединения	Режим точности			
	Virtual Screening	Pose Prediction		
	Prime Energy, ккал/моль	Prime Energy, ккал/моль	Cdock affinity, ккал/моль	MM-GBSA ΔG_{bind} , ккал/моль

Вопросы к защите

1. Зачем мутировать реактивный остаток на Ala в ходе ковалентного докинга?
2. Зачем в выражение для расчета CovDock Affinity включают значение GlideScore как до, так и после образования ковалентной связи?
3. Чем режим Pose Prediction отличается от режима Virtual Screening?
4. Какой информацией о структуре рецептора / механизме ингибирования необходимо владеть, чтобы выполнить ковалентную стыковку?
5. Какие достоинства стратегии ковалентного ингибирования вы можете выделить?

6. Какие недостатки стратегии ковалентного ингибирования вы можете выделить?
7. С какими аминокислотами могут образовывать связь ковалентные ингибиторы? С помощью каких хемотипов?
8. Чем ковалентные ингибиторы отличаются от нековалентных с точки зрения кинетики связывания?

Рекомендуемая литература

1. *Sutanto, F.* Covalent inhibitors: a rational approach to drug discovery / F. Sutanto, M. Konstantinidou, A. Dömling // RSC Medicinal Chemistry. – 2020. – Vol. 11, № 8. – P. 876–884.
2. *Lagoutte, R.* Covalent inhibitors: an opportunity for rational target selectivity / R. Lagoutte, R. Patouret, N. Winssinger // Current Opinion in Chemical Biology. – 2017. – Vol. 39. – P. 54–63.
3. Development of covalent inhibitors: Principle, design, and application in cancer / L. Zheng, Y. Li, D. Wu [et al.] // MedComm Oncology. – 2023. – Vol. 2, № 4. – Article number e56.
4. *Greig, S. L.* Osimertinib: First Global Approval // Drugs. – 2016. – Vol. 76, № 2. – P. 263–273.
5. Covalent Ligand Screening Uncovers a RNF₄ E₃ Ligase Recruiter for Targeted Protein Degradation Applications / C. C. Ward, J. I. Kleinman, S. M. Brittain [et al.] // ACS Chemical Biology. – 2019. – Vol. 14, № 11. – P. 2430–2440.
6. *Roskoski, R.* Orally effective FDA-approved protein kinase targeted covalent inhibitors (TCIs) // Pharmacological Research. – 2021. – Vol. 165. – Article number 105422.
7. *Bauer, R.* Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies // Drug Discovery Today. – 2015. – Vol. 20, № 9. – P. 1061–1073.
8. *Christoph, S.* Docking of Covalent Ligands: Challenges and Approaches // Molecular Informatics. – 2018. – Vol. 37, № 9–10. – Article number e1800062.
9. Development and Validation of a Genetic Algorithm for Flexible Docking / G. Jones, P. Willett, R. C. Glen [et al.] // Journal of Molecular Biology. – 1997. – Vol. 267. – P. 727–748.
10. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility / G. M. Morris, R. Huey, W. Lindstrom [et al.] // Journal of Computational Chemistry. – 2009. – Vol. 30, № 16. – P. 2785–2791.
11. CovDock // Schrödinger : [сайт]. – URL: [newsite.schrodinger.com/life-science/learn/white-papers/covdock/](https://www.schrodinger.com/life-science/learn/white-papers/covdock/) (дата обращения: 03.09.2025).
12. Docking Covalent Inhibitors: A Parameter Free Approach To Pose Prediction and Scoring / K. Zhu, K. W. Borrelli, J. R. Greenwood [et al.] // Journal of Chemical Information and Modeling. – 2014. – Vol. 54, № 7. – P. 1932–1940.
13. Structure-Based Virtual Screening Approach for Discovery of Covalently Bound Ligands / T. D. Warshaviak, G. Golan, K. W. Borrelli, K. Zhu // Journal of Chemical Information and Modeling. – 2014. – Vol. 54, № 7. – P. 1941–1950.
14. Covalent Docking and Molecular Dynamics Simulations Reveal the Specificity-Shifting Mutations Ala237Arg and Ala237Lys in TEM Beta-Lactamase / G. M. da Silva, J. Yang, B. Leang [et al.] // PLOS Computational Biology. – 2022. – Vol. 18, № 6. – Article number e1009944.

15. Modeling receptor flexibility in the structure-based design of KRAS^{G12C} inhibitors / K. Zhu, C. Li, K. Y. Wu [et al.] // *Journal of Computer-Aided Molecular Design*. – 2022. – Vol. 36, № 8. – P. 591–604.
16. Systematic Studies on the Protocol and Criteria for Selecting a Covalent Docking Tool / C. Wen, X. Yan, Q. Gu [et al.] // *Molecules*. – 2019. – Vol. 24, № 11. – Article number 24112183.
17. Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors / B. Kuhn, M. Tichy, L. Wang [et al.] // *Journal of Medicinal Chemistry*. – 2017. – Vol. 60, № 6. – P. 2485–2497.
18. Target Binding Properties and Cellular Activity of Afatinib (BIBW 2992), an Irreversible ErbB Family Blocker / F. Solca, G. Dahl, A. Zoephel [et al.] // *The Journal of Pharmacology and Experimental Therapeutics*. – 2012. – Vol. 343, № 2. – P. 342–350.
19. Afatinib: Emerging next-generation tyrosine kinase inhibitor for NSCLC / V. Nelson, J. Ziehr, M. Agulnik, M. Johnson // *Onco Targets and Therapy*. – 2013. – Vol. 6. – P. 135–143.



Лабораторная работа 9

УЧЕТ КОНФОРМАЦИОННОЙ ПОДВИЖНОСТИ РЕЦЕПТОРА В МОЛЕКУЛЯРНОМ ДОКИНГЕ

Цели работы: освоение основных подходов к молекулярной стыковке, позволяющих учесть конформационную подвижность рецептора.

Вам предстоит:

- 1) подготовить кристаллический комплекс нитрозфирредуктазы PETN с низкомолекулярным лигандом;
- 2) на основе информации о значениях температурных факторов провести стандартный молекулярный докинг Glide со смягчением ван-дер-ваальсовых и электростатических потенциалов;
- 3) с помощью модуля Induced Fit Docking выполнить индуцированную стыковку.

Теоретическое введение

Белки являются структурами, которые меняют свою конформацию под влиянием условий внешней среды или при взаимодействии с другими молекулами. Такая конформационная гибкость имеет решающее значение для многих клеточных процессов, таких как передача сигналов, иммунный ответ и ферментативный катализ.

Подвижность белка во многом опосредуется термическими колебаниями его структуры в клеточной среде, которые происходят в широком диапазоне временных масштабов (от пс до с) и с различными амплитудами, включая движения основной и боковых цепей или даже перегруппировки целых доменов [1].

Фокусируясь на взаимодействии белков с малыми молекулами, нельзя не отметить эволюцию взглядов на процесс связывания лиганда с рецептором с точки зрения подвижности последнего.

Концепция «замок-ключ», предложенная Фишером в 1984 году в попытках объяснить явление субстратной селективности ферментов, является самой простой моделью и предполагает, что белки распознают «свои» лиганды за счет взаимодополняемости молекулярных поверхностей (форм) активных сайтов и малых молекул [2].

Однако при проведении сравнения между апо- и холо-формами ферментов можно заметить порой существенные различия, что свидетельствует об изменении структуры макромолекулы при формировании белок-лигандных комплексов. С учетом того, что и лиганд как правило характеризуется динамическим поведением, можно говорить о взаимной «подгонке» конформаций партнеров по связыванию друг под друга. Объясняющая такое поведение белков **модель индуцированного соответствия** была описана Кошландом в 1958 г. [3].

Наконец, в конце XX века появилась на свет концепция, предполагающая, что белки существуют в виде ансамбля, включающего несколько конформационных состояний, между которыми колеблется белок. Причем вероятности занятия каждой из конформаций определяются их относительными свободными энергиями, в соответствии с распределением Больцмана. Встреча с партнером по связыванию смещает распределение в сторону одного или нескольких вполне определенных состояний, приводя к образованию конечной структуры комплекса. Эта концепция, названная **моделью выбора конформаций** или **моделью сдвига популяции** (англ. conformational-selection (population shift) model), постулирует, что холо-структура существует еще до события связывания [4, 5].

Схематично три описанных выше взгляда на процесс связывания лигандов с макромолекулами показаны на рис. 9.1.

Ключевой задачей вычислительных методов в контексте структурно-ориентированного дизайна лекарств является прогнозирование правильного способа связывания малой молекулы в сайте биомишени. Большая часть алгоритмов стыковки на сегодняшний день реализует подход «жесткий рецептор – гибкий лиганд», в котором не учитывается конформационная подвижность белка [6]. Если структура рецептора *существенно не меняется* при связывании лиганда, то такой подход к стыковке имеет место быть, однако в противном случае приходится учитывать индуцированную малой молекулой подвижность белка.

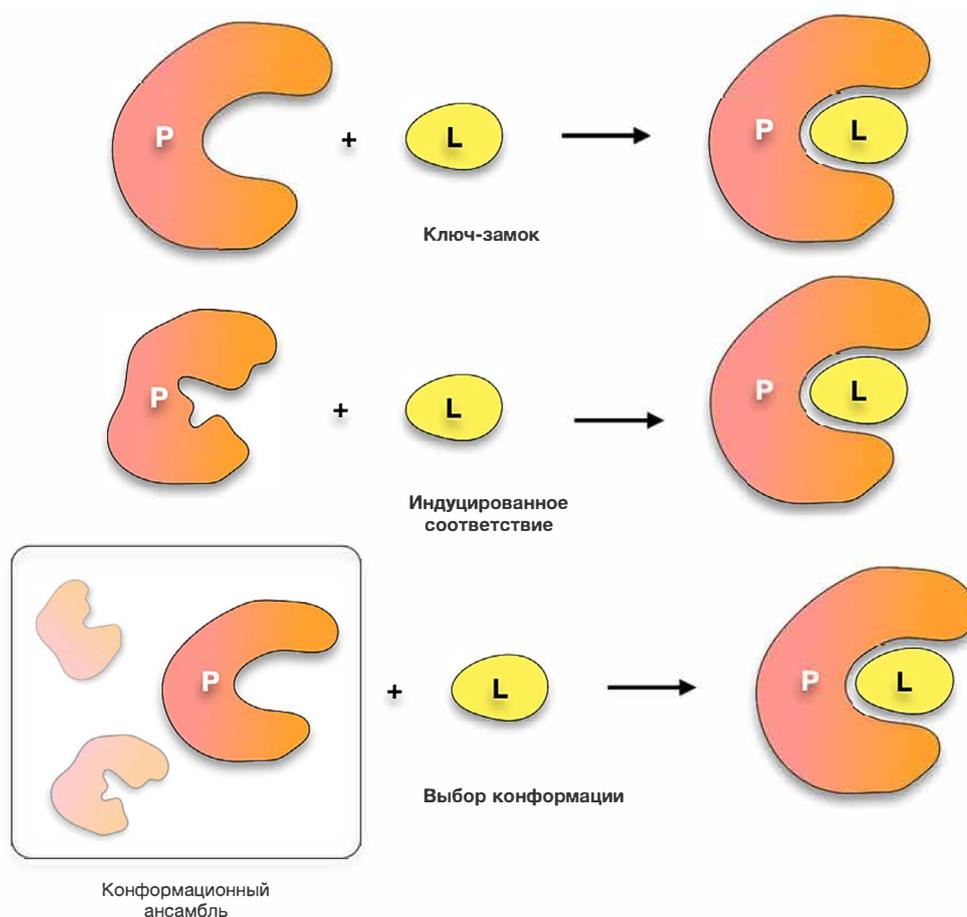


Рис. 9.1. Модели связывания лигандов с рецепторами.
Адаптировано из Comprehensive Medicinal Chemistry III [1]

В некоторых случаях можно применить подход уменьшения ван-дер-ваальсовых радиусов атомов белка/лиганда или даже удаления некоторых боковых цепей в активном сайте, таким образом «освобождая» место для лиганда в связывающем кармане и снижая вероятность возникновения нежелательных столкновений (*англ.* clashes) [7]. Выбрать остатки для масштабирования или обрезки можно с помощью значений В-фактора, отражающего зависимость от температуры атомные вибрации, т. е. являющегося мерой локальной конформационной подвижности белка (рис. 9.2).

Однако такой подход, очевидно, не позволяет получить конкретного понимания о белок-лигандных взаимодействиях, которые могут возникнуть в результате взаимных конформационных изменений малой молекулы и мишени.

Решить проблему можно с помощью метода индуцированной стыковки (*англ.* IFD, Induced-Fit Docking), суть которого заключается в явном учете как подвижности лиганда, так и подвижности белка при предсказании и оценке поз.

Schrödinger Maestro включает модуль Induced-Fit Docking, позволяющий проводить IF-стыковку.

Оригинальный алгоритм (Standard sampling) производит удаление некоторых боковых цепей (*англ.* trimming side chains) белка и смягчение потенциалов Ван-дер-Ваальса, таким образом предоставляя лиганду больший «простор» для размещения в сайте связывания. Выбор аминокислот для обрезки может происходить как автоматически (в зависимости от значений В-фактора, площади поверхности, доступной для сольватации, наличия солевых мостиков), так и в ручном режиме.

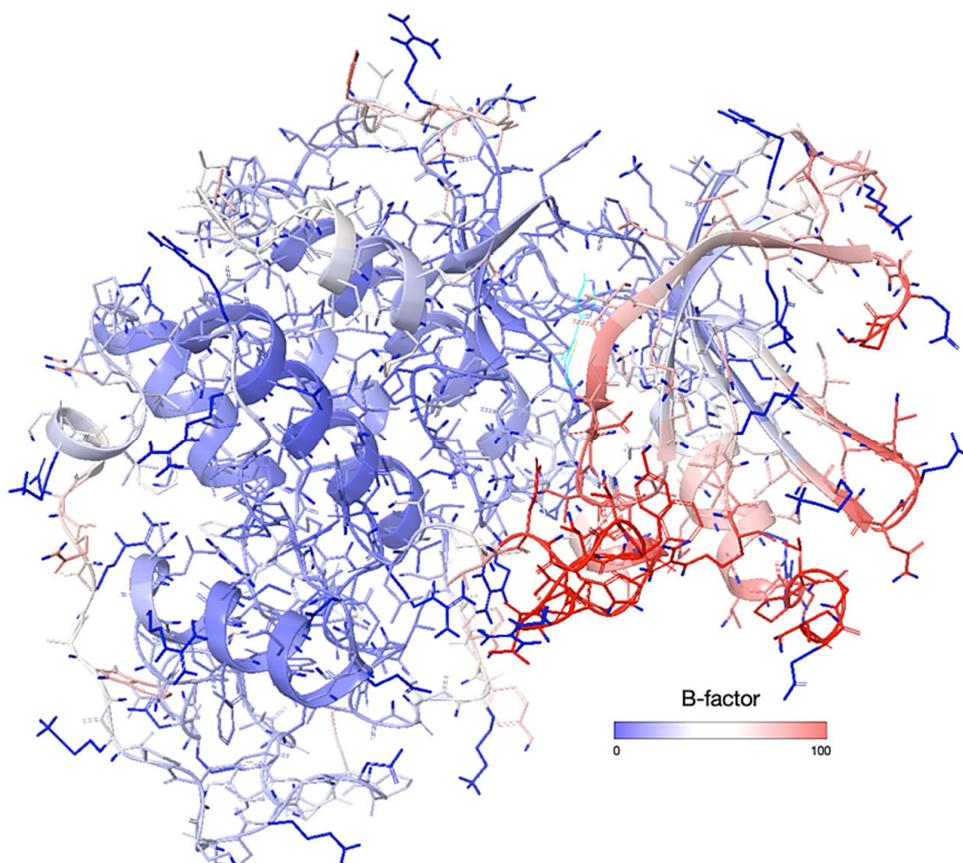


Рис. 9.2. Структура CDK2 (PDB ID: 1PXJ [3]): цвета петель, листов и спиралей, а также атомов белка отражают подвижность (красный – высокий B-фактор и высокая подвижность; синий – низкая)

Затем следует стыковка малой молекулы в «расширенный» сайт с помощью алгоритма Glide [8] с сохранением максимум 20 поз на лиганд (80 в режиме Extended sampling), которые проходят через процедуру предсказания конформации боковых цепей [9] (будет предсказана конформация всех остатков, находящихся на расстоянии не более чем 5 Å от лиганда). Дополнительно выполняется минимизация энергии комплексов. Структура рецепторов в каждой позе теперь отражает индуцированное соответствие конформациям лигандов.

На заключительной этапе реализуется повторная стыковка Glide с генерацией поз, находящихся в пределах энергетического окна в 30 ккал/моль относительно лучших поз, полученных после минимизации, а также расчет IFD Score:

$$IFD\ Score = 0.05Prime\ Energy + Glide\ GScore\#2,$$

где Prime Energy соответствует энергии комплекса, полученного после предсказания конформации остатков и минимизации, а Glide GScore – значение скоринговой функции Glide, рассчитанное во время второй итерации докинга (рис. 9.3) [10, 11].

Подробнее о применении методов индуцированного докинга в процессе дизайна лекарств вы можете узнать, ознакомившись с работами [12–14].

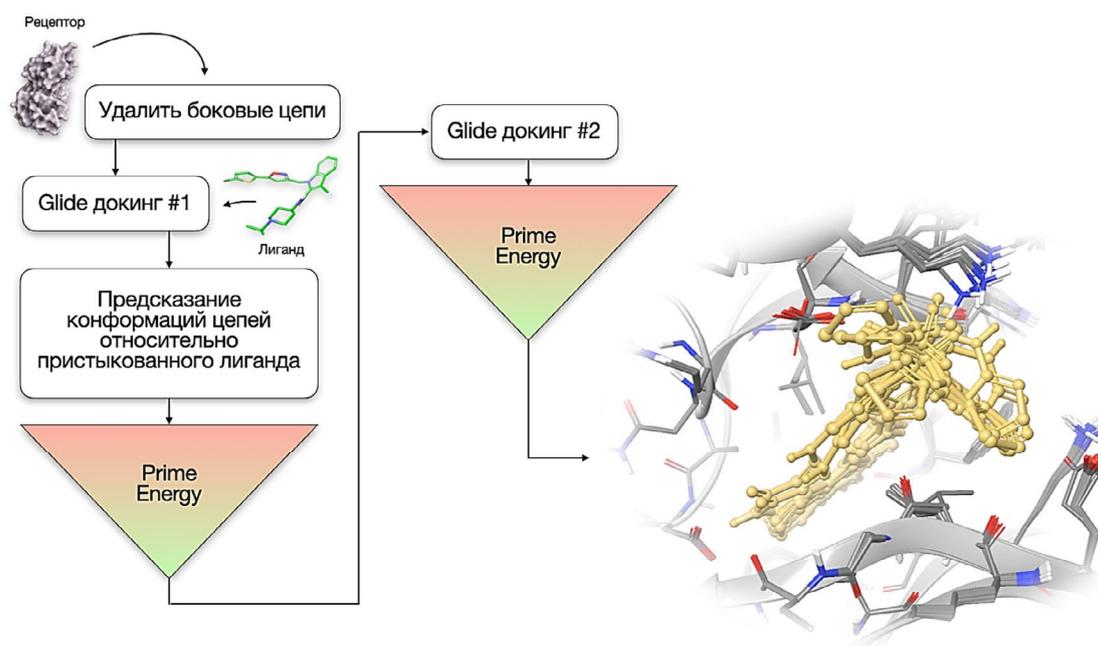


Рис. 9.3. Рабочий процесс Induced-Fit Docking

В этой работе вам будет предложено предсказать позы лигандов в сайтах белков, при этом попытавшись учесть подвижность, опосредованную индуцированным соответствием, как с помощью докинга со смягченными потенциалами, так и с помощью IFD.

Экспериментальная часть

9.1. Создание проекта

В своем личном каталоге создайте папку LW9. Дважды щелкните по ярлыку Maestro на рабочем столе. На верхней панели перейдите на вкладку **File > Change Working Directory**. В открывшемся окне укажите путь к рабочей директории LW9 в своей личной папке.

Затем перейдите **File > Save project as**, задайте проекту имя LW9 в графе **File name**, укажите путь к рабочей директории в графе **Location**.

9.2. Смягчение ван-дер-ваальсовых и электростатических потенциалов

С помощью модуля **Protein Preparation Workflow** загрузите структуру комплекса нитроэфирредуктазы PETN (*англ.* Pentaerythritol tetranitrate) в комплексе с циклогекс-2-ен-1-оном (PDB ID: **1GVQ** [15]), но не производите подготовку.

Отобразите в Workspace элементы вторичной структуры в виде лент с помощью панели **Style**. Нажмите **Color Atoms** и выберите **Atom PDB B factor (Temperature factor)**.

Теперь структура будет раскрашена в соответствии с температурными факторами: синий – низкий B-фактор, низкая подвижность и высокая определенность, красный, соответственно, наоборот (рис. 9.4).

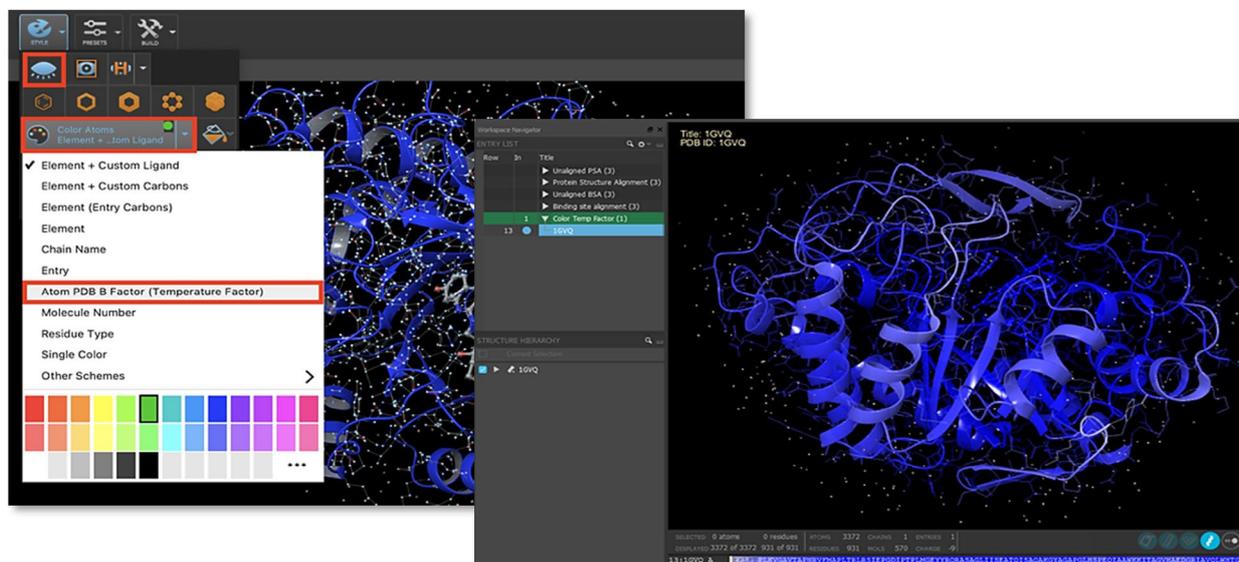


Рис. 9.4. Отображение цветовой кодировки структуры в зависимости от значения B-фактора

В панели **Quick Select** нажмите на ... и выберите пункт **Binding site + Ligands**. В панели **Style** выберите  (**Display only selected atoms**), нажмите **Apply Labels > Edit Custom Label > Add** и выберите **PDB B Factor (Temperature)**. Теперь атомам лиганда и остатков сайта связывания будут присвоены метки с численными значениями температурного фактора. Скройте отображение вторичной структуры для выбранных остатков сайта (рис. 9.5).

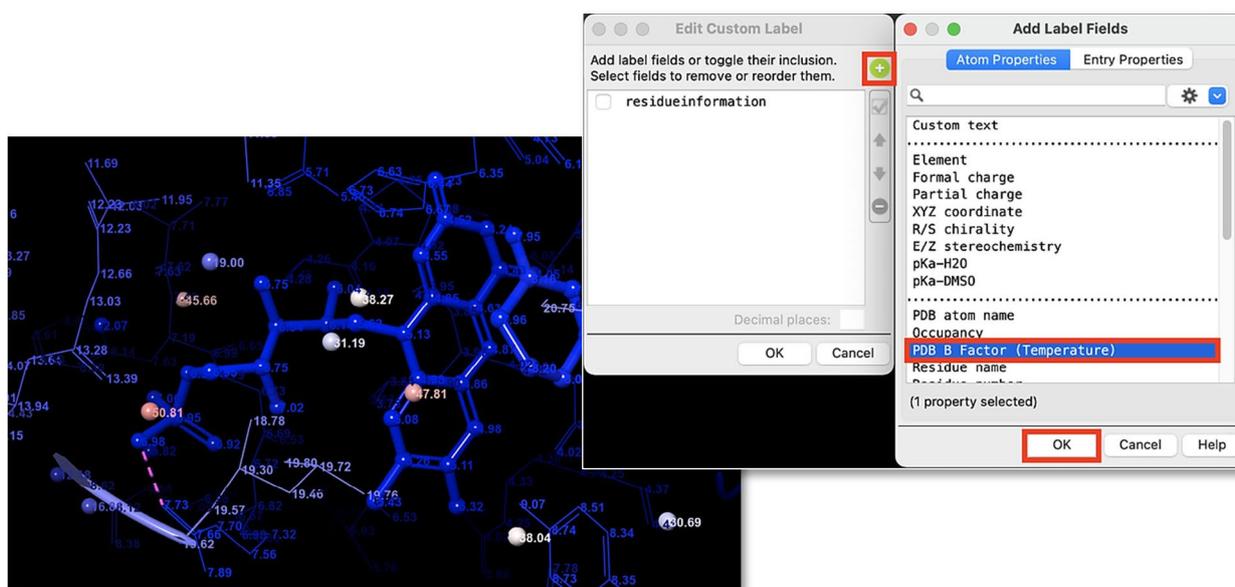


Рис. 9.5. Отображение меток атомов

Отразите в отчете: снимки Workspace, демонстрирующие целую структуру РЕТН редуктазы, а также сайта связывания отдельно в цветовой кодировке согласно значениям

В-фактора. Сделайте вывод о том, какие части белка характеризуются высокой степенью подвижности, а какие – низкой.

Важно: учитывайте, что малоподвижные остатки в ядре белка имеют значение В-фактора, приблизительно равное четырем, а масштабировать потенциалы каталитических остатков крайне нежелательно!

Подготовьте структуры белка с помощью **Protein Preparation Workflow** в соответствии со стандартным протоколом, описанным в лабораторной работе 1. Удалите воду и выполните минимизацию.

Откройте модуль генерации сетки потенциалов, определите область стыковки и циклогексанон в качестве лиганда. Находясь на вкладке **Receptor**, откройте расширенные настройки **Advanced Settings**. Основываясь на значениях В-фактора, выберите остатки, потенциалы которых будут подвергнуты масштабированию (рис. 9.6).

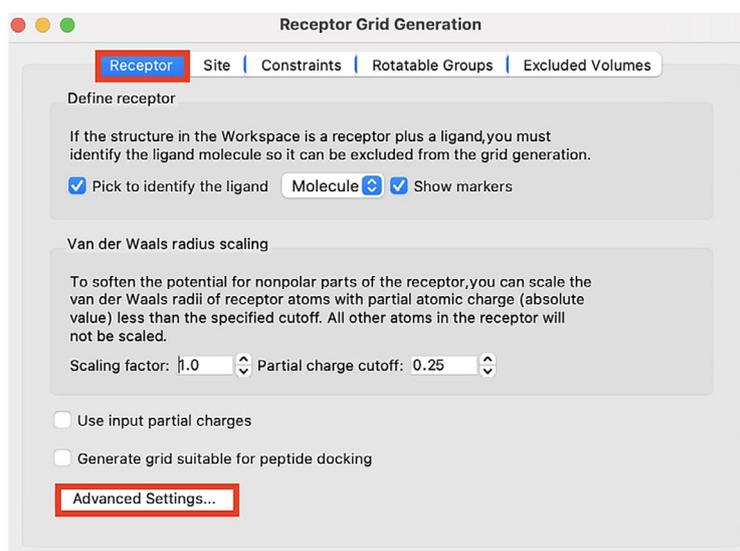


Рис. 9.6. Генерация сетки потенциалов

В разделе **Per-atom scale factors** выберите **Specify for selected atoms**. Установите значение **VDW scale factor** равным «0.8», а **Charge Scale factor** – равным «1». Выберите остатки, для которых будет применено масштабирование. Нажмите **Pick** и выберите **Residues** в выпадающем списке. Выделите в **Workspace** выбранные вами остатки и нажмите **OK**.

Альтернативно, можно не выбирать остатки в **Workspace**, а перечислить их с помощью выражения **ASL**: уберите галочку у пункта **Pick**, нажмите **Select**, а в открывшейся панели **Atom Selection** перейдите на вкладку **Residues**, где выберите **Residue number** и укажите номера интересующих вас остатков. Затем нажмите **Add** и **OK** (рис. 9.7). Выражение **ASL** должно принять вид:

res.num 1, 100, 200, 1000

Смените имя задания на «**Scaled_grid**» и нажмите **Run**. Удалите все настройки масштабирования и запустите расчет для стандартной сетки потенциалов, назвав задание «**Standard_grid**». Общее время расчет займет не более 10 минут.

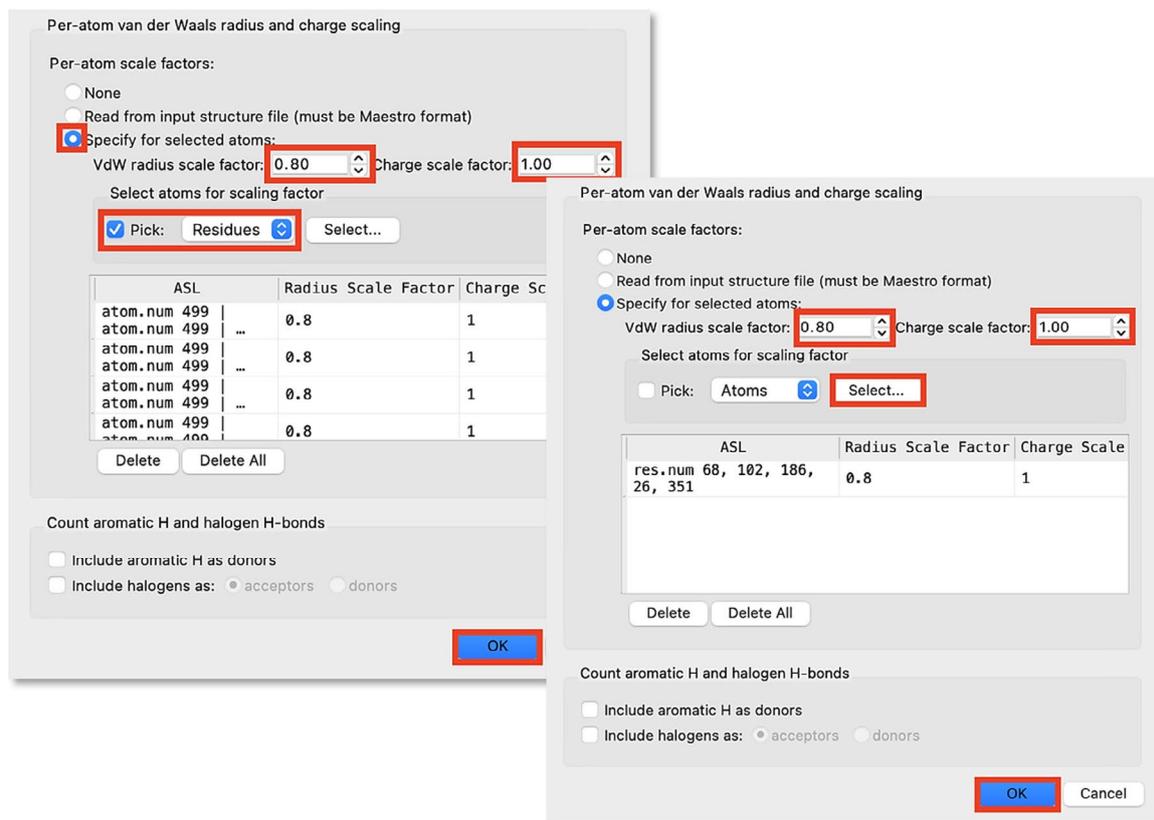


Рис. 9.7. Настройки параметров масштабирования

Импортируйте структуру пикриновой кислоты (**picric_acid.maegz**) и дважды выполните молекулярный докинг с помощью модуля **Ligand Docking**, используя протокол **flexible Glide SP**, указав в качестве грид-боксов созданные на предыдущем шаге сетки. Задания назовите «**SP_f_standard**» и «**SP_f_scaled**» соответственно. Расчет займет не более 15–20 минут.

Для оценки результатов стыковки импортируйте из PDB и подготовьте (в соответствии со стандартным протоколом) структуру комплекса редуктазы PETN с пикриновой кислотой (PDB ID: **1GVS** [16]).

Отразите в отчете:

- 1) наложение структур, полученных в результате двух запусков Glide, и кристаллического комплекса пикриновой кислоты с редуктазой. Лиганды обозначьте разными цветами, укажите значения Glide SP;
- 2) ответ на следующий вопрос: достоверно известно, что редуктаза PETN также способна связывать такие молекулы, как преднизолон и прогестерон. Можно ли будет использовать прием смягчения потенциалов в таком случае?

Заметка: для ответа на второй вопрос необходимо найти в PDB комплексы PETN с данными соединениями, импортировать и подготовить их, а затем выполнить их наложение, используя в качестве референса структуру 1GVS (с помощью модуля **Binding Site Alignment** и/или **Protein Structure Alignment**). Дополнительно можно провести расчет **SiteMap** (режим анализа: **Single binding site**) для нахождения объемов сайтов. Мотивируйте ответ снимками **Workspace**.

9.3. Индуцированный докинг

9.3.1. Импорт и подготовка структур

Импортируйте и подготовьте структуру комплекса CDK2 с ингибитором (PDB IDs: **1PXJ** [8]), выполнив достраивание петель с помощью Prime.

Заметка: данная запись PDB содержит структуру упомянутой выше киназы с небольшим ингибитором (практически фрагментом). Достоверно известно, что оккупировать сайт CDK2 могут молекулы гораздо большего размера. В этой работе вам предстоит предсказать позу связывания лиганда CDK2, представленного на рис. 9.8 [18].

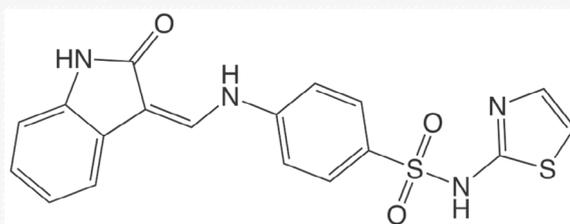


Рис. 9.8. Структура 2-оксииндол-содержащего ингибитора CDK2

С помощью модулей **2D-Sketcher** и **LigPrep** создайте и подготовьте структуру изображенного выше лиганда.

9.3.2. Выполнение индуцированной стыковки

Включите в **Workspace** структуру подготовленного комплекса **1KE8**, а в **Entry List** выделите подготовленный лиганд. Откройте модуль индуцированной стыковки: **Tasks > Browse > Receptor-Based Virtual Screening > Induced Fit Docking**.

На вкладке **Receptor** в разделе **Box Center** определите область стыковки: выберите пункт **Centroid of Workspace ligand**, поставьте галочку напротив **Pick**, а затем нажмите на любой атом лиганда в **Workspace**. Границы грид-бокса будут отображены в **Workspace** аналогично тому, как это происходит при генерации стеки потенциалов в модуле **Receptor Grid Generation**. Выберите **Project Table (1 selected entry)** в разделе **Set Ligand to be docked** (рис. 9.9).

Перейдите на вкладку **Glide Docking** для установки настроек стыковки. Выберите стандартный протокол в верхней части окна. Поставьте галочку напротив пункта **Trim Side Chains**, выберите автоматический режим выбора остатков (**Automatic (based on B-factor)**) (рис. 9.10).

Переименуйте задание как «**InducedFit_CDK2**» и нажмите **Run**. Расчет на одном процессоре занимает около часа. С разрешения преподавателя импортируйте файл с готовыми результатами расчета.

Визуально осмотрите позы. Также вы можете отобразить в **Entry List** значения **IFD Score**, **Glide GScore** и **Prime energy** для лучшего понимания энергетики полученных комплексов.

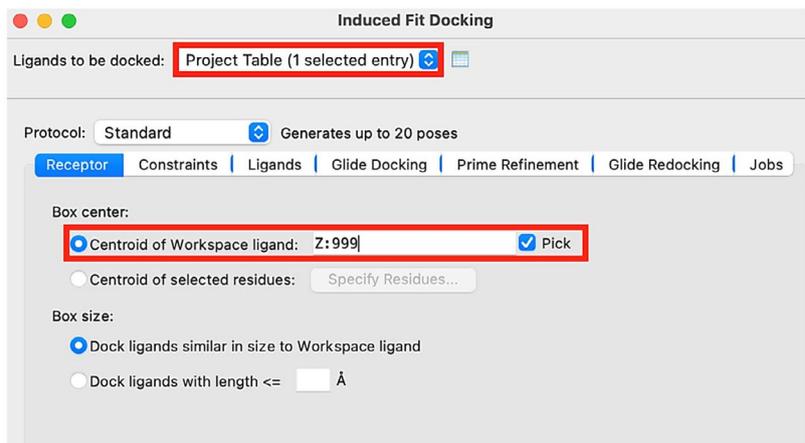


Рис. 9.9. Настройки области стыковки

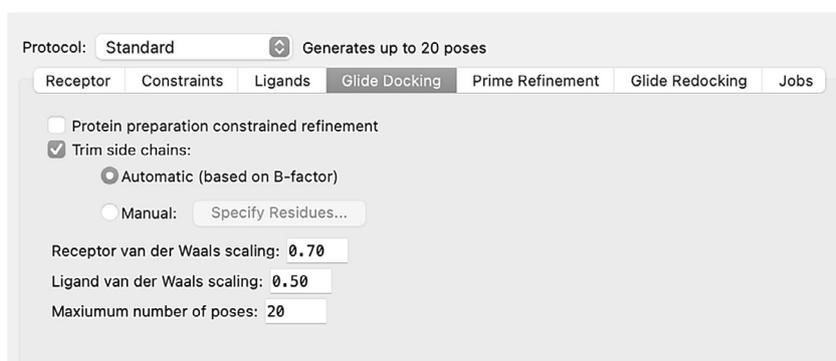


Рис. 9.10. Настройки обрезки цепей

Отразите в отчете:

- 1) ответ на следующий вопрос: удалось ли предсказать правильную позу связывания для лиганда? Импортируйте и подготовьте комплекс CDK2 (PDB ID: **1KE8**), содержащий 2-оксииндольный лиганд. Выполните выравнивание кристаллической и полученных в ходе IFD структур. Также отразите значение скоринга и ранг позы, показывающей лучшее соответствие;
- 2) результат следующего эксперимента: попробуйте выполнить стыковку с использованием приема смягчения потенциалов. Согласно значениям В-фактора, выберите остатки для масштабирования VDW радиусов на 30%. Приведите снимок **Workspace** с наложением полученной структуры и белкового комплекса из кристалла.

Вопросы к защите

1. Основные положения теории об индуцированном соответствии лиганда и рецептора.
2. Объясните физический смысл В-фактора.
3. В чем заключаются достоинства алгоритма индуцированного докинга, представленного в данной работе?
4. В чем заключаются недостатки и ограничения алгоритма индуцированного докинга, представленного в данной работе?

5. Какие еще есть методы молекулярного моделирования, позволяющие учитывать подвижность рецептора?
6. Из каких компонентов складывается значение скоринговой функции IFD Score?
7. Какие части белка наиболее сильно подвержены конформационным изменениям?
8. Опишите алгоритм, с помощью которого осуществляется индуцированная стыковка в ПО Schrödinger.
9. На что нужно обращать внимание при выборе боковых цепей для удаления?
10. Какое место занимают подходы индуцированной стыковки в процессе открытия лекарств?

Рекомендуемая литература

1. Comprehensive Medicinal Chemistry III / Eds.: S. Chackalamannil [et al.]. — 3rd Ed. — Amsterdam : Elsevier, 2017. — [4536] p. — ISBN 978-0-12-803200-8.
2. Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme // Berichte der deutschen chemischen Gesellschaft. — 1894. — Vol. 27, № 3. — P. 2985–2993.
3. Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis // Proceedings of the National Academy of Sciences of the United States of America. — 1958. — Vol. 44, № 2. — P. 98–104.
4. Folding and binding cascades: Dynamic landscapes and population shifts / S. Kumar, B. Ma, C.-J. Tsai [et al.] // Protein Science. — 2000. — Vol. 9, № 1. — P. 10–19.
5. Kumar, S. Factors enhancing protein thermostability / S. Kumar, C.-J. Tsai, R. Nussinov // Protein Engineering. — 2000. — Vol. 13, № 3. — P. 179–191.
6. Molecular Modeling : Basic Principles and Applications / H.-D. Holtje, W. Sippl, D. Rognan, G. Folkers. — 3rd, Revised and Expanded Ed. — Weinheim : Wiley-VCH, 2008. — X, 310 p. — ISBN 978-3-527-31568-0.
7. Carlson, H. A. Accommodating Protein Flexibility in Computational Drug Design / H. A. Carlson, J. A. McCammon // Molecular Pharmacology. — 2000. — Vol. 57. — P. 213–218.
8. Discovery of a Novel Family of CDK Inhibitors with the Program LIDAEUS: Structural Basis for Ligand-Induced Disordering of the Activation Loop / S. Y. Wu, I. McNaie, G. Kontopidis [et al.] // Structure. — 2003. — Vol. 11, № 4. — P. 399–410.
9. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy / R. A. Friesner, J. L. Banks, R. B. Murphy [et al.] // Journal of Medicinal Chemistry. — 2004. — Vol. 47, № 7. — P. 1739–1749.
10. A Hierarchical Approach to All-Atom Protein Loop Prediction / M. P. Jacobson, D. L. Pincus, C. S. Rapp [et al.] // PROTEINS: Structure, Function, and Bioinformatics. — 2004. — Vol. 55, № 2. — P. 351–367.
11. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects / W. Sherman, T. Day, M. P. Jacobson [et al.] // Journal of Medicinal Chemistry. — 2006. — Vol. 49, № 2. — P. 534–553.
12. Sherman, W. Use of an Induced Fit Receptor Structure in Virtual Screening / W. Sherman, H. S. Beard, R. Farid // Chemical Biology & Drug Design. — 2006. — Vol. 67. — P. 83–84.
13. Structure-Based Site of Metabolism (SOM) Prediction of Ligand for CYP3A4 Enzyme: Comparison of Glide XP and Induced Fit Docking (IFD) / D. K. Lokwani,

- A. P. Sarkate, K. S. Karnik [et al.] // *Molecules*. – 2020. – Vol. 25, № 7. – Article number 25071622.
14. Discovery of New Non-Steroidal FXR ligands via a virtual screening workflow based on Phase shape and induced fit docking / J. Fu, P. Si, M. Zheng [et al.] // *Bioorganic & Medicinal Chemistry Letters*. – 2012. – Vol. 22, № 22. – P. 6848–6853.
 15. *Zhong, H.* Induced-fit docking studies of the active and inactive states of protein tyrosine kinases / H. Zhong, L. M. Tran, J. L. Stang // *Journal of Molecular Graphics and Modelling*. – 2009. – Vol. 28, № 4. – P. 336–346.
 16. Kinetic and Structural Basis of Reactivity of Pentaerythritol Tetranitrate Reductase with NADPH, 2-Cyclohexenone, Nitroesters, and Nitroaromatic Explosives / H. Khan, R. J. Harris, T. Barna [et al.] // *Journal of Biological Chemistry*. – 2002. – Vol. 277, № 24. – P. 21906–21912.
 17. Oxindole-Based Inhibitors of Cyclin-Dependent Kinase 2 (CDK2): Design, Synthesis, Enzymatic Activities, and X-ray Crystallographic Analysis / H. N. Bramson, J. Corona, S. T. Davis [et al.] // *Journal of Medicinal Chemistry*. – 2001. – Vol. 44, № 25. – P. 4339–4358.



Лабораторная работа 10

ПРЕДСКАЗАНИЕ ЭНЕРГИИ СВЯЗЫВАНИЯ МЕТОДОМ MM-GBSA

Цель работы: ознакомление с методами предсказания энергии связывания лиганда с белком-мишенью с помощью подходов обобщенной механики Борна.

Вам предстоит:

- 1) подготовить структуру фактора свертывания крови Ха и провести молекулярный докинг библиотеки лигандов;
- 2) произвести расчет свободной энергии связывания на основе полученных поз методом MM-GBSA с помощью модуля Prime;
- 3) проанализировать соответствие рассчитанных и определенных экспериментально значений ΔG_{bind} .

Теоретическое введение

Возможность соотнести структуру белково-лигандных комплексов (полученных, например, с помощью методов молекулярной стыковки) с их аффинностью связывания представляет собой задачу фундаментальной важности [1]. Однако возникает закономерный вопрос: как подойти к оценке свободной энергии связывания лиганда с рецептором? Простой подсчет вкладов взаимодействий может не дать результатов – средство не коррелирует с количеством образующихся водородных связей, площадью поверхности гидрофобных взаимодействий или многими другими простыми терминами [2]. Кроме того, в первом приближении, расчет свободной энергии должен включать не только оценку энтальпийного вклада в связывание (который и обусловлен возникновением нековалентных взаимодействий), но также и энтропийного, описывающего изменение степеней свободы лиганда/рецептора при образовании комплекса и сольватации/десольватации:

$$\Delta G_{bind} = \Delta H - T\Delta S,$$

где ΔH – изменение энтальпии, ΔS – изменение энтропии, T – абсолютная температура.

На первый взгляд кажется, что скоринговые функции (в особенности эмпирические и основанные на силовых полях) при ранжировании поз в докинге производят учет как штрафа за потерю конформационной энтропии лиганда при связывании, так и грубый учет растворителя [3, 4]. Однако оценки алгоритмов стыковки являются лишь оценками и далеко не всегда коррелируют с экспериментальными значениями [5, 6] (при том что позы генерируются относительно правильно).

Довольно логичный подход к оценке ΔG_{bind} может быть представлен следующим выражением:

$$\Delta G_{bind} = E_{tot, PL} - E_{tot, free L} - E_{tot, free P},$$

где $E_{tot, PL}$, $E_{tot, free L}$, $E_{tot, free P}$ – энергии комплекса, свободных лиганда и белка соответственно.

С учетом того, что как лиганд-белковый комплекс, так и его компоненты по отдельности, существуют в водной среде, необходимо производить расчет их энергий с учетом влияния среды.

Подход к такой оценке можно осуществлять путем явного моделирования молекул растворителя (например, с помощью симуляций молекулярной динамики); однако следует помнить о вычислительной дороговизне таких методов [7].

Другим способом является неявное представление растворителя, основанное на замене реальной водной среды, состоящей из дискретных молекул, бесконечным континуумом с определенными зависимыми от расстояний диэлектрическими свойствами [8]. В таком случае выражение для расчета энергии комплекса и свободных белка с лигандом примет вид:

$$E_{tot} = E_{vac} + \Delta G_{solv},$$

где E_{vac} представляет потенциальную энергию молекулы в вакууме (газовой фазе), а ΔG_{solv} определяется как свободная энергия перехода молекулы из вакуума в растворитель, т. е. свободная энергия сольватации.

В то время как вычисление потенциальной энергии в газовой фазе не представляет особых проблем и осуществляется с помощью классических подходов молекулярной механики, вопрос расчета энергии сольватации требует более подробного рассмотрения.

В общем случае ΔG_{solv} можно аппроксимировать как сумму энергии сольватации незаряженной молекулы ($\Delta G_{nonpolar}$) и свободной энергии поляризации (ΔG_{polar} – электростатическая составляющая). неполярный вклад в общую энергию, кодирующий образование полости в среде растворителя, считают пропорциональным площади поверхности молекулы, доступной для растворителя (*англ.* SASA, Solvent-accessible surface area) [9, 10]:

$$\Delta G_{solv} = \Delta G_{nonpolar} + \Delta G_{polar},$$

$$\Delta G_{nonpolar} = \sigma SASA,$$

где σ – коэффициент пропорциональности.

Расчет электростатической составляющей можно провести, основываясь на уравнении Пуассона-Больцмана [11] (*англ.* PB, Poisson-Boltzmann), однако его решение для макромолекулярных систем требует больших вычислительных затрат, а потому используется некоторое число аппроксимаций.

Одним из приближений является обобщенная модель Борна (*англ.* GB, Generalized Born), которая подразумевает рассмотрение молекулы растворенного вещества в качестве совокупности сфер со значением внутренней диэлектрической проницаемости меньше, чем диэлектрическая проницаемость среды:

$$\Delta G_{polar} = -\frac{1}{2} \left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}} \right) \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j e^{-\frac{r_{ij}^2}{4\alpha_i \alpha_j}}}},$$

где ϵ_{in} и ϵ_{out} – значения диэлектрической проницаемости внутри поверхности сфер и снаружи соответственно, q – частичные заряды на атомах i и j , r_{ij} – расстояние между теми же атомами, а α – эффективные Борновские радиусы атомов, которые могут быть интерпретированы примерно как расстояние от каждого атома до поверхности молекулы, образованной сферами [12].

Расчет эффективных радиусов Борна сильно влияет на точность модели и не будет рассмотрен здесь, но заинтересованный читатель сможет подробнее ознакомиться с этим вопросом в [9, 13].

Сам метод неявного учета сольватации с помощью модели Борна сокращенно именуют ММ/GBSA (*англ.* Molecular Mechanics/Generalized Born surface area, молекулярная механика с обобщенной моделью Борна и площадью поверхности).

Модуль Prime [14, 15], используемый в данной работе, позволяет оценить свободную энергию связывания лиганда как раз с помощью данного подхода. В качестве входной структуры требуется либо экспериментально определенные координаты комплекса, либо поза, полученная в ходе докинга. Модель сольватации VSGB (*англ.* Variable-Dielectric Solvation Generalized Born) позволяет провести оценку энергетики как раз с помощью ММ/GBSA:

$$E_{total} = \sum_{bonds} k_b (r - r_0)^2 + \sum_{angels} k_\theta (\theta - \theta_0)^2 +$$

$$+ \sum_{tors} \frac{V_n}{2} [1 + \cos(n\varphi - \delta)] +$$

$$+ \sum_{electro} \frac{q_i q_j}{r_{ij} \epsilon_{in}} + \sum_{vdW} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + E_{sol} + \sum E_{corr},$$

где \sum_{bonds} – вклад энергии деформации связей, r, r_0 – текущие и равновесные значения межатомных расстояний между непосредственно связанными атомами; \sum_{angles} – вклад энергии искажения валентных углов, θ_0 – текущие и равновесные значения валентных углов; \sum_{tors} – вклад энергии искажения торсионных углов, n – период, ϕ, δ – текущие и равновесные значения торсионных углов; $\sum_{electro}$ – вклад энергии всех попарных электростатических взаимодействий $q_{i,j}, r_{ij}, \epsilon$ – частичные заряды атомов i и j , расстояние между атомами i и j и диэлектрическая проницаемость; \sum_{vdW} – вклад энергии всех попарных ван-дер-ваальсовых взаимодействий A_{ij}, B_{ij} – коэффициенты вкладов отталкивания и притяжения, r_{ij} – расстояние между атомами i и j ; E_{sol} – терм, учитывающий сольватацию с помощью модели GBSA; $\sum E_{corr}$ – термы коррекции.

Последние позволяют точнее оценить такие взаимодействия, как, например, энергию водородных связей или стэкинга (пример для водородной связи – ниже):

$$E_{HB,corr} = \sum_i \sum_j \alpha_i \alpha_j e^{-(r^{HA} - r_0)^2} \cos(\theta^{DHA} - \theta_0),$$

где i и j – тяжелые атомы, участвующие в образовании водородной связи, r_0 – оптимальное расстояние между атомом водорода и атомом-акцептором, (1.94 Å), θ_0 – оптимальный угол связи, (160°), r и θ – текущая длина и угол, а α – коэффициенты, связанные с типом тяжелых атомов, участвующих в образовании водородной связи.

Подробнее о поправочных компонентах и в целом о модели VSGB можно узнать в оригинальной публикации [16].

В заключение стоит отметить, что значения энергии связывания, полученные методом ММ/GBSA, не являются абсолютными и вряд ли будут четко коррелировать с экспериментально определенными [17], однако расчет Prime в большинстве случаев позволяет верно ранжировать лиганды по увеличению сродства и прекрасно подходит для повторного скоринга и обогащения результатов стыковки [18, 19].

Расчет абсолютных значений лучше производить методами возмущений свободной энергии (*англ.* FEP, Free Energy Perturbations) [20–22], которые гораздо лучше соотносятся с экспериментами, но и гораздо более вычислительно затратны.

Экспериментальная часть

10.1. Создание проекта

В своем личном каталоге создайте папку LW10. Дважды щелкните по ярлыку Maestro на рабочем столе. На верхней панели перейдите на вкладку **File > Change Working Directory**. В открывшемся окне укажите путь к рабочей директории LW10 в своей личной папке.

Затем перейдите **File > Save project as**, задайте проекту имя LW10 в графе **File name**, укажите путь к рабочей директории в графе **Location**.

10.2. Проведение виртуального скрининга

10.2.1. Импорт и подготовка структур

Загрузите из PDB и подготовьте структуру комплекса фактора свертывания крови Ха (PDB ID: **2ВОН** [23]). Не удаляйте цепь В и ион Ca²⁺, а затем выполните расчет сетки потенциалов.

Импортируйте файл **leadseries_prepped.maegz**, содержащий подготовленные лиганды, которые будут использоваться для скрининга.

10.2.2. Первичный виртуальный скрининг с использованием Glide SP

Откройте модуль стыковки **Ligand Docking**, укажите **Entry List (selected)** в качестве источника лигандов и путь к файлу грид-бокса (не забудьте выделить группу **leadseries_prepped**). Выполните стыковку в режиме точности **SP flexible**. Расчет займет не более часа при расчете на одном процессоре. Вы можете изменить число доступных CPU и число подзаданий в разделе **Job Settings** для уменьшения времени выполнения задания. Также вы можете использовать файл с готовыми результатами скрининга (с разрешения преподавателя).

Отразите в отчете:

- 1) таблицу с результатами стыковки, в которой приведите структуру соединения, значение Glide GScore, экспериментально определенные значения pK_i [23], а также ранги соединений. Сортировку выполните по возрастанию значения скоринговой функции (табл. 10.1);

Таблица 10.1

Результаты скрининга

Номер соединения	Glide GScore, ккал/моль	Ранг Glide GScore	Exp. pK _i	Exp Affinity, nM/L	Ранг Exp

- 2) начения фактора обогащения EF (10%):

$$EF_{10\%} = \frac{N_{hits,10\%} / N_{total,10\%}}{N_{hits,100\%} / N_{total,100\%}},$$

где $N_{hits,10\%}$ и $N_{hits,100\%}$ — число молекул-хитов в первых 10% рейтинга и общее число хитов в наборе, $N_{total,10\%}$ и $N_{total,100\%}$ — общее число соединений в первых 10% рейтинга и в наборе соответственно.

10.2.3. Оценка с помощью MM-GBSA

Проведем повторную оценку и ранжирование результатов виртуального скрининга с помощью модуля предсказания энергии связывания Prime. Откройте панель **Prime MM-GBSA: Tasks > Browse > Lead Optimization > MM-GBSA**.

В разделе **Structures** выберите **Take complexes from a Maestro Pose Viewer file**, затем нажмите **Browse** и дважды кликните по файлу с результатами докинга. Измените название задания на «**prime_mmgsa_SP**», нажмите **Run**. Расчет займет около 2 часов на одном процессоре. Вы можете изменить число доступных CPU и число подзадач в разделе **Job Settings** для уменьшения времени выполнения задания. Также вы можете использовать файл с готовыми результатами скрининга (с разрешения преподавателя) (рис. 10.1).

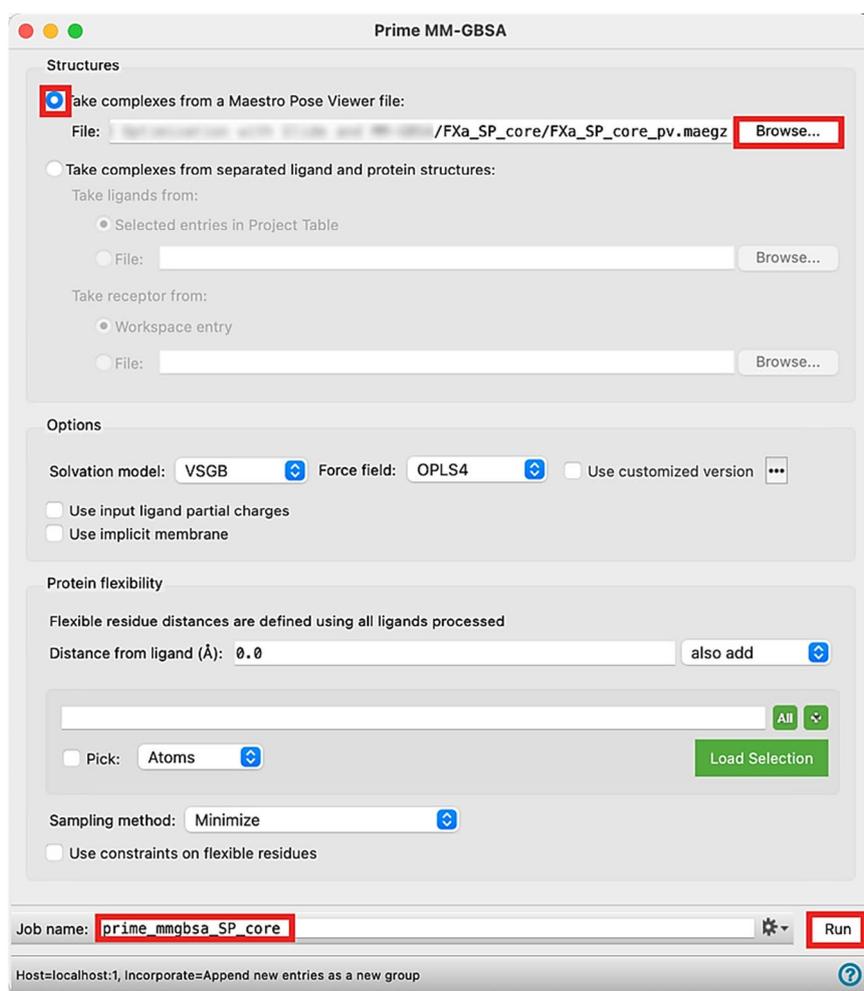


Рис. 10.1. Запуск расчета MM-GBSA Prime

Отразите в отчете:

- 1) дополненную столбцами со значениями MM-GBSA dG_{bind} и рангами относительно предсказанных энергий связывания табл. 10.1. Сортировку выполните по возрастанию dG_{bind} ;
- 2) значения фактора обогащения (10%); удалось ли добиться улучшения обогащения?
- 3) диаграмму рассеяния экспериментальных значений сродства (pK_i) и MM-GBSA dG_{bind} . Для этого воспользуйтесь модулем построения графиков в **Project Table**. Выбрав группу, содержащую результаты расчета Prime, откройте **Project Table**, затем **Plots > New Scatter Plot** (рис. 10.2).

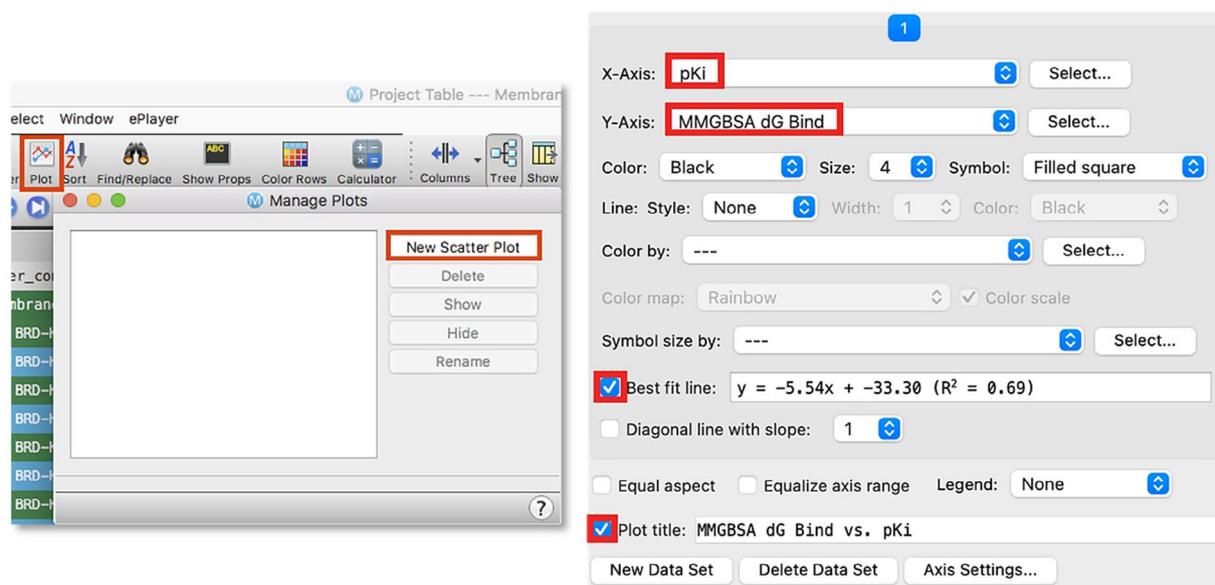


Рис. 10.2. Модуль построения графиков и диаграмм

Выберите необходимые свойства для отображения на оси X и Y. Не забудьте поставить галочку у пункта **Best fit line**. Укажите уравнение прямой и R^2 в отчете.

Вопросы к защите

1. Чем опосредован энтальпийный вклад в свободную энергию связывания лиганда?
2. Чем опосредован энтропийный вклад в свободную энергию связывания лиганда?
3. В чем заключаются преимущества и недостатки явного моделирования растворителя?
4. Из каких компонентов складывается энергия сольватации лиганда? Поясните смысл каждого из них.
5. Изложите суть подхода обобщения Борна при оценке вклада поляризации в энергию сольватации.
6. Что такое эффективный радиус Борна?
7. Почему значения большей части эмпирических оценочных функций не позволяют верно предсказать энергию связывания, хотя термины учета сольватации и энтропийные штрафы включены в уравнения, а коэффициенты вкладов взаимодействий получены методами регрессии на выборках комплексов с экспериментально определенными значениями аффинности?
8. Дайте определение понятию «фактор обогащения» и приведите формулу для его расчета.
9. К какому типу оценочных функций можно отнести функцию, осуществляющую расчет свободной энергии комплекса «белок – лиганд» в Prime?

Рекомендуемая литература

1. Comprehensive Medicinal Chemistry III / Eds.: S. Chackalamannil [et al.]. – 3rd Ed. – Amsterdam : Elsevier, 2017. – [4536] p. – ISBN 978-0-12-803200-8.

2. The Handbook of Medicinal Chemistry : Principles and Practice / Edited by S. E. Ward, A. Davis. – 2nd Ed. – Croydon : Royal Society of Chemistry, 2023. – LV, 994 p. – ISBN 978-1-78801-898-2.
3. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes / R. A. Friesner, R. B. Murphy, M. P. Repasky [et al.] // Journal of Medicinal Chemistry. – 2006. – Vol. 49, № 21. – P. 6177–6196.
4. *Pantsar, T.* Binding Affinity via Docking: Fact and Fiction / T. Pantsar, A. Poso // Molecules. – 2018. – Vol. 23, № 8. – Article number 23081899.
5. Comparative Assessment of Scoring Functions on a Diverse Test Set / T. Cheng, X. Li, Y. Li [et al.] // Journal of Chemical Information and Modeling. – 2009. – Vol. 49, № 4. – P. 1079–1093.
6. A Critical Assessment of Docking Programs and Scoring Functions / G. L. Warren, C. W. Andrews, A.-M. Capelli [et al.] // Journal of Medicinal Chemistry. – 2006. – Vol. 49, № 20. – P. 5912–5931.
7. Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents / J. Zhang, H. Zhang, T. Wu [et al.] // Journal of Chemical Theory and Computation. – 2017. – Vol. 13, № 3. – P. 1034–1043.
8. *Kleinjung, J.* Design and application of implicit solvent models in biomolecular simulations / J. Kleinjung, F. Fraternali // Current Opinion in Structural Biology. – 2014. – Vol. 25. – P. 126–134.
9. *Onufriev, A. V.* Generalized Born Implicit Solvent Models for Biomolecules / A. V. Onufriev, D. A. Case // Annual Review of Biophysics. – 2019. – Vol. 48. – P. 275–296.
10. *Gallicchio, E.* AGBNP: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-Resolution Modeling / E. Gallicchio, R. M. Levy // Journal of Computational Chemistry. – 2004. – Vol. 25, № 4. – P. 479–499.
11. *Fogolari, F.* The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology / F. Fogolari, A. Brigo, H. Molinari // Journal of Molecular Recognition. – 2002. – Vol. 15, № 6. – P. 377–392.
12. *Feig, M.* Implicit solvation based on generalized Born theory in different dielectric environments / M. Feig, W. Im, C. L. Brooks III // The Journal of Chemical Physics. – 2004. – Vol. 120, № 2. – P. 903–911.
13. *Onufriev, A.* Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect / A. Onufriev, D. A. Case, D. Bashford // Journal of Computational Chemistry. – 2002. – Vol. 23, № 14. – P. 1297–1304.
14. Schrödinger Release 2023-4: Prime. – New York : Schrödinger, 2023.
15. A Hierarchical Approach to All-Atom Protein Loop Prediction / M. P. Jacobson, D. L. Pincus, C. S. Rapp [et al.] // PROTEINS: Structure, Function, and Bioinformatics. – 2004. – Vol. 55, № 2. – P. 351–367.
16. The VSGB 2.0 model: A Next Generation Energy Model for High Resolution Protein Structure Modeling / J. Li, R. Abel, K. Zhu [et al.] // PROTEINS: Structure, Function, and Bioinformatics. – 2011. – Vol. 79, № 10. – P. 2794–2812.
17. Can I relate MM-GBSA energies to binding affinity? // Schrödinger : [сайт]. – URL: www.schrodinger.com/kb/1647/ (дата обращения: 03.12.2024).
18. *Zhang, X.* A Comprehensive Docking and MM/GBSA Rescoring Study of Ligand Recognition upon Binding Antithrombin / X. Zhang, H. Perez-Sanchez, F. C. Lightstone // Current Topics in Medicinal Chemistry. – 2017. – Vol. 17, № 14. – P. 1631–1639.

19. *Mulakala, C.* Could MM-GBSA be accurate enough for calculation of absolute protein/ligand binding free energies? /C. Mulakala, V.N. Viswanadhan // *Journal of Molecular Graphics and Modelling.* – 2013. – Vol. 46. – P. 41–51.
20. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field /L. Wang, Y. Wu, Y. Deng [et al.] // *Journal of American Chemical Society.* – 2015. – Vol. 137, № 7. – P. 2695–2703.
21. Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery /L. Wang, Y. Deng, Y. Wu [et al.] // *Journal of Chemical Theory and Computation.* – 2017. – Vol. 13, № 1. – P. 42–54.
22. *Wang, L.* Protein-Ligand Binding Free Energy Calculations with FEP /L. Wang, J. Chambers, R. Abel // *Methods in Molecular Biology.* – 2019. Vol. 2022. – P. 201–232.
23. Probing the Subpockets of Factor Xa Reveals Two Binding Modes for Inhibitors Based on a 2-Carboxyindole Scaffold: A Study Combining Structure-Activity Relationship and X-ray Crystallography /M. Nazaré, D. W. Will, H. Matter [et al.] // *Journal of Medicinal Chemistry.* – 2005. – Vol. 48, № 14. – P. 4511–4525.

Приложение А

Таблица

Название, трехбуквенные и однобуквенные коды 20 протеиногенных аминокислот, а также аннотирование типов взаимодействий, которые могут образовывать их боковые цепи (количество «+» показывает относительную силу того или иного взаимодействия)

Название	Трехбуквенный код	Однобуквенный код	Гидрофобные взаимодействия	Водородные связи	Солевой мостик
Аланин	Ala	A	+	–	–
Аргинин	Arg	R	+	+++	+++
Аспаргин	Asn	N	–	+++	–
Аспаргиновая кислота	Asp	D	–	++	+++
Цистеин	Cys	C	+	+	+
Глутамин	Gln	Q	+	++	–
Глутаминовая кислота	Glu	E	+	++	+++
Глицин	Gly	G	+	–	–
Гистидин	His	H	–	+++	+
Лизин	Lys	K	++	+++	–
Лейцин	Leu	F	+++	–	–
Изолейцин	Ile	I	+++	–	–
Метионин	Met	M	++	–	–
Фенилаланин	Phe	F	+++	–	–
Пролин	Pro	P	++	–	–
Серин	Ser	S	+	+	–
Треонин	Thr	T	++	+	–

Окончание таблицы

Название	Трехбуквенный код	Однбуквенный код	Гидрофобные взаимодействия	Водородные связи	Солевой мостик
Триптофан	Trp	W	++	+	–
Тирозин	Tyr	T	++	+	–
Валин	Val	V	++	–	–

Приложение Б

Образец оформления титульного листа к бланку отчета

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Институт машиностроения, химии и энергетики

(Наименование института)

Центр медицинской химии

(Наименование кафедры, центра, департамента)

ОТЧЕТ

о лабораторной работе № __
«Название лабораторной работы»

по дисциплине

«Основы молекулярного моделирования биологически активных веществ»

Выполнил:

Студент группы

Шифр группы

Фамилия И. О.

Проверил:

Преподаватель

Личная подпись

(ученая степень (при наличии), ученое звание (при наличии),

Фамилия И. О.)

Тольятти 20__

Приложение В

Образец оформления бланка отчета

Цель работы: *ознакомление с процессами анализа молекулярной динамики, освоение основных подходов к проведению молекулярного докинга и т. д.*

Ход работы:

1. *В личном каталоге создана новая рабочая директория «working_directory» и проект «working_project.prj»...*
2. *Установлена переменная среды SCHRODINGER=/opt/schrodinger...*
3. *Загружен и подготовлен комплекс мишени интереса с ингибитором (PDB ID: ...) (добавлены атомы водорода, с помощью алгоритма Epik учтено состояние протонирования при pH = ...)*
4. ...

Используемые модули:

Protein Preparation Workflow – подготовка белков...

Receptor Grid Generation – создание сетки потенциалов...

Содержательная часть отчета

Ответы на поставленные вопросы, графики и диаграммы, рисунки и таблицы.

Выводы:

В ходе выполнения лабораторной работы...

Содержание

Введение	3
Лабораторная работа 1	
Основы работы в Maestro. Подготовка структур лигандов и белка	4
Теоретическое введение	5
Экспериментальная часть	13
Лабораторная работа 2	
Конформационный анализ малых молекул	32
Теоретическое введение	33
Экспериментальная часть	38
Лабораторная работа 3	
Фармакофорный поиск	49
Теоретическое введение	50
Экспериментальная часть	57
Лабораторная работа 4	
Молекулярный докинг	77
Теоретическое введение	78
Экспериментальная часть	81
Лабораторная работа 5	
Идентификация сайта связывания	89
Теоретическое введение	90
Экспериментальная часть	91
Лабораторная работа 6	
Создание и фильтрация виртуальных библиотек соединений	97
Теоретическое введение	98
Экспериментальная часть	105
Лабораторная работа 7	
Оптимизация взаимодействий лиганда с белком	113
Теоретическое введение	114
Экспериментальная часть	117

Лабораторная работа 8

Ковалентный докинг	124
Теоретическое введение	125
Экспериментальная часть	129

Лабораторная работа 9

Учет конформационной подвижности рецептора в молекулярном докинге	140
Теоретическое введение	141
Экспериментальная часть	144

Лабораторная работа 10

Предсказание энергии связывания методом MM-GBSA	152
Теоретическое введение	153
Экспериментальная часть	155

Приложение А.....	161
-------------------	-----

Приложение Б.....	163
-------------------	-----

Приложение В.....	164
-------------------	-----

Для заметок

A large, empty rectangular area with a light blue background and a decorative border, intended for notes. The border consists of a thin blue line followed by a hatched pattern. The corners of the inner area are rounded.

Учебное издание

Кудинов Андрей Юрьевич,
Бунев Александр Сиясатович

ОСНОВЫ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ БИОЛОГИЧЕСКИ АКТИВНЫХ ВЕЩЕСТВ

Лабораторный практикум

Редактор *Т.В. Антонова*
Компьютерная верстка: *Е.В. Веселова*
Дизайн макета и обложки: *Е.В. Веселова*

Подписано в печать 29.01.2026. Формат 60×84/8.
Печать оперативная. Усл. п. л. 19,53.
Тираж 100 экз. Заказ № 1-31-25.

Издательство Тольяттинского государственного университета
445020, г. Тольятти, ул. Белорусская, 14,
тел. 8 (8482) 44-91-47, www.tltsu.ru