

МОЛЕКУЛЯРНЫЙ ДОКИНГ

Руководство пользователя

Редакция 2024 года

Молекулярный докинг – это позиционирование низкомолекулярного соединения (лиганда) в активный центр белка-мишени. Оптимизация положения лиганда внутри активного центра производится по результатам оценки энергии взаимодействия лиганда с макромолекулой. Вычисление энергии производится исходя из заранее приготовленной сетки потенциалов.

Лиганды, закристаллизованные вместе с белком в активном центре, называются нативными. Их координаты приведены в PDB-файле белка с пометкой HETATM (следует обратить внимание, что с пометкой HETATM в PDB-файле представлены также другие низкомолекулярные соединения, например, молекулы растворителя, воды). Докинг нативных лигандов позволяет определить, насколько точно программа позиционирует данный лиганд, сравнивая его координаты после докинга с координатами, представленными в структуре комплекса белок+лиганд.

Содержание

1. Скачивание молекулярных структур.....	2
2. Подготовка лиганда.....	4
3. Подготовка белка	8
4. Определение активного центра белка.....	11
5. Построение сетки	11
6. Докинг	12
7. Анализ результатов докинга.....	12
8. Просмотр результатов докинга	14
9. Сравнение скоринг-функции с экспериментальными данными	15
Приложение А. Протонирование программой Marvin Suite	16
Приложение В. Восстановление структуры белка	17
Приложение С. Сравнение с экспериментом с помощью $\log(K_d)$	18

1. Скачивание молекулярных структур

На сайте PDB (база данных Protein Data Bank <https://www.rcsb.org>) доступны для свободного скачивания данные о трехмерной структуре нативного комплекса лиганд-белок, полученные из экспериментальных данных ядерно-магнитного резонанса или рентген-кристаллографии.

- Поиск комплексов осуществляется по четырехбуквенному идентификатору, который вводится в указанное ниже поле.



- Помимо четырехбуквенного идентификатора в строку поиска можно ввести ключевое слово – название белка или лиганда, которые нужно найти.

1C5W Номер структуры в PDB

STRUCTURAL BASIS FOR SELECTIVITY OF A SMALL MOLECULE SUB-MICROMOLAR INHIBITOR OF UROKINASE TYPE PLASMINOGEN ACTIVATOR

DOI: 10.2210/pdb1c5w/pdb

Classification BLOOD CLOTTING

Deposited: 1999-12-22 Released: 2000-12-22

Deposition author(s) Katz, B.A., Mackman, R., Luong, C., Radika, K., Martelli, A., S.H., Wong, L.

Organism Homo sapiens Организм, из которого был выделен белок

Expression System: Pichia pastoris, Pichia pastoris

Mutation(s): 1

Structural Biology Knowledgebase: 1C5W (1 model >23 annotations) [SBI@rcsb.org](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.94 Å < Разрешение (в ангстремах)

R-Value Free: 0.211

R-Value Work: 0.177

wwPDB Validation

Metric	Percentile Ranks	Value
Clashscore		6
Ramachandran outliers		2.4%
Sidechain outliers		4.0%
RSRZ outliers		2.3%

Literature

Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator.

- При выборе структуры белка-мишени следует обратить внимание на разрешение структуры - Resolution: от 1 до 2 Å - хорошее разрешение; от 2 до 3 Å - удовлетворительное; свыше 3 Å - плохое.
- В шапке также приведено название организма, из которого был выделен данный белок (**Organism**).

- В секциях, приведенных ниже, дано более подробное описание структур молекул входящих в комплекс:

Macromolecules

Classification: [BLOOD CLOTTING](#) Sequence Display for IC5W

Total Structure Weight: 32014.25

Macromolecule Entities Toggle Protein Feature View

Molecule	Chains	Length	Organism	Details
PROTEIN (UROKINASE-TYPE PLASMINOGEN ACTIVATOR)	A	23	Homo sapiens	EC#: 3.4.21.73 IU BMB Fragment: SHORT CHAIN PLAU Get View

Protein Feature View - UniProtKB AC: [P00749](#) UniProt Find similar proteins by: [Sequence](#) | [Structure](#)

P00749 - UROK_HUMAN - Urokinase-type plasminogen activator

Molecule	Chains	Length	Organism	Details
PROTEIN (UROKINASE-TYPE PLASMINOGEN ACTIVATOR)	B	253	Homo sapiens	EC#: 3.4.21.73 IU BMB Fragment: CATALYTIC DOMAIN Mutation: N145A PLAU Get View

Small Molecules

Ligands 2 Unique

ID	Chains	Name / Formula / InChI Key	2D Diagram & Interactions	3D Interactions
ESI Query on ESI Download SDF File Download CCD File	B	4-iodobenzo[b]thiophene-2-carboxamide C ₉ H ₈ I N ₂ S YERQOXAYAFWFEJ-UHFFFAOYSA-D		Ligand Explorer Binding Pocket (JSmol) Electron Density (JSmol)
FLC Query on FLC Download SDF File Download CCD File	B	CITRATE ANION C ₆ H ₅ O ₇ KRKNYBCHXYNG-OX-UHFFFAOYSA-K		Ligand Explorer Binding Pocket (JSmol) Electron Density (JSmol)

- Секция **Macromolecules** содержит: названия макромолекул (белков, ДНК и РНК) – Molecule; буквы цепей, входящих в эти макромолекулы – Chains; длины цепей – Length (количество аминокислотных остатков); организм, из которого были экспрессированы молекулы – Organism; и более подробное описание цепей и названия доменов белка, в которые они входят – Details.
- Секция **Small Molecules** содержит описания лигандов, входящих в состав комплекса, в том числе ингибиторов, молекул растворителя и других молекул, оставшихся в комплексе при кристаллизации: ID –

трехбуквенное обозначение лиганда (в PDB-файле это обозначение заменяет название аминокислотного остатка в секции HETATM); Chains – цепь, к которой принадлежит лиганд (обычно, это ближайшая цепь макромолекулы); Name – полное название лиганда в соответствии с химической номенклатурой; 2D-Diagram and Interactions – структура лиганда и описание взаимодействия его химических групп с аминокислотными остатками белка.

- Если в комплексе присутствуют лиганды с измеренными экспериментальными активностями, то ниже можно найти секцию **External Ligand Annotations**. В ней содержатся константы ингибирования K_i (или концентрация полуингибирования IC_{50}) белка данным лигандом.

External Ligand Annotations	
ID	Binding Affinity (Sequence Identity %)
ESI	IC50: 210 - 320 nM (99) <small>BindingDB</small> Ki: 79.43 - 1700 nM (99 - 100) <small>BindingDB</small>
	N/A in BindingMoad N/A in PDBbind

- Для скачивания структуры на локальный компьютер необходимо выбрать Download Files > PDB Format

2. Подготовка лиганда

Из скаченного PDB-файла в текстовом редакторе вырезаем записи, начинающиеся HETATM, которые соответствуют нативному лиганду. Следует обратить внимание, что с пометкой HETATM в PDB-файле представлен не только нативный лиганд. Также пометка HETATM может быть у молекул ионов (ZN, MG, K, NA и т.д.), молекул воды (HOH), и других низкомолекулярных соединений, оставшихся вокруг комплекса при кристаллизации (SO4, NAG, BCT, CIT и т.д.). Все атомы с пометкой HETATM не относящиеся в лиганду мы не используем.

- **Как определить, какая из представленных в разделе HETATM молекул, является нужным нам лигандом – ингибитором?**

Самый надежный способ – обратиться к соответствующей статье, в которой проводился эксперимент (кристаллизация комплекса). Название и авторы этой статьи приведены в шапке PDB-файла и на странице комплекса. Часто узнать название молекулы-ингибитора можно именно из названия статьи. Далее, по названию молекулы можно найти ее трехбуквенное обозначение в секции «Small Molecules» на сайте комплекса. Помимо этого, именно для молекулы ингибитора в секции **External Ligand Annotations** будут приведены константы взаимодействия с белком, и, следовательно, указанный там ID и будет трехбуквенным идентификатором интересующего нас лиганда.

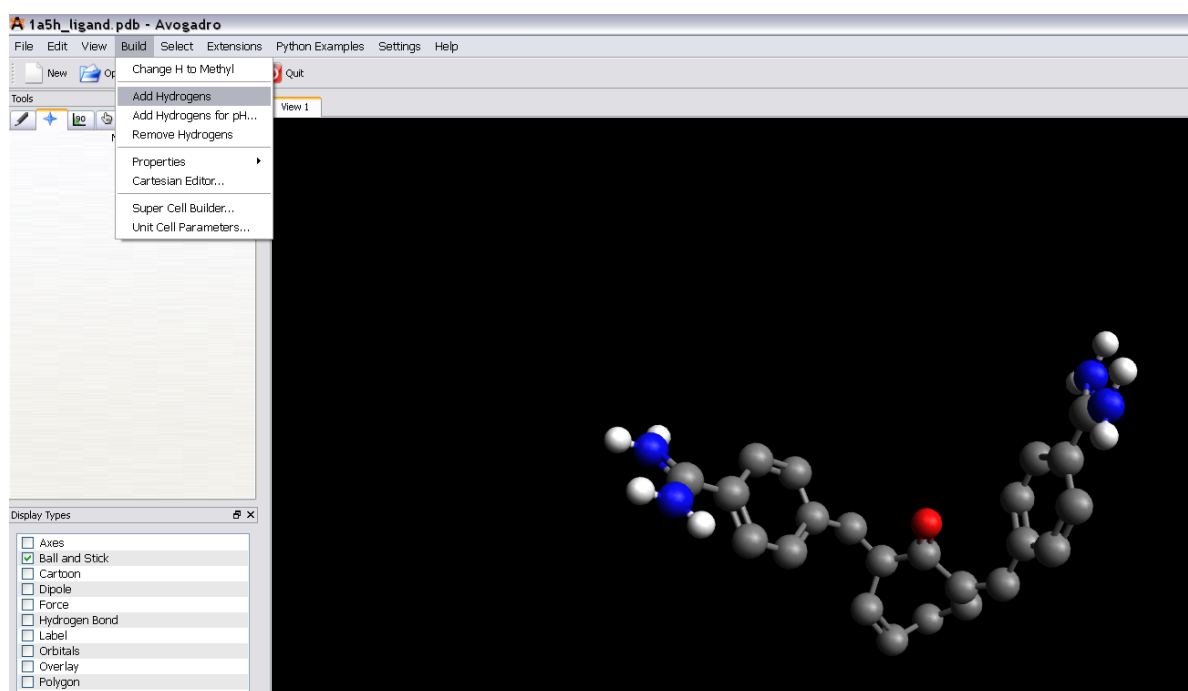
Другой способ – визуально оценить расположение лигандов в белке. Ингибитор располагается в активном центре белка, в отличие от молекул растворителя, которые обычно расположены вокруг белка на его поверхности.

Молекулу ингибитора копируем в отдельный файл (копируем целиком строки с атомами лиганда). Называем новый файл лиганда `2DN1_ligand.pdb`, где 2DN1 – четырехбуквенный идентификатор белка.

НЕТАТМ	2208	CAD	HEM	A	142	43.261	25.099	17.566	1.00	15.21	C
НЕТАТМ	2209	CBD	HEM	A	142	42.839	24.846	18.977	1.00	17.87	C
НЕТАТМ	2210	CGD	HEM	A	142	43.678	24.272	20.010	1.00	18.19	C
НЕТАТМ	2211	O1D	HEM	A	142	43.808	24.575	21.209	1.00	22.11	O
НЕТАТМ	2212	O2D	HEM	A	142	44.130	23.051	19.594	1.00	20.95	O
НЕТАТМ	2213	NA	HEM	A	142	41.564	30.074	16.347	1.00	12.40	N
НЕТАТМ	2214	NB	HEM	A	142	39.387	30.988	14.744	1.00	11.55	N
НЕТАТМ	2215	NC	HEM	A	142	38.757	28.265	14.112	1.00	12.84	N
НЕТАТМ	2216	ND	HEM	A	142	40.884	27.376	15.721	1.00	13.27	N
НЕТАТМ	2217	FE	HEM	A	142	40.164	29.186	15.187	1.00	11.43	FE

Вырезанная из PDB нативная структура лиганда не содержит атомов водорода. Для дальнейшей работы необходимо осуществить протонирование нативного лиганда. Протонирование осуществляется с помощью модуля OpenBabel, который встроен в визуальный редактор Avogadro (версии 1.2):

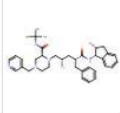
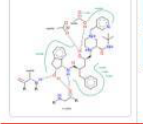
- 1) Установите программу с сайта <https://avogadro.cc>
- 2) Откройте PDB-файл лиганда с помощью меню:
File > Open
- 3) Добавьте водороды (проведите протонирование) через меню:
Build > Add Hydrogens for pH... > 7.4 > Ok.
- 4) Сохраните лиганд с помощью меню в MOL-формате (`2DN1_ligand.mol`):
File > Save as... > MDL SDFFile(*.mol)



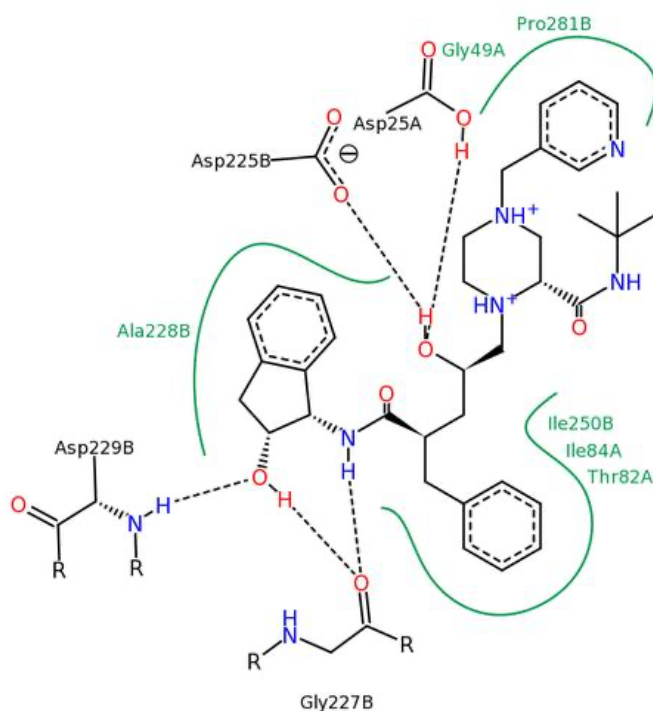
■ Как проверить правильность расстановки зарядов и навешивания водородов на лиганд?

Самый верный способ проверки протонирования – посмотреть статью на сайте PDB, в некоторых статьях приводится подробная информация с рисунками. Для корректировки протонирования лиганда необходимо с помощью визуального редактора Avogadro вручную изменить количество водородов (см. далее).

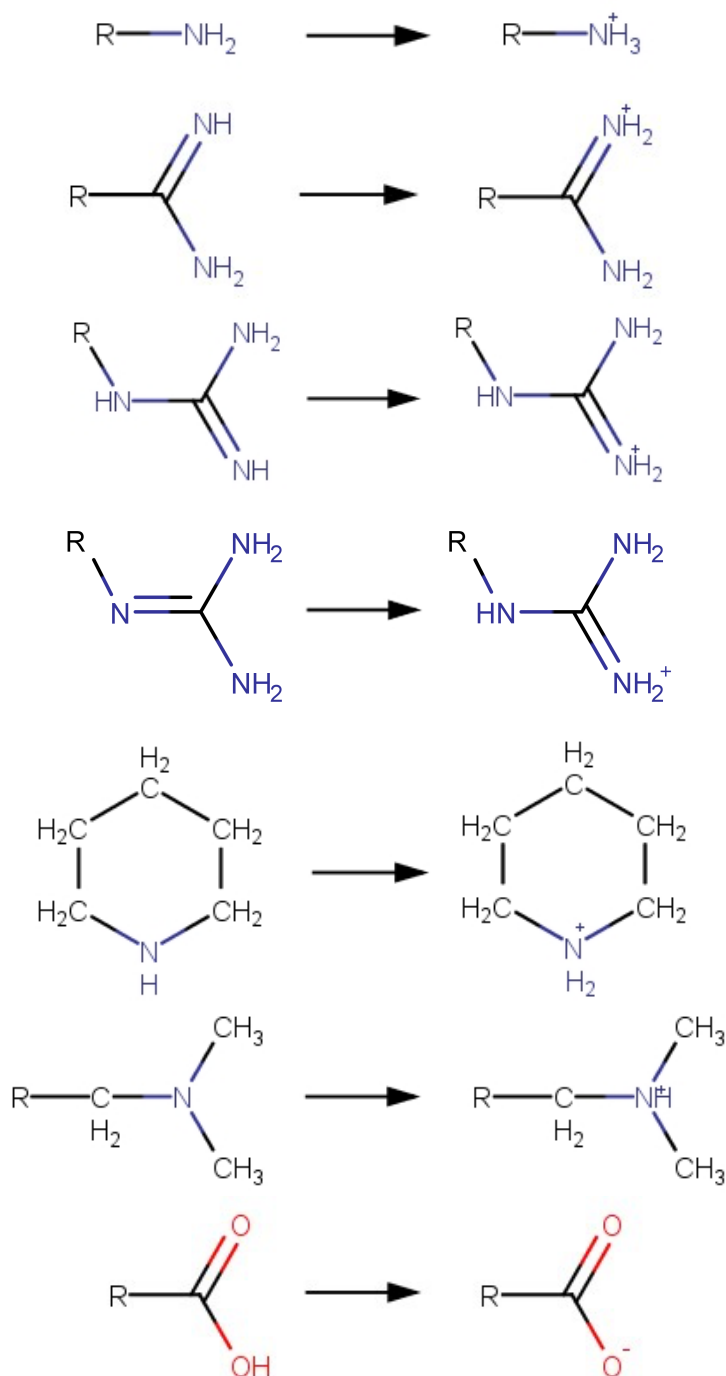
Также, на сайте есть раздел **Small molecules**, в котором представлена как двумерная структура молекулы «2D Diagram», так и схема взаимодействий лиганда с аминокислотными остатками активного центра белка «2D Interactions». Проверять правильность расстановки водородов по «2D Diagram» не стоит, так как последнее время все представленные молекулы рисуются на сайте PDB в нейтральном положении. Если же присутствует «2D Interactions», то по ней можно увидеть какие химические группы лиганда имеют положительный или отрицательный заряд, и каким образом они участвуют во взаимодействии с белком.

Small Molecules				
Ligands 1 Unique				
ID	Chains	Name / Formula / InChI Key	2D Diagram & Interactions	3D Interactions
MK1 Query on MK1	B	N-[2(R)-HYDROXY-1(S)-INDANYL]-5-[(2(S)-TERTIARY BUTYLAMINOCARBONYL)-4(3-PYRIDYLMETHYL)PIPERAZINO]-4(S)-HYDROXY-2(R)-PHENYLMETHYLPENTANAMIDE INDINAVIR (<i>Synonym</i>) C ₃₆ H ₄₇ N ₅ O ₄ CBVCZFGXHXORBI-PXQQMZJSSA-N		 Ligand Explorer NGL Binding Pocket (JSmol)

Poseview Image of MK1 in 1C6Y





Также, обязательно нужно проверить правильность расстановки зарядов на указанных ниже химических группах при нормальном pH = 7.4. Если в лиганде встречаются такие группы, то нужно изменить положения водородов (добавить или убрать их вручную) следующим образом:



Дополнительно существуют молекулярные редакторы, позволяющие определять протонированное состояние определенных химических групп или целых молекул при различных значениях pH среды, например, программа Marvin Suite (см. Приложение А). Их использование в данной практической работе не является обязательным.

■ Использование программы Avogadro для удаления или добавления атомов водорода

Если для подготовленной молекулы лиганда `2DN1_ligand.mol` требуется убрать или добавить атомы водорода, то это можно сделать в программе Avogadro в режиме редактирования молекулы. Удалить лишний атом водорода в программе Avogadro можно в режиме  «Draw Tool» щелкнув по лишнему атому правой клавишей мыши. Добавить недостающий атом водорода можно в этом же режиме, выбрав «Element: Hydrogen (1)» и дорисовав нужный атом и связь в окне View. Поскольку в этом случае есть высокая вероятность поставить атом не очень точно, далее следует провести следующие действия:

- 1) Выбрать все атомы молекулы в режиме Selection Tool , кроме добавленного атома водорода (можно сделать это через меню Select > Select All, а затем убрать выделение с добавленного атома).
- 2) Зафиксировать выделенные атомы (меню Extensions > Molecular Mechanics > Fix Selected Atoms)
- 3) Провести оптимизацию геометрии (меню Extensions > Optimize Geometry). Изменить положение при этом должен только добавленный атом водорода. После этого молекулу можно сохранить в формате mol: `2DN1_ligand.mol`

3. Подготовка белка

Из скаченного PDB-файла удаляем записи, начинающиеся с HETATM, которые описывают низкомолекулярные соединения: ингибиторы (нативный лиганд), молекулы растворителя (молекулы воды), кофакторы и т.п.

Записи, начинающиеся с HETATM, и относящиеся к атомам металла или нестандартным аминокислотам, оставляем. В некоторых комплексах между строк АТОМ могут располагаться записи, начинающиеся с ANISOU (информация об анизотропном температурном факторе). Записи ANISOU не нужны для дальнейшей подготовки белка, однако удалять их не нужно, так как используемая программа для протонирования белка APLITE их игнорирует.

Иногда в PDB оказывается закристаллизованной не одна структура комплекса белок-лиганд, а несколько копий идентичных структур. В этом случае, помимо идентичных копий белка в структуре будет закристаллизовано и столько же копий лиганда. Проверить, так это или нет, можно визуально в редакторе, а также в шапке PDB-файла для каждого MOL_ID будет указано наличие нескольких цепей. На приведенном ниже примере 6 идентичных цепей с одним ID:

```
COMPND    MOL_ID: 1;  
COMPND    2 MOLECULE: GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2;  
COMPND    3 CHAIN: A, B, C, D, E, F;
```

Если структура белка содержит идентичные цепи, то оставляем их в единственном экземпляре. Обычно оставляются наиболее полные (т.е. содержащие меньше пропущенных атомов и аминокислот) последовательности. Для того чтобы удалить идентичную цепь белка, необходимо удалить все атомы (то есть все записи, начинающиеся с «АТОМ»), относящиеся к этой цепи.

Некоторые структуры белка состоят из двух и более различных цепей, например ВИЧ-протеаза, которая является гомодимером, составленным из двух идентичных белковых молекул. В приведенном ниже примере белок содержит две различные цепи с разными ID. В этом примере каждая из цепей А и В присутствует в единственном экземпляре и удалять их записи «АТОМ» не нужно.

```
COMPND    MOL_ID: 1;
COMPND    2 MOLECULE: COAGULATION FACTOR XII;
COMPND    3 CHAIN: A;
COMPND    8 MOL_ID: 2;
COMPND    9 MOLECULE: COAGULATION FACTOR XII;
COMPND   10 CHAIN: B;
```

Оставшиеся записи, в том числе записи, начинающиеся с CONECT (их наличие важно), сохраняем в файле 2DN1_prot.pdb.

```
АТОМ      1  N  LEU A  2      34.835  47.594  4.410  1.00  39.80      N
АТОМ      2  CA LEU A  2      34.159  48.253  5.522  1.00  31.22      C
АТОМ      3  C  LEU A  2      34.154  49.768  5.289  1.00  25.26      C
АТОМ      4  O  LEU A  2      35.136  50.451  5.041  1.00  29.93      O
АТОМ      5  CB LEU A  2      34.752  47.945  6.891  1.00  35.80      C
АТОМ      6  CG LEU A  2      34.313  46.773  7.765  1.00  39.40      C
АТОМ      7  CD1 LEU A  2      33.483  45.732  7.045  1.00  43.91      C
АТОМ      8  CD2 LEU A  2      35.517  46.056  8.360  1.00  38.99      C
АТОМ      9  N  SER A  3      32.931  50.295  5.371  1.00  23.94      N
АТОМ     10  CA SER A  3      32.738  51.731  5.262  1.00  22.01      C
АТОМ     11  C  SER A  3      33.023  52.406  6.594  1.00  19.93      C
АТОМ     12  O  SER A  3      33.079  51.773  7.658  1.00  20.60      O
АТОМ     13  CB SER A  3      31.292  52.068  4.872  1.00  21.54      C
АТОМ     14  OG SER A  3      30.447  51.757  5.995  1.00  22.21      O
АТОМ     15  N  PRO A  4      33.188  53.729  6.572  1.00  21.12      N
АТОМ     16  CA PRO A  4      33.325  54.447  7.835  1.00  20.51      C
АТОМ     17  C  PRO A  4      32.196  54.163  8.815  1.00  21.75      C
АТОМ     18  O  PRO A  4      32.438  53.987  10.031 1.00  22.36      O
АТОМ     22  N  ALA A  5      30.954  54.113  8.333  1.00  20.31      N
АТОМ     23  CA ALA A  5      29.832  53.798  9.220  1.00  18.34      C
АТОМ     24  C  ALA A  5      29.936  52.395  9.793  1.00  18.36      C
АТОМ     25  O  ALA A  5      29.608  52.183  10.951 1.00  19.86      O
АТОМ     26  CB ALA A  5      28.505  53.965  8.505  1.00  18.79      C
...
CONECT   643 2217
CONECT   1751 2276
CONECT   2175 2179 2206
CONECT   2176 2182 2189
...
```

Далее, внимательно просматриваем PDB-файл белка (удобнее в текстовом редакторе) на предмет поиска пропущенных аминокислотных остатков и пропущенных атомов в аминокислотных остатках.

Пропущенные аминокислотные остатки указаны в начале файла в качестве:

REMARK 465 MISSING RESIDUES

Пропущенные атомы также указываются в начале файла:

REMARK 470 MISSING ATOM

Если обе секции отсутствуют, то это означает, что в структуре белка нет пропущенных атомов или аминокислотных остатков.

Если есть пропущенные аминокислотные остатки, то необходимо проверить их расположение. Для этого на сайте PDB необходимо перейти в раздел **Structure** (вверху страницы) и открыть 3D визуализацию комплекса белка и лиганда. Последовательность аминокислот в цепи белка указана сверху визуализации, выбранная аминокислота выделяется цветом на 3D структуре. Номера отсутствующих аминокислот указаны в секции «REMARK 465 MISSING RESIDUES».



The screenshot shows the RCSB PDB website interface for entry 4CRC. The main content area displays a 3D ribbon model of the protein structure in green, with red dots representing missing residues. The sequence of the protein is shown at the top, with the missing residues highlighted in red. The interface includes a navigation menu at the top, a search bar, and a sidebar with various tools and options.

Если пропущенные аминокислотные остатки находятся вдали от сайта связывания лиганда с белком, то их отсутствие игнорируется. Если пропущенные аминокислотные остатки или атомы находятся вблизи активного центра, то их необходимо восстановить (см. 0).

Рентгеноструктурный анализ позволяет определять положения только тяжелых атомов (всех, за исключением атомов водорода). Поэтому для проведения докинга в структуру белка необходимо добавить атомы водорода. С помощью программы APLITE в режиме расстановки водородов преобразуем 2DN1_prot.pdb в 2DN1_prot_h.pdb (параметры для файла aplite-work.par можно найти по ссылке: <http://moldesign.ru/files/files/aplite-full.par>):

```
aplite h2pdb aplite-work.par 2DN1_prot.pdb
```

4. Определение активного центра белка

Поскольку ингибитор, закристаллизованный с белком, обычно располагается именно в активном центре белка, то центр активного центра белка должен приблизительно совпадать с геометрическим центром закристаллизованного лиганда. Таким образом, можно или визуальнo определить центральный атом лиганда, или (что лучше) определить его точно с помощью программы APLITE.

Для определения активного центра белка используется молекулярная структура нативного лиганда в формате MOL или PDB. Наличие атомов водорода в нативном лиганде не обязательно. С помощью программы APLITE осуществляется вычисление координат ближайшего атома к геометрическому центру молекулы лиганда. Полученные координаты принимаются за активный центр белка.

```
aplite par4grid 2DN1_ligand.mol
```

В результате работы программы создается файл параметров с координатами активного центра белка для программы SOLGRID:

```
CENTER X: 32.729  
CENTER Y: 36.202  
CENTER Z: -32.098
```

Область докинга у нас имеет форму куба, с центром в найденных выше координатах и гранью (по умолчанию) равной 22 Å.

5. Построение сетки

С помощью программы SOLGRID осуществляется построение сетки потенциалов 2DN1_prot_h.bin. Входными файлами программы SOLGRID является файл параметров с координатами активного центра белка 2DN1.par и подготовленный белок 2DN1_prot_h.pdb

```
solgrid 2DN1.par 2DN1_prot_h.pdb
```

Построение сетки займет некоторое время (от часа и более, в зависимости от мощности компьютера). В конце работы программа запишет сетку потенциалов в файл 2DN1_prot_h.bin, размер его будет составлять около 180-200 МБ.

6. Докинг

С помощью программы SOL осуществляем задочивание нативного лиганда 2DN1_ligand.mol в сетку 2DN1_prot_h.bin. Для работы программы SOL необходим файл параметров sol-work.par

```
sol sol-work.par 2DN1_prot_h.bin 2DN1_ligand.mol
```

Файл параметров sol-work.par можно найти по ссылке: <http://moldesign.ru/files/files/sol-work.par>. Процедура докинга в зависимости от мощности компьютера и размера лиганда может занять от 2-3 часов до суток.

7. Анализ результатов докинга

Результаты работы программы SOL по умолчанию выводятся в файл 2DN1_ligand.out. Также в конце работы программа записывает конечные конформации лигандов в файлы в одном из доступных форматов файлов (.mol, .mrk, .hin или .sdf). Количество выдаваемых файлов – от 0 до количества независимых запусков программы (1 по умолчанию). Сколько конечных конформаций лиганда будет записано и в каком формате, определяется в файле параметров SOL.

В out-файле программы приведены названия файлов белка и лиганда, значения параметров докинг (программы и генетического алгоритма), параметры построения сетки, параметры лиганды. После завершения работы в данный файл записывается суммарная статистика по всем независимым запускам программы, отсортированным по значению Docked Energy.

Final statistics for 50 runs, ordered by increasing energy

N	Docked energy	Cluster N	initial position RMS	Inner Energy	Grid Energy	order0 grid	order1 grid	order2 grid
1	-31.1879	1	0.3458	31.2229	-62.4109	-30.5483	-41.2035	9.3410
2	-31.1843	1	0.3047	31.1998	-62.3841	-30.4866	-41.1743	9.2768
3	-31.1296	1	0.3176	31.3806	-62.5102	-30.1995	-41.6346	9.3239

Эта статистика включает в себя номер независимого запуска (N), значение Docked Energy, номер кластера, в который попала данная задоченная конформация лиганда, среднеквадратичное отклонение этой конформации, от конформации лиганда, поданной на вход (initial position RMS), внутреннюю энергию напряжений лиганда (Inner Energy), а также составляющие сеточной энергии, которые не зависят от заряда атома лиганда (order0 grid), пропорциональны заряду атома лиганда (order1 grid) и пропорциональны квадрату заряда атома лиганда (order2 grid) соответственно.

Далее, выходной файл содержит итог по этой статистике:

- Количество кластеров (RMSD лигандов в которых составляет менее 1 Å),
- Информацию по каждому кластеру:
 - номер кластера (Clust.N);
 - населенность кластера – число лигандов, попавших в кластер (popul.);
 - минимальное, максимальное и среднее среднеквадратичный расстояния между лигандами в кластере (min RMS, max RMS, mean RMS);
 - минимальное значение Docked Energy для этого кластера (min docked energy);
 - минимальное значение сеточной энергии для этого кластера (min grid energy);
 - минимальное значение скоринг-функции для этого кластера (min scoring – оно соответствует минимальному значению сеточной энергии):

```
*****
Founding 1 clusters, separated by more than 1.0000 angstroms
Clust.  popul.  min RMS  max RMS  mean RMS  min docked  min grid  min scoring
N                                     energy      energy
1           50      0.0152  1.2464  0.2950  -31.1879  -62.6112  -4.6111
```

- И, наконец, выходной файл содержит информацию о «лучшем» лиганде – лиганде, обладающем минимальной Docked Energy:
 - скоринг-функцию этого лиганда (BEST SCORING);
 - сеточную энергию этого лиганда (BEST GRID ENERGY);
 - число вращательных степеней свободы этого лиганда (NUMBER OF ROTATORS).

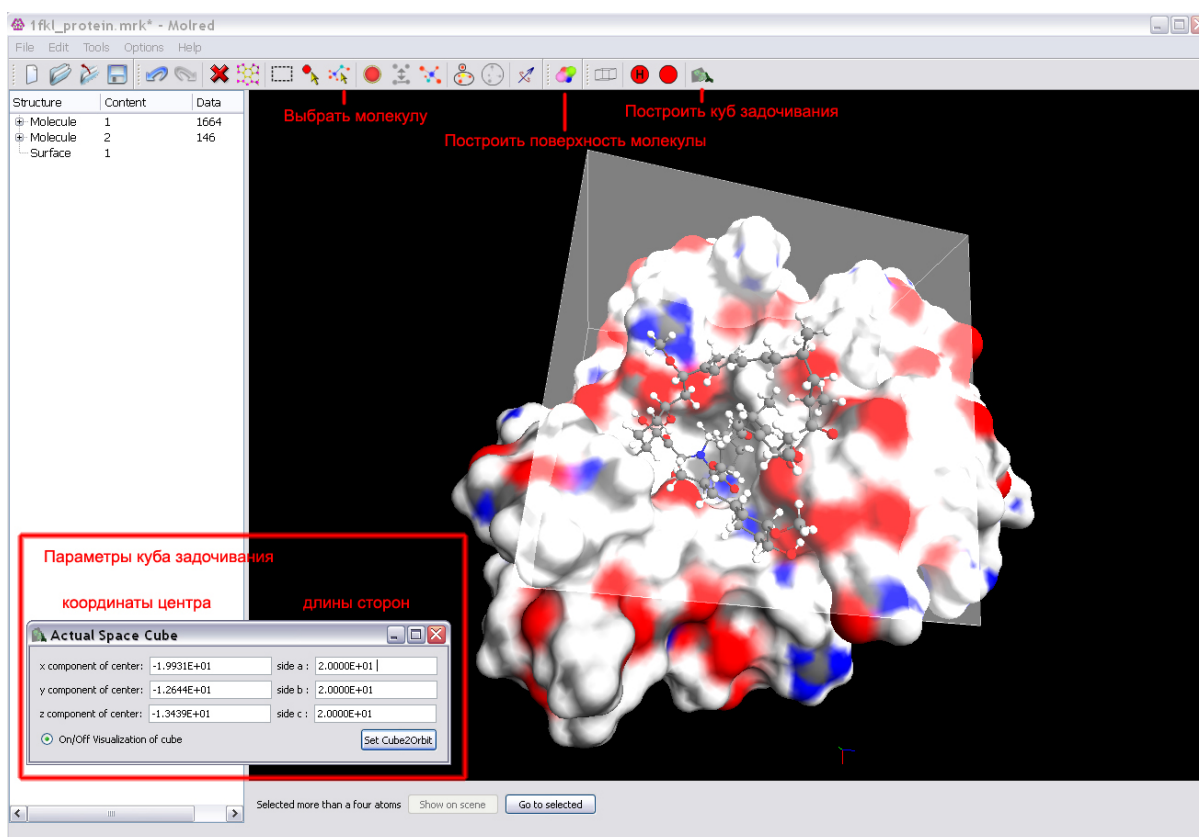
Скоринг-функция (BEST SCORING) приводится в программе SOL в ккал/моль. Конформация «лучшего» лиганда записывается в файл «2DN1_ligand_01.*».

В конце out-файла выводится время работы программы:

```
*****
BEST SCORING: -4.591086
BEST GRID ENERGY: -62.410857
NUMBER OF ROTATORS: 5
*****
Total time (6382.000000 seconds):
  0 days
  1 hours
  46 minutes
  22 seconds
```

8. Просмотр результатов докинга

Увидеть положения лигандов в кубе задочивания после докинга можно с помощью молекулярного редактора MolRed.



Для этого с помощью команды File > Open следует открыть файл белка в формате PDB или MRK (2DN1_prot_h.pdb). Далее, можно выбрать целиком всю молекулу белка (Molecule Select на командной панели > дважды кликнуть на белок) и построить для нее поверхность (кнопка Build Surface на командной панели или в меню Tools). Использование поверхности позволяет лучше определить активный центр и место позиционирования лиганда.

Затем с помощью команды File > Merge следует добавить в программу лиганд в формате MOL, HIN или MRK. С помощью этой команды можно добавить в программу несколько лигандов, например, нативный лиганд 2DN1_ligand.mol и лиганд после докинга 2DN1_ligand_01.mrk.

Чтобы отобразить куб задочивания, нужно вызвать окно «Actual Space» на командной панели или через Options > Display > Actual Space и в нем нажать радиокнопку «On/Off Visualization of cube». Здесь же необходимо будет ввести координаты центра куба докинга и его размеры. Размер грани куба докинга по умолчанию – 22 Å.

9. Сравнение скоринг-функции с экспериментальными данными

Для численного сравнения результатов работы программы SOL с экспериментальными данными, необходимо найти соответствие между теоретически рассчитанной энергией связывания белка и лиганда и ее экспериментальным значением. Программа SOL позволяет рассчитать теоретическое значение ΔG – скоринг-функцию (она имеет размерность ккал/моль). Экспериментальное значение энергии связывания белка и лиганда ΔG можно получить из экспериментальных данных (величин K_d , K_i или IC_{50}).

Если в качестве экспериментальных данных дано значение K_d , K_i или IC_{50} , то экспериментальную энергию связывания можно представить как:

$$\Delta G = R * T * \ln(K_d), \text{ где}$$

R – универсальная газовая постоянная (8,31 Дж/(моль*К)),

T – температура в К (берем комнатную темп. 300 К),

ΔG – энергия связывания (Дж/моль),

K_d – константа диссоциации (можно также использовать K_i или IC_{50}).

При конкурентном ингибировании константу диссоциации комплекса белка с ингибитором называют константой ингибирования K_i . Иногда в экспериментах измеряется величина IC_{50} : вообще говоря, она зависит от условий эксперимента, однако в случае если концентрация субстрата при конкурентном ингибировании много меньше константы Михаэлиса для субстрата, то IC_{50} можно считать приблизительно равной K_i . Таким образом, подставляя в приведенную выше формулу K_d , K_i или IC_{50} (если в PDB не указано других экспериментальных данных), можно получить экспериментальное значение свободной энергии Гиббса ΔG и сравнить его с теоретическим значением ΔG (скоринг-функцией программы SOL), полученным при докинге.

Приложение А. Протонирование программой Marvin Suite

Случаи, когда при нейтральном рН определённые группы заряжаются положительно или отрицательно (за счет присоединения или потери водорода), конечно, этим не ограничиваются. При отсутствии хороших знаний в области органической химии, которые помогли бы предсказать способ протонирования в более сложных случаях (предсказать в том числе с использованием таблиц со значениями рKa), можно воспользоваться возможностями программы Marvin Suite (можно скачать здесь <https://www.chemaxon.com/download/marvin-suite/#marvin>). Программу можно установить и без лицензии, но нужная нам функция без лицензии, скорее всего, не будет работать. Можно легко получить пробную (“Evaluation”) лицензию. Внимание: она действует только 3 дня. Для того, чтобы получить её, нужно зарегистрироваться на сайте, с которого скачивали программу. Будучи зарегистрированным, Вы увидите в шапке сайта Ваш e-mail. Наведя мышью на него, выберите “My Evaluations”. Здесь Вы сможете скачать пробную лицензию (или оформить академическую, но для этого нужно подать заявку и дождаться ответа от администрации сайта): Bundle Name > “Marvin”, Option > “Download Free License”.

Итак, после того, как Вы открыли в программе Avogadro лиганд в формате PDB, добавили атомы водорода (Build > Add Hydrogens) и сохранили лиганд в формате MOL (получили файл 2DN1_ligand.mol), откройте его с помощью MarvinView. Далее: Tools > Protonation > Major Microspecies. В появившемся окне впишите рН, для которого Вы хотите узнать преобладающее зарядовое состояние данного лиганда. Оставьте флажок напротив опции Keep explicit hydrogens и оставьте без флажка опцию Take major tautomeric form. После того как нажмёте ОК, появится текстовая информация и картинка. Нажмите на картинку правой кнопкой мыши и выберите Save All..., в качестве формата сохранения выберите MDL Molfile (.mol). Сохраните эту молекулу под другим именем, например 2DN1_ligand_Marvin.mol. Теперь Вы можете открыть этот файл с помощью MarvinView и увидеть, какие группы при данном рН заряжены отрицательно или положительно. **ВНИМАНИЕ:** Далее возвращайтесь к основному файлу 2DN1_ligand.mol, в который Вы ещё никаких изменений не вносили после добавления водородов ко всем тяжелым атомам в Avogadro функцией Build > Add Hydrogens. (Одна из причин в том, что в файле 2DN1_ligand_Marvin.mol не будет явно добавлен атом водорода, несмотря на то, что в MarvinView Вы увидите положительный заряд на какой-либо группе).

Приложение В. Восстановление структуры белка

Если белок содержит пропущенные атомы или аминокислотные остатки вблизи активного центра, то необходимо их восстановить. Для этого в базе данных PDB проводим поиск альтернативных структур исследуемого белка, содержащих пропущенные аминокислотные остатки и атомы. Найденную структуру, в которой присутствуют пропущенные аминокислоты или атомы, назовем эталонной.

Далее переведем эталонную структуру белка в систему координат исследуемого белка путем вращения и трансляции эталонного белка как жесткого тела. Это преобразование производится с помощью программы `superimpose-protein`. Первый аргумент программы – это эталонная структура (которую будем вращать). Второй аргумент – это исследуемая структура, содержащая пропуски (на которую будет наложена эталонная структура).

Строка запуска программы:

```
superimpose-protein.exe etalon.pdb to_reduce.pdb
```

В ходе работы программа сравнивает расположение атомов аминокислотных остатков друг относительно друга и выводит на экран совпадающие атомы. Далее, программа предлагает убрать из сравнения часть атомов и заново проводит сравнение. Если убирать атомы из сравнения больше не нужно, то осуществляется вращение эталонной структуры, которая сохраняется в файл с именем `etalon_rotated.pdb`. Дополнительно программа предлагает перевести в новую систему координат структуру нативного лиганда, соответствующую эталонной структуре белка. Данная опция необходима для проведения кросс-докинга.

Полученная структура `etalon_rotated.pdb` соответствует эталонному белку, в котором координаты атомов записаны в системе координат исследуемого белка. Далее открываем полученный файл в текстовом редакторе и вырезаем из него строки с пропущенными в исследуемом белке аминокислотными остатками и/или атомами. После чего вставляем вырезанные строки в соответствующий файл исследуемого белка `to_reduce.pdb`.

Приложение С. Сравнение с экспериментом с помощью $\log(Kd)$

В случае если в качестве экспериментальных данных дано значение $\log(Kd)$, экспериментальная величина будет равняться:

$$-\log(Kd) = -(\Delta G * \log(e))/(R * T)$$

Для того чтобы сравнить величину $-\log(Kd)$ с ее табличным значением надо перевести ΔG в Дж/моль (1 кал = 4.2 Дж): ΔG (ккал/моль) = $4.18 * 10^3 \Delta G$ (Дж/моль)

Если подставить значения констант, то получим выражение:

$$-\log(Kd) = -(4.18 * 10^3 * \Delta G * 0.43)/(8.31 * 300)$$

или

$$-\log(Kd) = -0.7282 * \Delta G$$

при этом ΔG – значение скоринг-функции программы SOL.

На основании данного выражения возможно дальнейшее сравнение скоринг-функции с экспериментальной величиной $-\log(Kd)$.