

МОЛЕКУЛЯРНЫЙ ДОКИНГ

Руководство пользователя

Редакция 2023 года

Молекулярный докинг – это позиционирование низкомолекулярного соединения (лиганда) в активный центр протеина. 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 НЕТАТМ (следует обратить внимание, что с пометкой НЕТАТМ 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 протеина	8
4. Определение активного центра протеина	10
5. Построение сетки	10
6. Докинг	11
7. Анализ результатов докинга	11
8. Просмотр результатов докинга	13
9. Сравнение скоринг-функции с экспериментальными данными	14
Приложение А. Протонирование программой Marvin Suite	15
Приложение В. Сравнение с экспериментом с помощью $\log(K_d)$	16

1. Скачивание молекулярных структур

На сайте базы данных Protein Data Bank <http://www.rcsb.org/pdb/home/home.do> доступны для свободного скачивания данные о трехмерной структуре нативного комплекса лиганд-протеин, 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 протеина 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/pdb/1c5w/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) [SBIK.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

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

Literature

Download Primary Citation

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

- При выборе структуры протеина 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](#)

Q

[Full Protein Feature View for P00749](#)

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-CARBOXAMIDINE <chem>C9H8I N2S</chem> YERQOXAYAFWFEJ-UHFFFAOYSA-D		Ligand Explorer Binding Pocket (JSmol) Electron Density (JSmol)
FLC Query on FLC Download SDF File Download CCD File	B	CITRATE ANION <chem>C6H5O7</chem> 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.).

- **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.

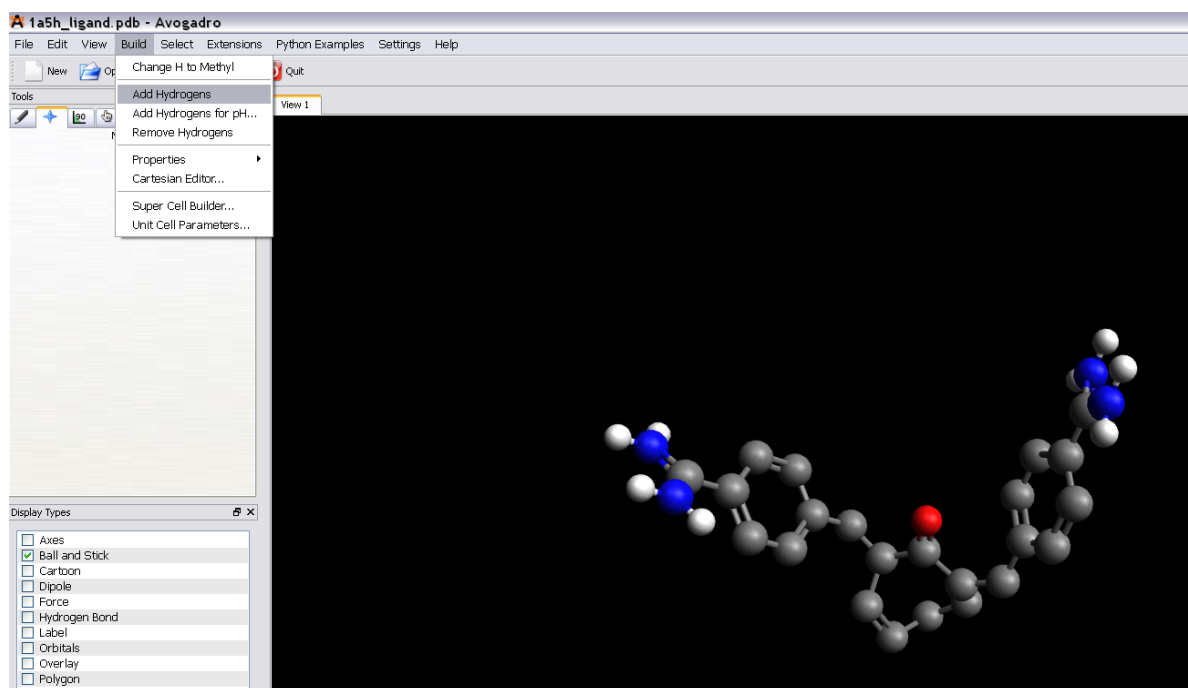
HETATM	1893	C	BCT	A	302	14.831	40.554	52.211	1.00	43.04	C
HETATM	1894	O1	BCT	A	302	16.055	40.790	52.432	1.00	42.31	O
HETATM	1895	O2	BCT	A	302	13.962	40.525	33.138	1.00	44.06	O
HETATM	1896	O3	BCT	A	302	14.479	40.555	50.995	1.00	44.48	O
HETATM	1897	S	S04	A	303	7.755	60.848	22.468	0.50	41.79	S
HETATM	1898	O1	S04	A	303	8.521	59.976	21.567	0.50	42.02	O
HETATM	1899	O2	S04	A	303	7.398	60.082	23.678	0.50	41.82	O
HETATM	1900	O3	S04	A	303	8.595	62.012	22.824	0.50	41.67	O
HETATM	1901	O4	S04	A	303	6.534	61.308	21.787	0.50	41.48	O
HETATM	1902	C	BCT	A	304	15.176	33.156	39.946	1.00	50.21	C
HETATM	1903	O1	BCT	A	304	16.334	33.321	40.438	1.00	50.32	O
HETATM	1904	O2	BCT	A	304	14.079	33.366	40.594	1.00	49.95	O
HETATM	1905	O3	BCT	A	304	15.138	32.759	38.744	1.00	49.47	O
HETATM	1906	CX	682	A	1	11.303	51.284	25.507	1.00	18.08	C
HETATM	1907	CY	682	A	1	11.559	51.401	23.988	1.00	18.93	C
HETATM	1908	CZ	682	A	1	11.685	52.530	26.319	1.00	21.27	C

Описание структуры лигандов

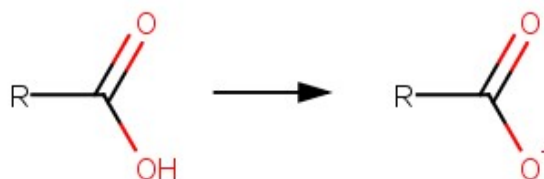
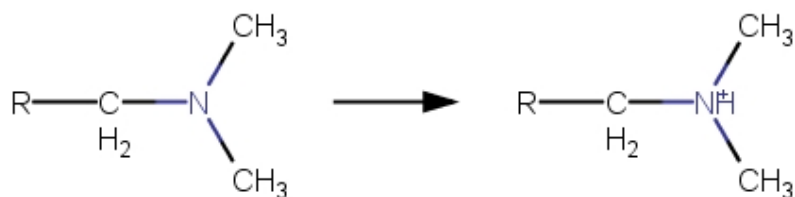
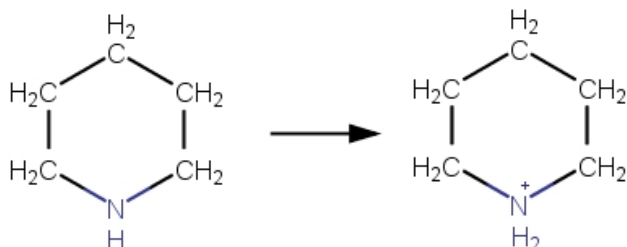
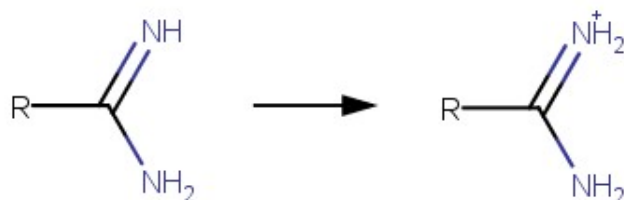
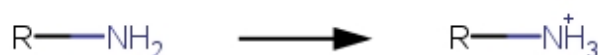
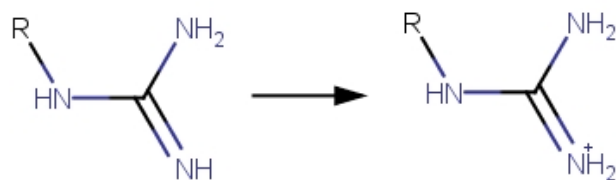
Другой способ – визуально оценить расположение лигандов в белке. Ингибитор располагается в активном центре белка (молекулы растворителя обычно расположены вокруг белка на его поверхности) и часто (но не всегда!) больше остальных лигандов; на странице комплекса в PDB может быть раздел с константами связывания именно для ингибитора.

Молекулу ингибитора выносим в отдельный файл, например, 2DN1_ligand.pdb (копируем целиком строчки).

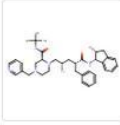
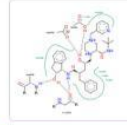
Вырезанная из PDB нативная структура лиганда, как правило, не содержит атомов водорода. Для дальнейшей работы необходимо осуществить протонирование нативного лиганда. Протонирование осуществляется с помощью различных программных инструментов, например, с помощью визуального редактора Avogadro или программы OpenBabel. В Avogadro для этого используйте опцию Build > Add Hydrogens for pH... > 7.4 > Ok. Протонированный лиганд сохраняется в формате MOL (2DN1_ligand.mol) командами «Save as» (в случае Avogadro) или «Export» (в случае OpenBabel).



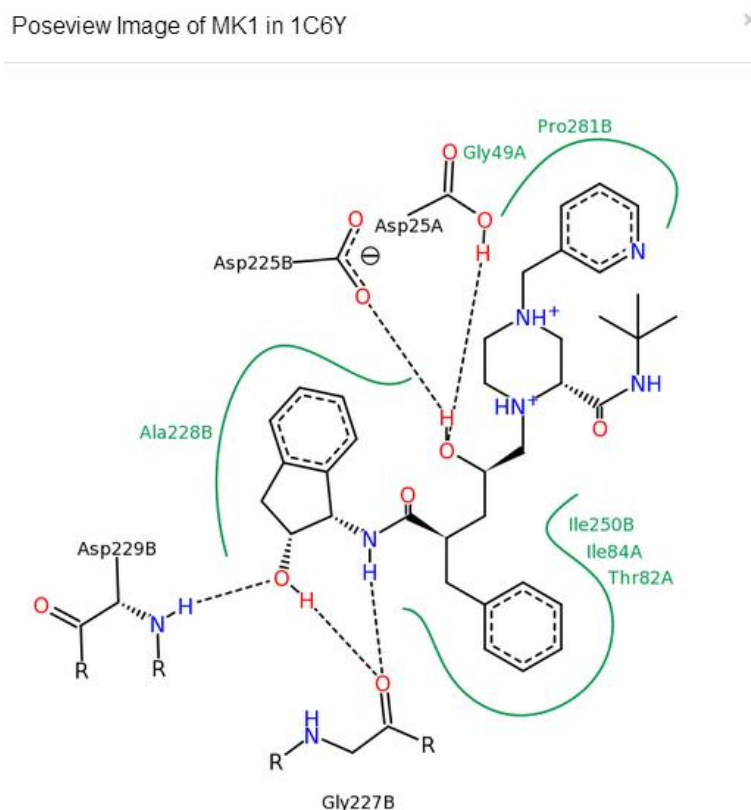
Проверяется правильность протонирования – при нормальном pH некоторые химические группы должны быть положительно или отрицательно заряжены. Если в лиганде встречаются такие группы, то нужно изменить положения водородов (добавить или убрать вручную) следующим образом:



В некоторых случаях проверить, в каком протонированном состоянии находится молекула лиганда при связывании с белком, можно при помощи сайта PDB. Для этого на странице комплекса необходимо найти раздел «2D Diagram & Interactions» в секции **Small molecules**:


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


В данном разделе представлена двумерная структура молекулы лиганда (слева), а также может быть изображена схема взаимодействий лиганда и аминокислотных остатков активного центра белка (справа). Если схема присутствует, то нажатие на нее позволит более подробно рассмотреть, какие химические группы лиганда имеют положительный или отрицательный заряд, и каким образом они участвуют во взаимодействии с белком:



Существуют также 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 протеина

Из скаченного PDB файла удаляем записи, начинающиеся HETATM, которые описывают такие низкомолекулярные соединения, как молекулы растворителя или молекулы воды. Записи, начинающиеся HETATM, и относящиеся к ионам металлов или нестандартным аминокислотам, оставляем (лиганд мы уже вырезали из этого файла). Сохраняем все оставшиеся записи (в том числе записи, начинающиеся CONECT, их наличие важно) в файле `2DN1_prot.pdb`.

АТОМ	1877	CG2	THR	A	242	17.692	36.238	40.400	1.00	35.32	C
АТОМ	1878	N	GLN	A	243	20.667	34.857	43.806	1.00	41.62	N
АТОМ	1879	CA	GLN	A	243	21.058	34.084	44.982	1.00	44.46	C
АТОМ	1880	C	GLN	A	243	21.682	34.971	46.046	1.00	44.95	C
АТОМ	1881	O	GLN	A	243	22.764	35.534	45.854	1.00	44.72	O
АТОМ	1882	CB	GLN	A	243	22.041	32.981	44.596	1.00	46.44	C
АТОМ	1883	CG	GLN	A	243	21.390	31.621	44.419	1.00	51.18	C
АТОМ	1884	CD	GLN	A	243	22.405	30.496	44.474	1.00	54.04	C
АТОМ	1885	OE1	GLN	A	243	22.756	30.002	45.550	1.00	55.68	O
АТОМ	1886	NE2	GLN	A	243	22.882	30.079	43.306	1.00	55.60	N
АТОМ	1887	N	ALA	A	244	20.991	35.091	47.174	1.00	46.11	N
АТОМ	1888	CA	ALA	A	244	21.481	35.899	48.284	1.00	47.32	C
АТОМ	1889	C	ALA	A	244	22.731	35.270	48.896	1.00	47.89	C
АТОМ	1890	O	ALA	A	244	23.751	35.968	49.041	1.00	48.36	O
АТОМ	1891	CB	ALA	A	244	20.393	36.031	49.339	1.00	46.76	C

Иногда в PDB оказывается закристаллизованной не одна структура комплекса белок-лиганд, а несколько копий идентичных структур. Проверить, так это или нет, можно визуально в редакторе, например, MOLRED, а также в этом случае 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;
```

В этом случае, помимо идентичных копий белка в структуре будет закристаллизовано и столько же копий лиганда.

Если структура протеина 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. Исключение из правила удаления всех цепей, кроме одной - подготовка белков ВИЧ-протеазы. В них мы оставляем две идентичные цепи, так как ВИЧ-протеаза является гомодимером, составленным из двух идентичных белковых молекул.

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 никаких пропусков. Если пропущенных атомов нет (нет секции «REMARK 470 MISSING ATOM»), а пропущенные аминокислотные остатки являются крайними в белке (это или первые аминокислотные остатки цепей, или последние) и находятся далеко от активного центра, то их отсутствие игнорируется.

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 протеина

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 протеина 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 протеина.

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

В результате работы программы создается файл параметров с координатами активного центра протеина 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 протеина 2DN1.par и подготовленный протеин 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 «имя_лиганда.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, и 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) и 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
```

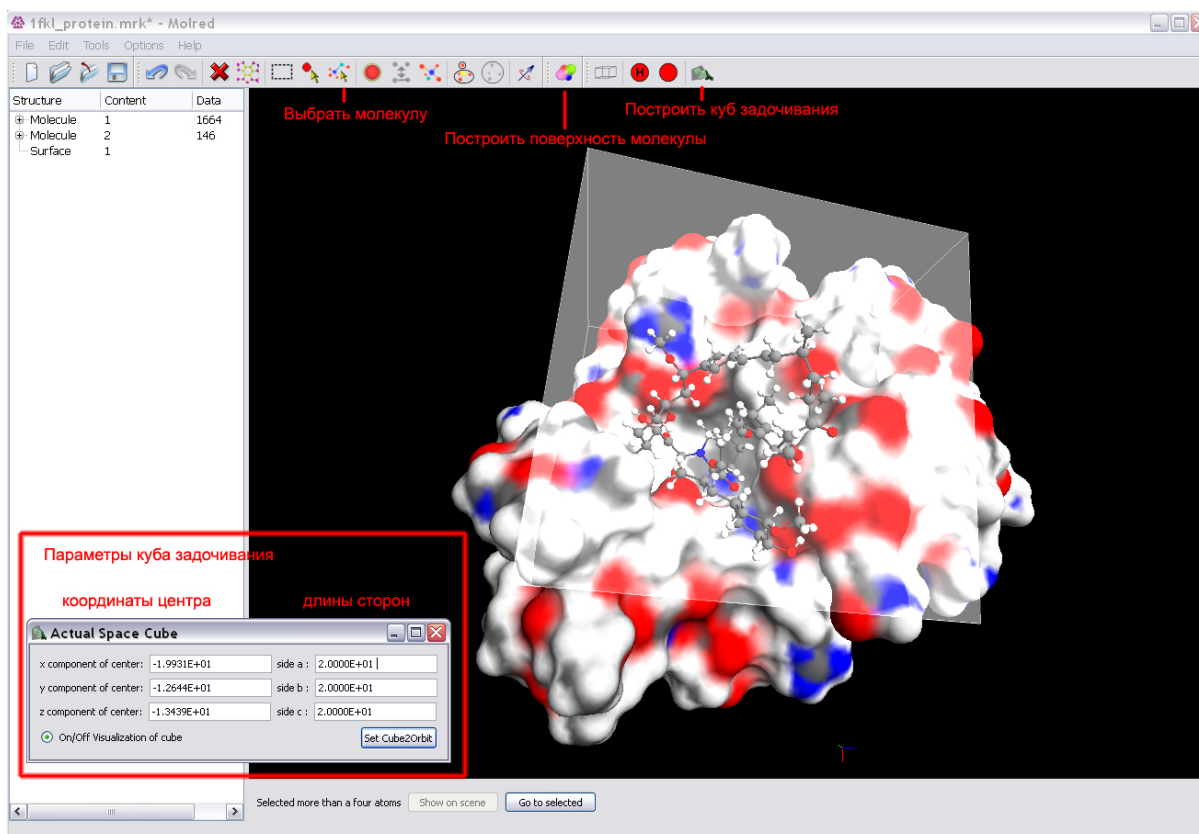
И, наконец, выходной файл содержит информацию о «лучшем» лиганде – лиганде, обладающем минимальной Docked Energy: скоринг-функцию этого лиганда (BEST SCORING), его сеточную энергию (BEST GRID ENERGY) и число вращательных степеней свободы этого лиганда (NUMBER OF ROTATORS). Конформация этого лиганда записывается в файл «имя_лиганда_01.*». BEST SCORING приводится в программе SOL в ккал/моль.

В конце 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 и построить для нее поверхность (Build Surface на командной панели или в меню Tools), на которой будет удобнее рассмотреть лиганд. 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:

$$\Delta G = R \cdot T \cdot \ln(K_d),$$

где R - универсальная газовая постоянная = 8,31 Дж/(моль*К)
 T - температура в К (берем комнатную темп. 300 К),
 ΔG - энергия связывания в 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 рКа), 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. Сравнение с экспериментом с помощью $\log(Kd)$

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

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

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

Если подставить значения констант, то получим выражение:

$$-\log(Kd) = -(4.18 \cdot 10^3 \cdot \Delta G \cdot 0.43) / (8.31 \cdot 300), \text{ или}$$

$-\log(Kd) = -0.7282 \cdot \Delta G$, при этом ΔG - значение скоринг-функции программы SOL.

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