

Компьютерные методы в фармакологии



В.Б.Сулимов

НИВЦ МГУ

Лекция № 7

**Квазидокинг, Тензорные поезда, токсичность,
заряды и 3D-структуры молекул, LogP**

Квазидокинг – путь к квантово-химическому докингу

Рельеф многомерной поверхности энергии белок-лиганд слабо зависит от выбора силового поля или метода квантовой химии – минимумы находятся примерно там же

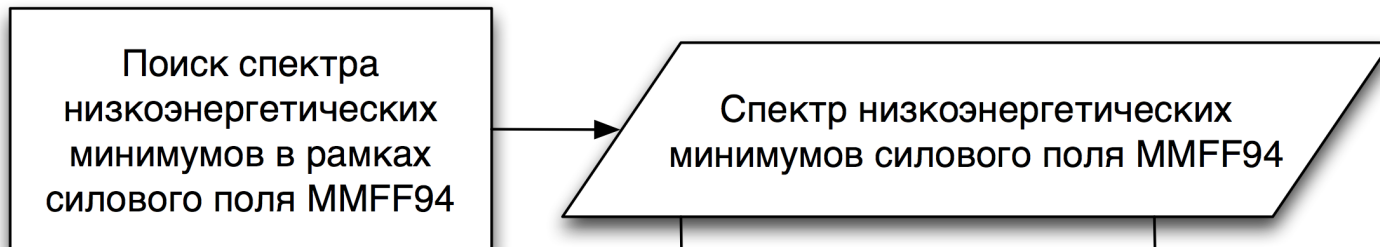
- ▶ Используя программу **FLM** в вакууме, находим несколько тысяч низкоэнергетических минимумов комплекса белок-лиганд в силовом поле MMFF94 – обычно 4 тысяч минимумов хватает – это показали исследования десятков комплексов белок-лиганд
- ▶ Энергия каждого найденного минимума вычисляется с помощью полуэмпирического квантово-химического метода **PM7** без дополнительной локальной оптимизации
- ▶ Выявляется поза с самой низкой энергией PM7. Парадигма докинга выполняется почти для всех (нескольких десятков) тестовых комплексов
- ▶ Проводится локальная оптимизация энергии **PM7** из позы лиганда с наименьшей энергией при варьировании положений атомов лиганда. Это и есть лучшая поза лиганда

Квантово-химический пакет MORAS

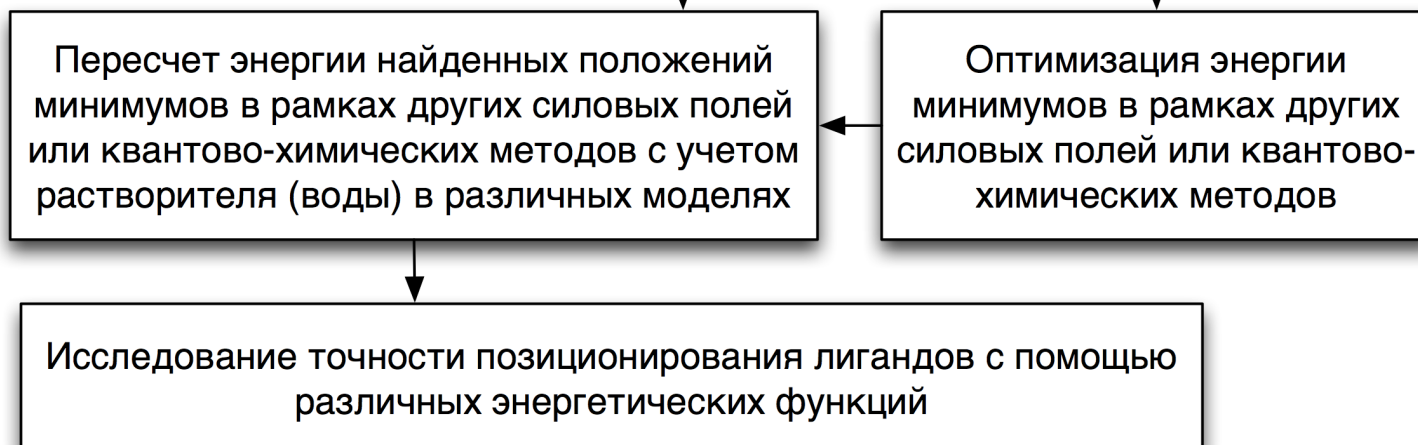
- ▶ Полуэмпирические методы квантовой химии: на основе физических соображений некоторые интегралы в матричном уравнении Шредингера заменяются нулями, некоторые вычисляются, другие заменяются аналитическими формулами; эти методы развиваются около 50 лет
- ▶ Недостатки этих упрощений компенсируются введением параметров, которые подгоняются, либо под эксперимент, либо под точные расчеты *ab initio*
- ▶ Новый квантово-химический полуэмпирический метод PM7:
 - Улучшенные описания межмолекулярных взаимодействий
- ▶ Можно проводить расчеты комплексов белок-лиганд – несколько тысяч атомов
- ▶ Для каждого расчета комплекса белок-лиганд необходимо:
 - При оптимизации лиганда около **10 часов на 1 вычислительном ядре**
 - Без оптимизации лиганда (1SCF) **несколько минут**

Процедура квазидокинга

Первый шаг



Второй шаг



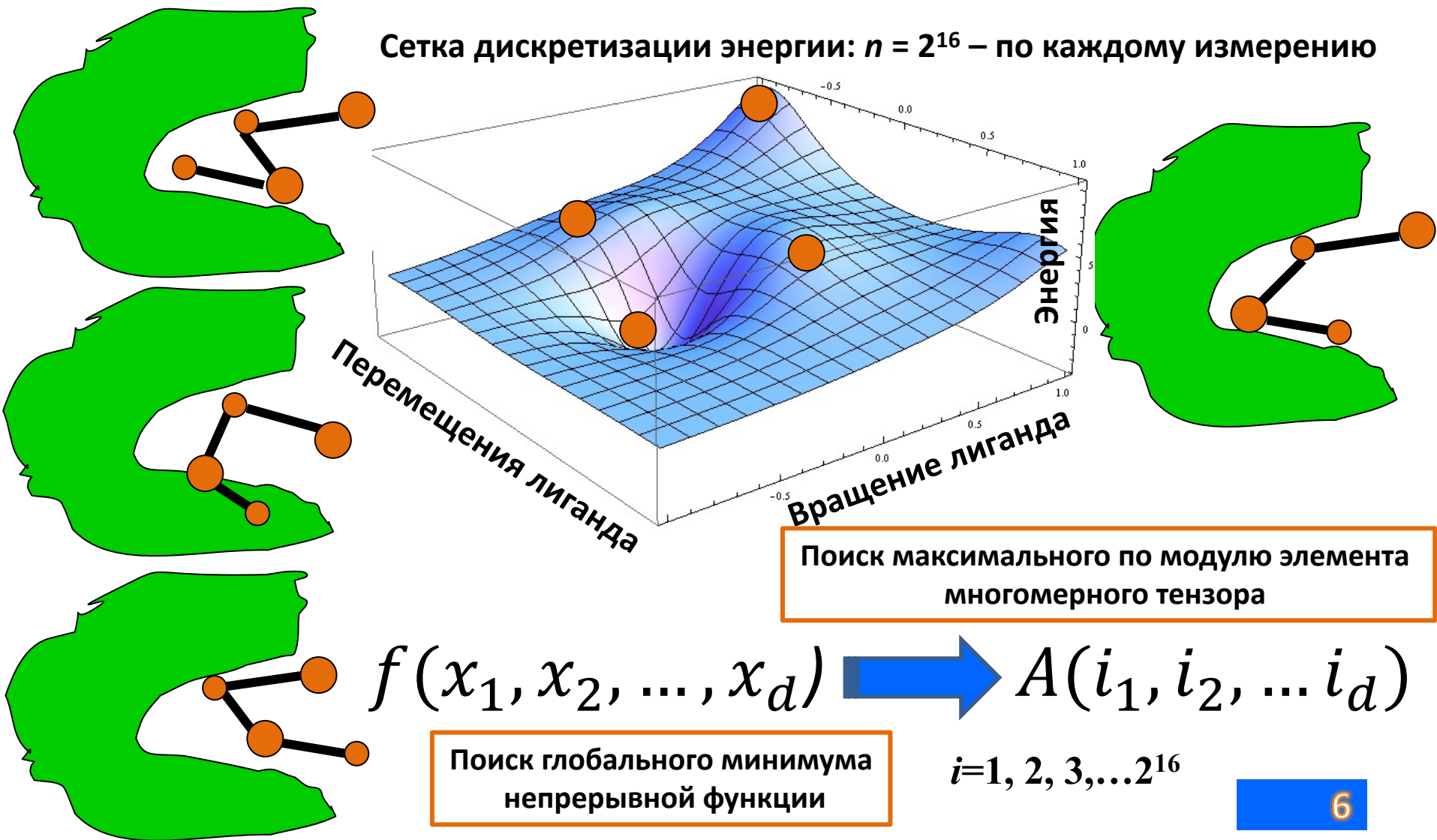
Суперкомпьютерная программа докинга SOL-P: ТТ-алгоритм докинга основан на методе глобальной оптимизации с использованием разложение тензоров в тензорные поезда – Tensor Trains

- ▶ Классическое силовое поле MMFF94, без упрощений, без подгоночных параметров
- ▶ Нет сетки потенциалов взаимодействия белок-лиганд
- ▶ Многопроцессорные параллельные вычисления: несколько сотен вычислительных ядер
- ▶ Непрерывная функция энергии системы белок-лиганд преобразуется с помощью очень мелкой сетки в многомерный тензор большого размера
- ▶ К этому тензору применяются современные методы тензорного анализа, развитые в Институте вычислительной математики РАН, для поиска максимального по модулю элемента тензора – он и будет соответствовать глобальному оптимуму исходной непрерывной функции

Программа докинга SOL-P: Преобразование непрерывной функции энергии в многомерный тензор

Число измерений = Число степеней свободы системы

Сетка дискретизации энергии: $n = 2^{16}$ – по каждому измерению



Разложение тензора в тензорный поезд

Многомерный массив (тензор) $A \in \mathbb{R}^{n_1 \times \dots \times n_d}$ может быть представлен в форме:

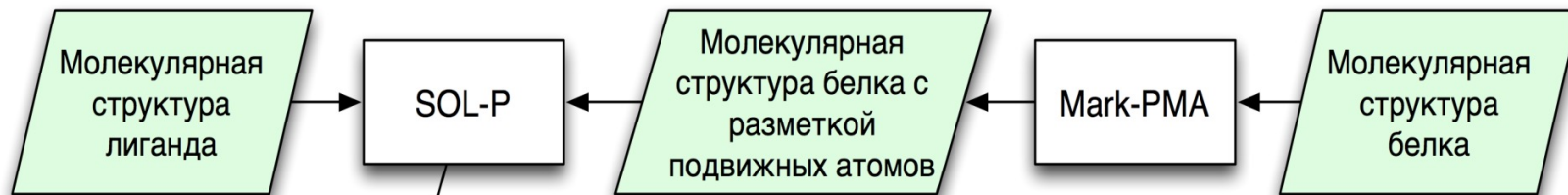
$$A(i_1, \dots, i_d) = \sum_{\alpha_1=1, \dots, \alpha_{d-1}=1}^{r_1, \dots, r_d} G_1(i_1, \alpha_1) G_2(\alpha_1, i_2, \alpha_2) \dots G_{d-1}(\alpha_{d-2}, i_{d-1}, \alpha_{d-1}) G_d(\alpha_{d-1}, i_d)$$

- ▶ r_1, \dots, r_{d-1} - **ТТ**-ранги данного тензора
- ▶ 3х мерные тензоры $G_i \in \mathbb{R}^{r_{i-1} \times n_i \times r_i}$ - вагоны

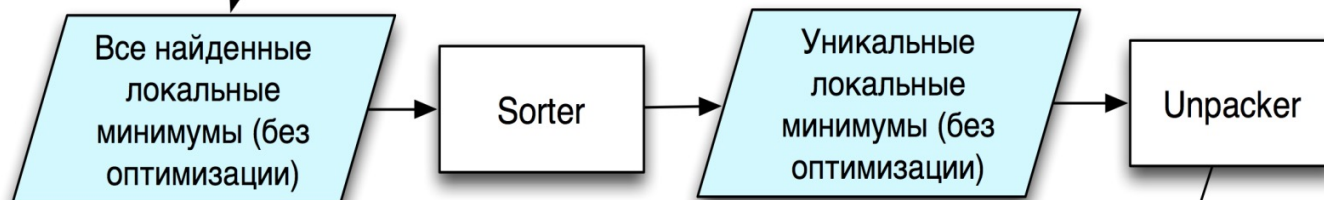
Если тензор получен при дискретизации регулярной гладкой функции – например, энергии комплекса белок-лиганд, то **ТТ** ранги – небольшие числа

Блок-схема программы SOL-P

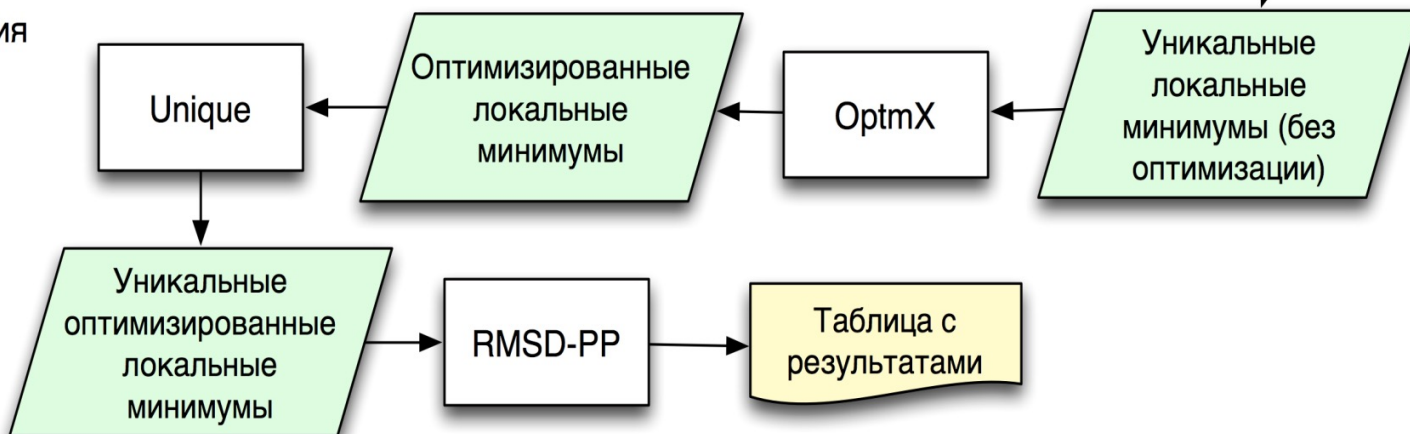
I стадия



II стадия



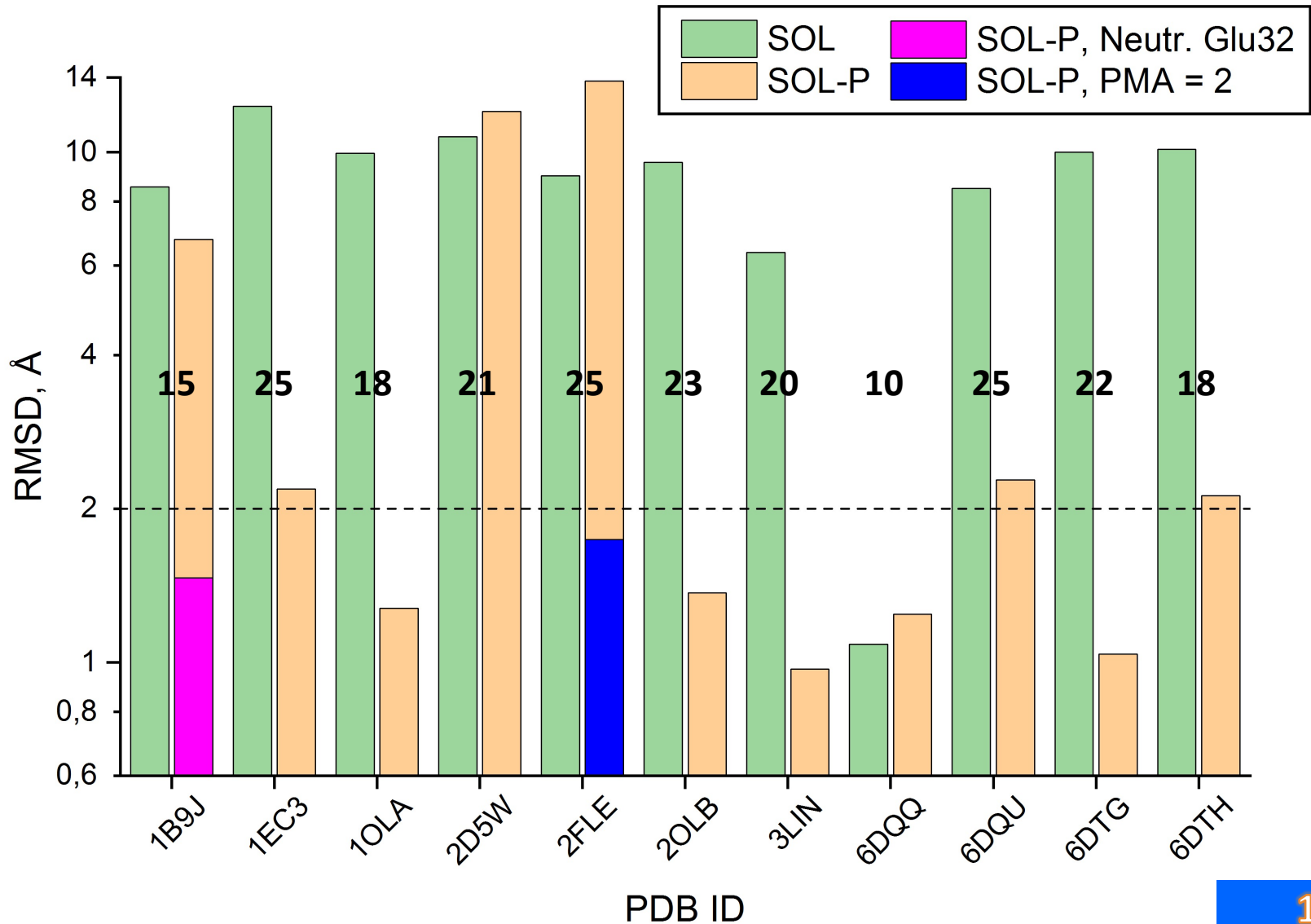
III стадия



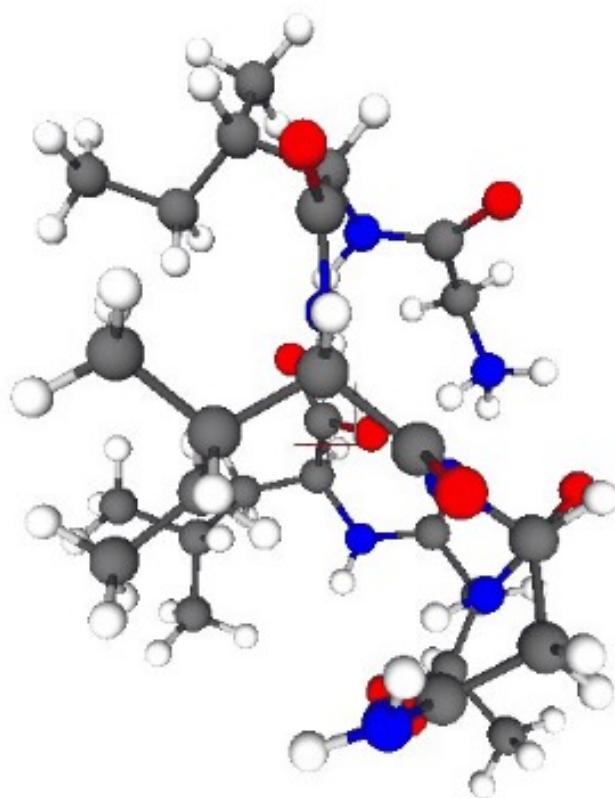
SOL-P суперкомпьютерная программа докинга Tensor Глобальная оптимизация энергии с помощью тензорных поездов Tensor Trains

- ▶ Для значительной части тестовых комплексов SOL-P находит тот же глобальный минимум энергии, который находит FLM
- ▶ Для жесткого белка SOL-P работает гораздо быстрее, чем FLM:
 - SOL-P нужно 100 CPU*часов
 - FLM нужно 10 000 CPU*часов
- ▶ SOL-P успешно проводит докинг лигандов, имеющих до 25 торсионов; для сравнения
 - Для SOL максимальное число торсионов 10-15,
 - Для Glide, GOLD and ICM max торсионов 10
- ▶ SOL-P может проводить докинг гибких лигандов в белки с несколькими десятками подвижных атомов: до 157 степеней свободы белка

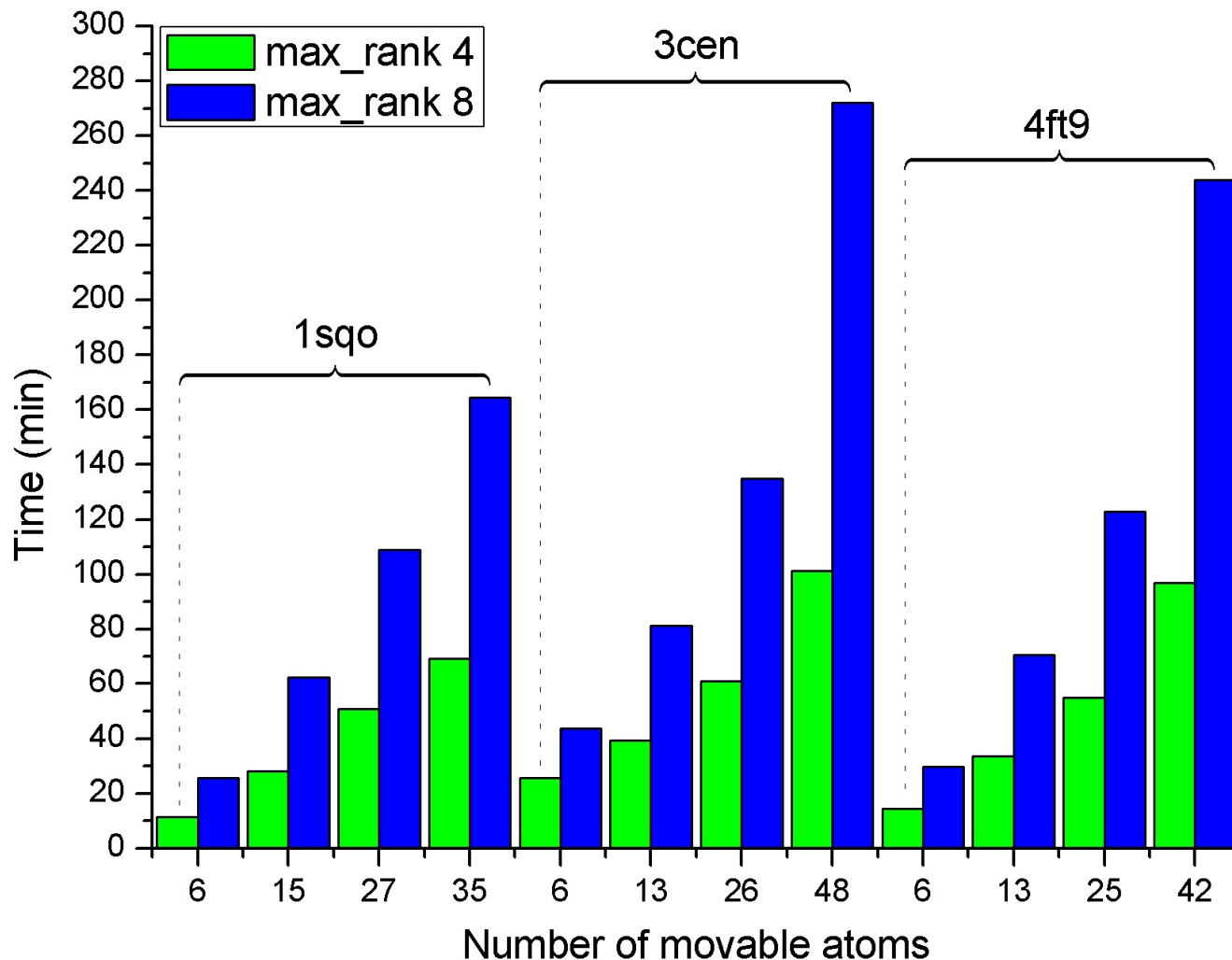
Докинг гибких лигандов



Гибкий олигопептид Gly-Ile-Ile-Asn-Thr-Leu 25 внутренних вращательных степеней свободы



SOL-P: времена докинга 1 лиганда на 256 cores



Наши программы докинга

- SOL_GRID – SOL – докинг на сетке заранее рассчитанных потенциалов. Генетический алгоритм. Находит глобальный минимум. Без локальной оптимизации. Жесткий белок. Это рабочая программа докинга, в том числе для виртуального скрининга больших баз данных лигандов.
- FLM – докинг без сетки. Монте Карло. Локальная оптимизация по атомам лиганда. Жёсткий белок. При наличии достаточных суперкомпьютерных ресурсов находит весь спектр многих тысяч низкоэнергетических локальных минимумов системы белок-лиганд.

Наши программы докинга

- SOL-P - бессеточный докинг. Новый ТТ алгоритм глобальной оптимизации на основе тензорных поездов. Гибкий лиганд и выбранные атомы белка-мишени могут быть подвижны. Локальная оптимизация по всем атомам лиганда и выбранным атомам белка. Время докинга: 100 минут при лиганде с 7 торсионными и 48 подвижными атомами белка на 256 вычислительных ядрах

Источники информации о трехмерной структуре малых молекул:

- Рентгеноструктурный анализ, Спектроскопия ЯМР
- Информация о трехмерных структурах малых молекул собрана в Кембриджской базе данных – Cambridge Structural Database (CSD)
<http://www.ccdc.cam.ac.uk/>
- Protein Data Bank - PDB: www.rcsb.org/pdb

Однако, если мы разрабатываем новое лекарственное средство, то оно еще не получено, и следовательно, мы не можем использовать экспериментальную трехмерную структуру!
- Самый верный способ правильно найти 3D структуру лиганда – это провести *ab initio* квантово-химический расчет
- А если таких лигандов тысячи, что делать?

Какие молекулы могут быть лекарствами?

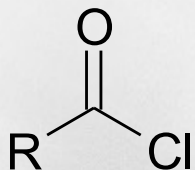
- Одна молекула может быть в различных пространственных 3D-конформациях
- Изомеры – молекулы одинакового химического состава, но разной структуры
- Таутомерия – обратимая изомерия: изомеры легко переходят друг в друга
- Растворимость в воде
- А если молекула в воде нестабильна и распадается на части?
- Коэффициент распределения октанол-вода
- Corina – преобразование молекул 2D -> 3D
- Зарядовые состояния молекулярных групп – основные идеи, протонирование

Не все органические вещества могут быть потенциальными лекарствами

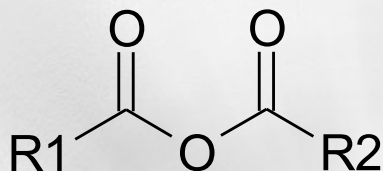
Причины, по которым вещество не может быть лекарством:

- **Вещество нестабильно в водных растворах или в присутствии компонентов плазмы крови**
- **Вещество вступает в неспецифичные химические реакции с биологическими субстратами организма**
- **Вещество нерастворимо или очень плохо растворимо в воде – не удастся достичь терапевтических концентраций**
- **Вещество обладает плохими фармакокинетическими характеристиками (ADME – Adsorption, Distribution, Metabolism, Excretion)**
- **Вещество токсично, канцерогенно и т.д.**
- **Содержат тяжелые или токсичные элементы – свинец, таллий, ртуть, сурьму, мышьяк и т.д.**

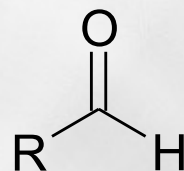
Вещества, нестабильные в водных растворах или в присутствии компонентов плазмы крови, вступающие в неспецифичные реакции с биологическими субстратами:



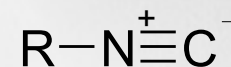
Хлорангидриды



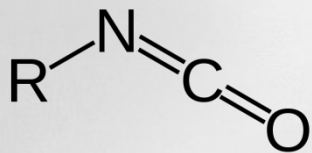
ангидриды



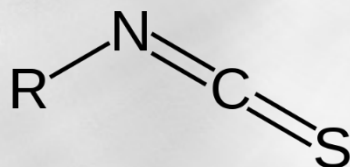
альдегиды



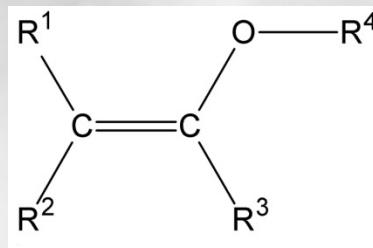
изоцианиды



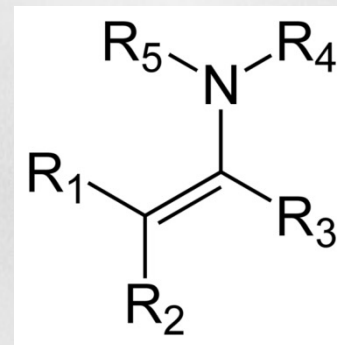
Изоцианаты



Изотиоцианаты



эфиры енолов

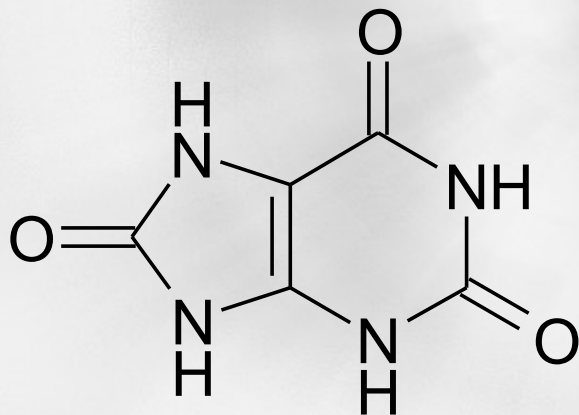


енамины

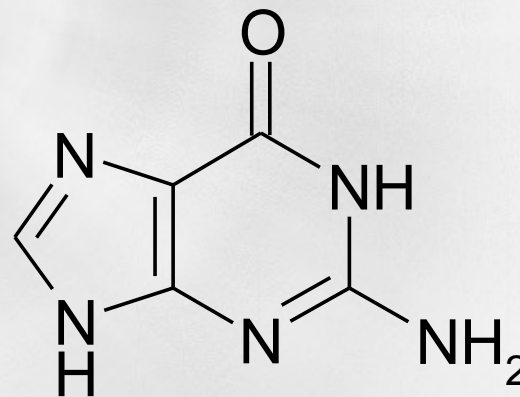
И многие другие реакционноспособные органические соединения

Растворимость в воде – проблема!

1. Следует избегать молекул составленных лишь из неполярных функциональных групп, таких как углеводородные, незамещенная ароматика и т.д.
2. Однако даже наличие в составе молекулы большого количества «гидрофильных» групп не означает, что вещество хорошо растворяется. Например, следующие вещества практически нерастворимы в воде:



Мочевая кислота

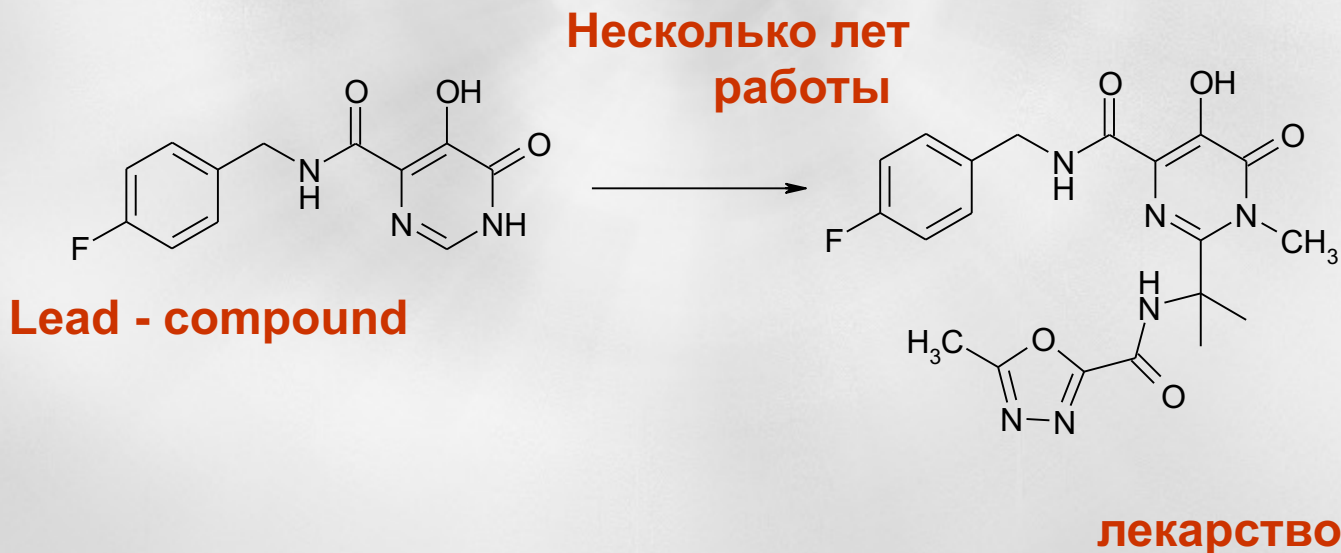


Гуанин

Растворимость в воде – проблема!

- Предсказать заранее растворимость вещества достаточно сложно, так как для этого необходимо оценить изменение свободной энергии молекулы при переходе из кристаллического состояния в растворенное:
$$\Delta G = (G_{\text{solv}} - G_{\text{cryst}}).$$
- В свою очередь, свободная энергия молекулы в кристаллическом состоянии зависит от упаковки молекул в кристалле, предсказать которую практически невозможно.
- Более того, есть лекарственные средства, существующие в нескольких полиморфных кристаллических модификациях, каждая из которых характеризуется своим параметром растворимости в воде!

Большинство биологически-активных веществ, найденных в процессе поиска новых лекарственных веществ (**Lead compounds**) обладают недостаточной растворимостью. В процессе оптимизации **lead compounds** обычно встает задача повышения растворимости для создания из lead compound настоящего лекарственного средства. Обычно это достигается введением гибких гидрофильных групп в молекулу, но так, чтобы требуемая биологическая активность молекулы не ухудшилась. Этот процесс – своего рода искусство ☺



Приемлемые фармакокинетические характеристики вещества (ADME – Adsorption, Distribution, Metabolism, Excretion) – поглощение, распределение, метаболизм, выведение.

В 1997 году Кристофер Липински (Christopher Lipinski, Pfizer) проанализировав структуры известных пероральных лекарственных средств получил 4 правила, которым должны подчиняться структуры молекул, чтобы с большой долей вероятности иметь приемлемые параметры ADME и быть лекарственно-подобными. Эти 4 правила называются Правилами Пяти

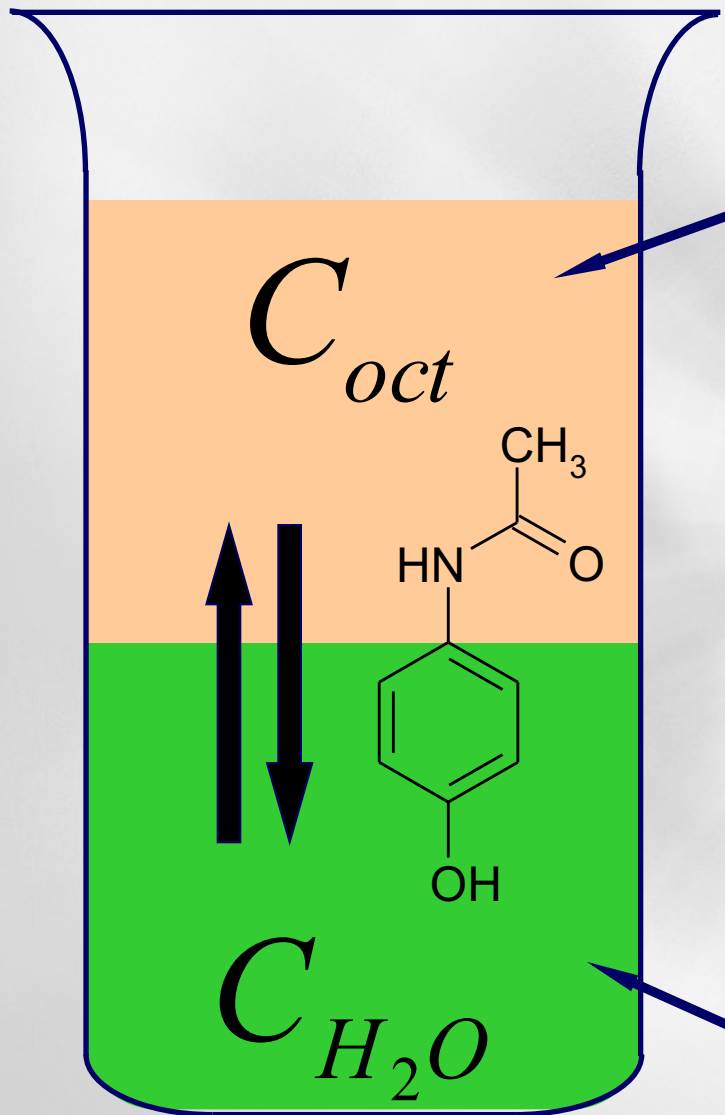
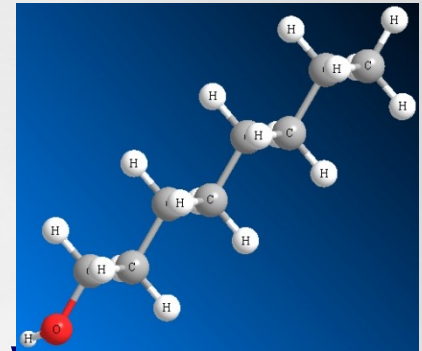
Правила Липинского

Lipinski's Rule-of-Five (Ro5)

Молекула перорального лекарства должна удовлетворять правилам.

1. Молекула должна содержать не более 5 доноров водородной связи
2. Молекула должна содержать не более 10 акцепторов водородной связи
3. Молекулярный вес вещества не должен превышать 500
4. Коэффициент распределения октанол-вода в пределах:
 $\log P < 5$ (распределение в системе октанол-вода имитирует распределение вещества в бислоидной липидной мембране – вещества с указанным $\log P$ способны проходить через такие мембраны и проникать внутрь клеток)

Что такое коэффициент распределения октано-вода (LogP) и почему он так важен?



N-октанол
(гидрофобная среда)

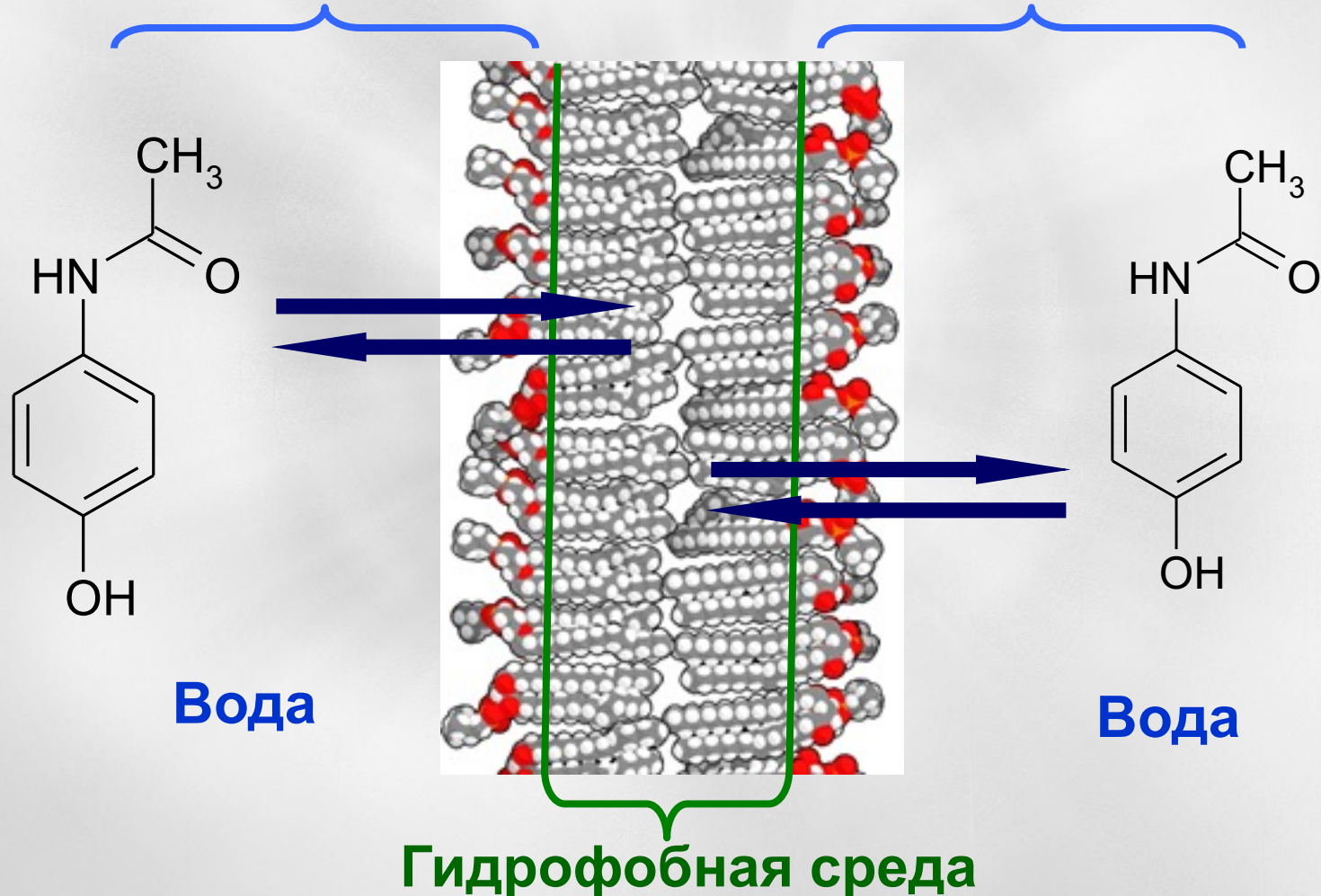
$$\log P = \log_{10} \frac{C_{oct}}{C_{H_2O}}$$

Вода
(гидрофильная среда)

$\log P$ характеризует процесс диффузии лекарственного вещества через бислоюную липидную мембрану. Для эффективной диффузии необходимо, чтобы вещество находилось в достаточных концентрациях как в гидрофильной, так и в гидрофобной средах.

Гидрофильная среда

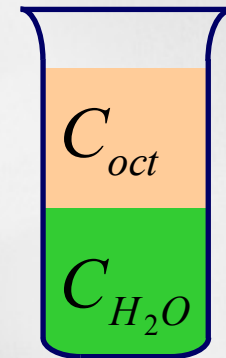
Гидрофильная среда



Способы определения LogP

• Экспериментальное измерение

(смешивание, встряхивание, определение концентрации в обеих фракциях, хроматографические методы)



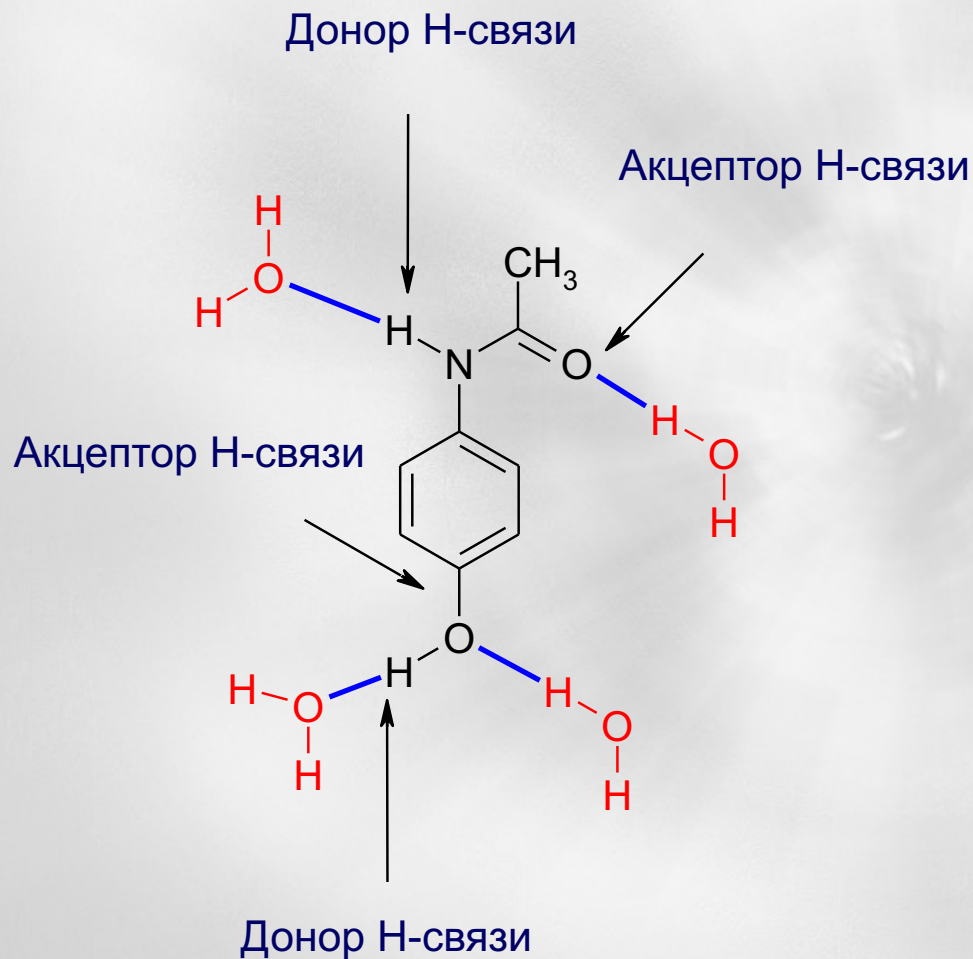
• Расчетные методы

Программы для расчета LogP (например, CLogP)

A.Leo *Chem. Rev.* 1993, **93**, 1281-1308



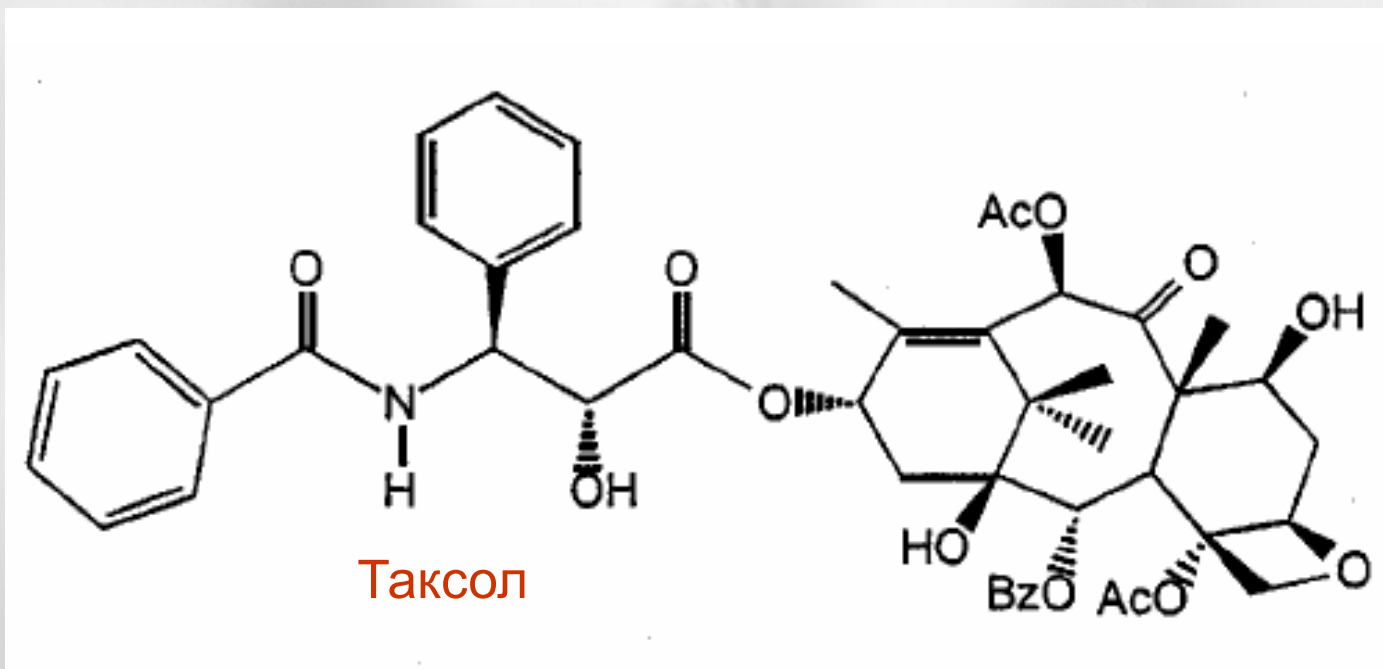
Рассмотрим выполнение правил Липински на примере молекулы простого лекарственного вещества – парацетамола (панадол)



1. Количество доноров Н-связи: 2 (проходит!)
2. Количество акцепторов Н-связи: 2 (проходит!)
3. Молекулярная масса: $M = 151.16$ (проходит!)
4. $\text{Log } P = 0.34$ (проходит!)

Итак, молекула парацетамола удовлетворяет всем правилам Липински и может гипотетически рассматриваться, как лекарственное средство!

Правила Липински – не физические законы, а эмпирические, поэтому возможны и исключения, например – **один из лидеров продаж среди лекарственных веществ – таксол** – противоопухолевый препарат (**паклитаксел** компании Bristol-Myers Squibb) – имеет молекулярную массу $M=854$, что **нарушает 3 правило**: $M < 500$. Существуют и более точные способы оценки лекарственно-подобия органических веществ, однако они достаточно сложны или требуют привлечения экспериментальных методик

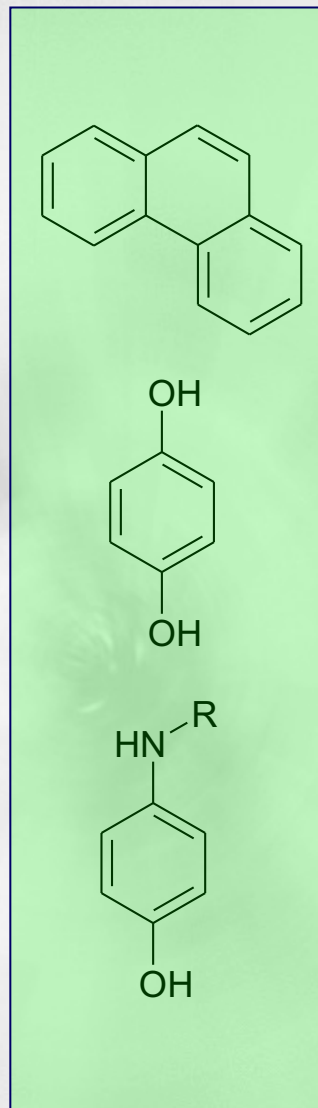


Токсичность.

Многие вещества в процессе химических реакций (метаболизма) в организме человека могут превращаться в токсичные метаболиты.

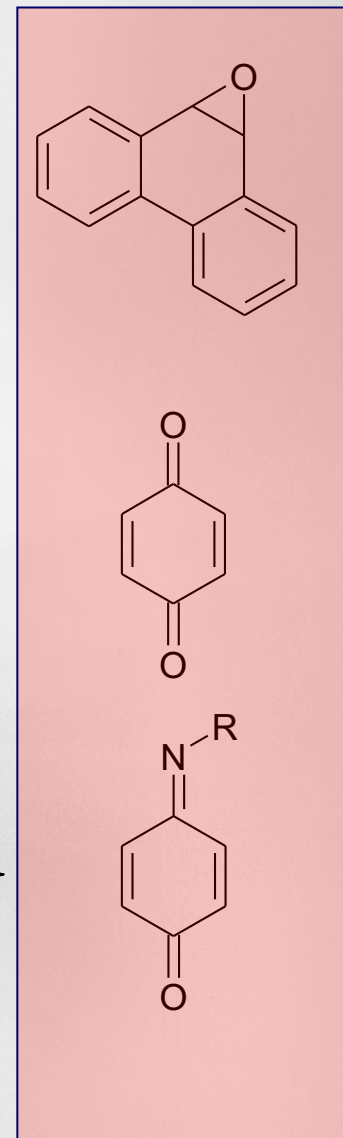
Известно несколько молекулярных мотивов, способных претерпевать такие нежелательные превращения.

Их необходимо избегать!



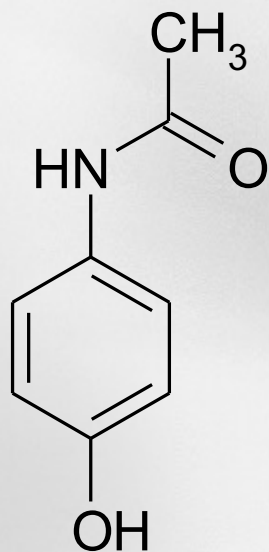
Безвредные
исходные
вещества

→
Превращения,
осуществляемые
→
ферментами
печени человека
→

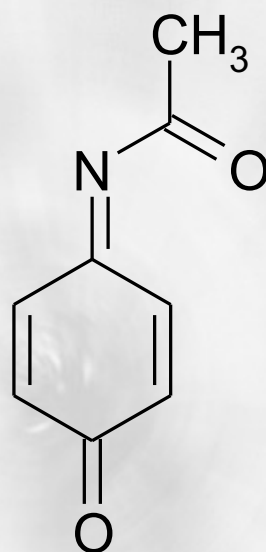
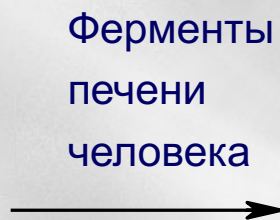


Токсичные
метаболиты

Можно заметить, что структура парацетамола такова, что может проявлять нежелательную **ТОКСИЧНОСТЬ**:



Парацетамол

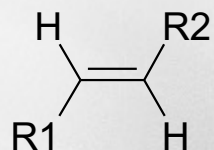


Токсичный
хинонимин

Действительно, показанное превращение обуславливает высокую токсичность парацетамола. Парацетамол довольно токсичен – в дозах выше 200 мг/кг вызывает необратимые повреждения печени и почек. Для кошек он – сильный яд. Если бы он разрабатывался сейчас, а не 100 лет назад – то был бы снят с испытаний из-за токсичности!

Итак, чтобы стать лекарством, вещество должно одновременно удовлетворять многим требованиям – по растворимости, фармакокинетике и токсичности!

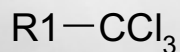
Нежелательные структуры, которые могут привести к общей токсичности лекарственного вещества



алкены



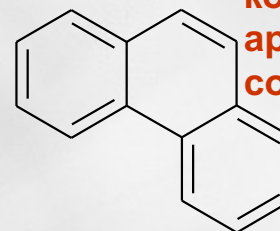
эпоксиды



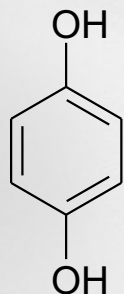
полигалоген-
производные



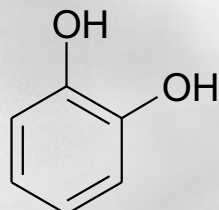
нитрозо-
соединения



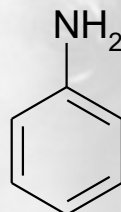
конденсированные
ароматические
соединения



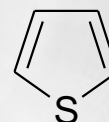
гидрохиноны



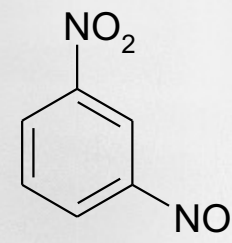
катехолы



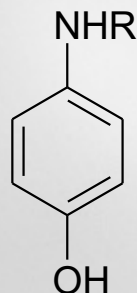
незамещенные
ароматические
амины



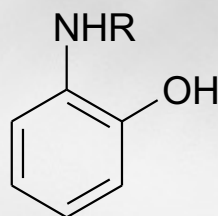
тиофены



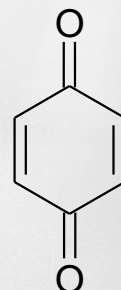
полинитросоединения



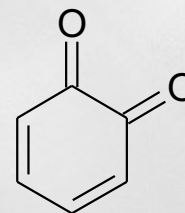
пара-аминофенолы
Лекция № 7



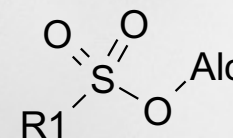
орто-аминофенолы



пара-хиноны



орто-хиноны



алкил-сульфаты

Спасибо за внимание

- *...Surely every medicine is an innovation; and he that will not apply new remedies, must expect new evils...*
- *...Каждый медицинский метод есть инновация; а кто не хочет применять новые средства, должен ждать новых бед...*

Sir Francis Bacon (1561-1626)



OF INNOVATIONS