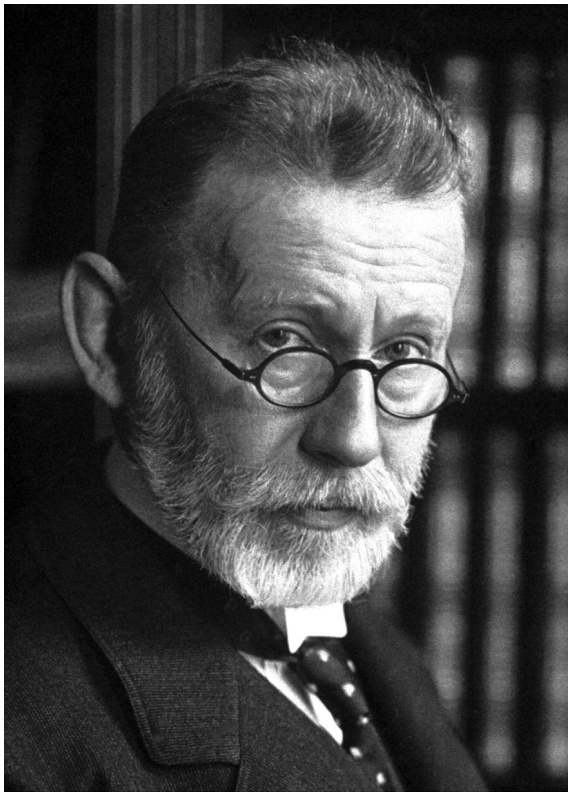
Молекулярное моделирование

Методы и применение

1. Введение
2. Теоретические основы
3. Моделирование малых молекул
4. Моделирование больших молекул
5. Моделирование комплексов
6. Дизайн комплексов

Лекарство

Химическое соединение, которое позволяет остановить патологический процесс в организме, не нарушая течение нормальных процессов (и не убивая сам организм)



Пауль Эрлих (1854–1915)

Низкомолекулярные лекарства (MW ~ 300÷700):

- органические соединения
- небольшие пептиды

Высокомолекулярные лекарства:

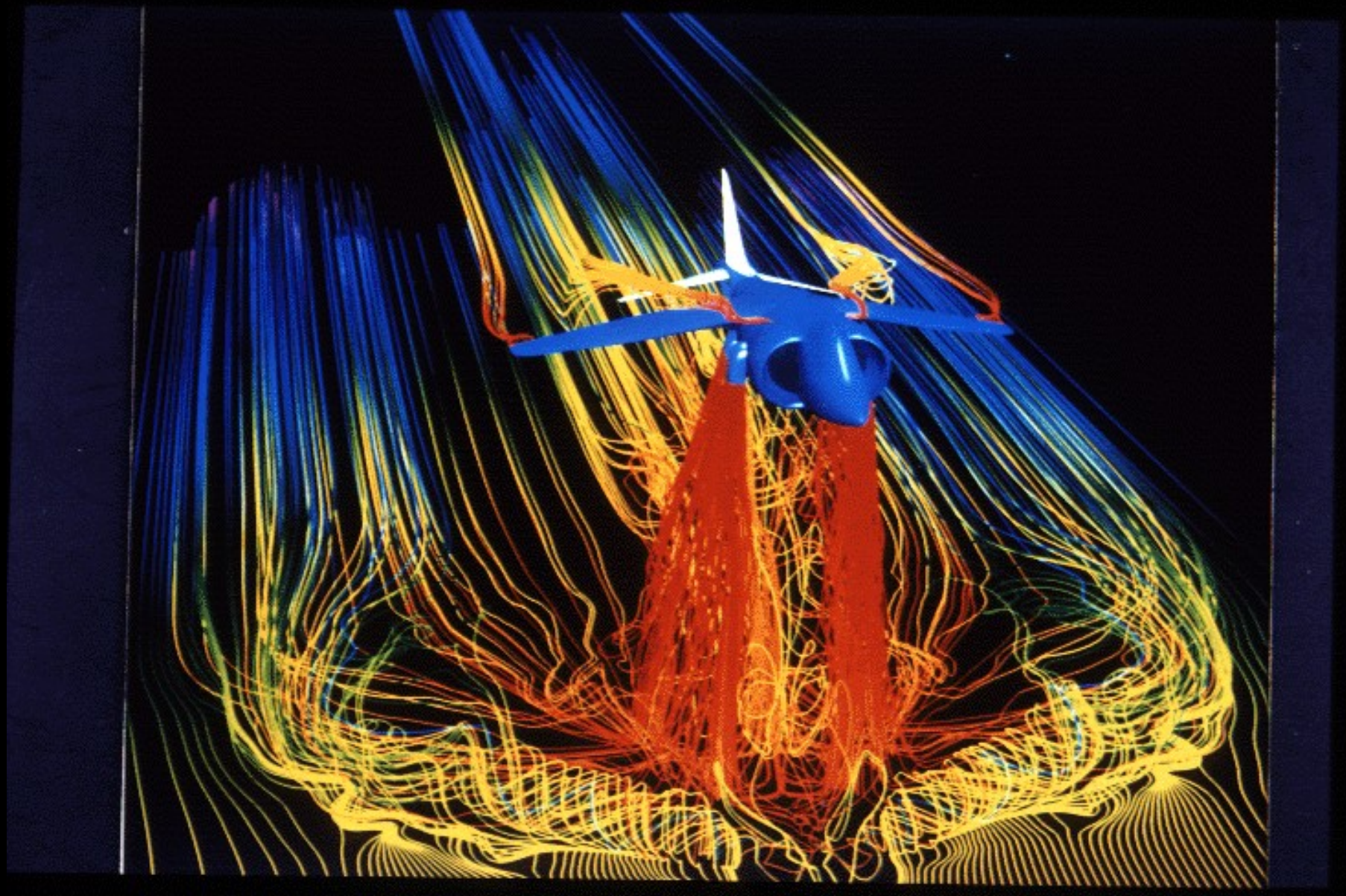
- полимеры
- белки
- нуклеиновые кислоты
- антитела

Концепция *волшебной пули*: каждое лекарство действует на определённую мишень в организме.

www.calxibe.com

http://www.1024x768wallpapers.com

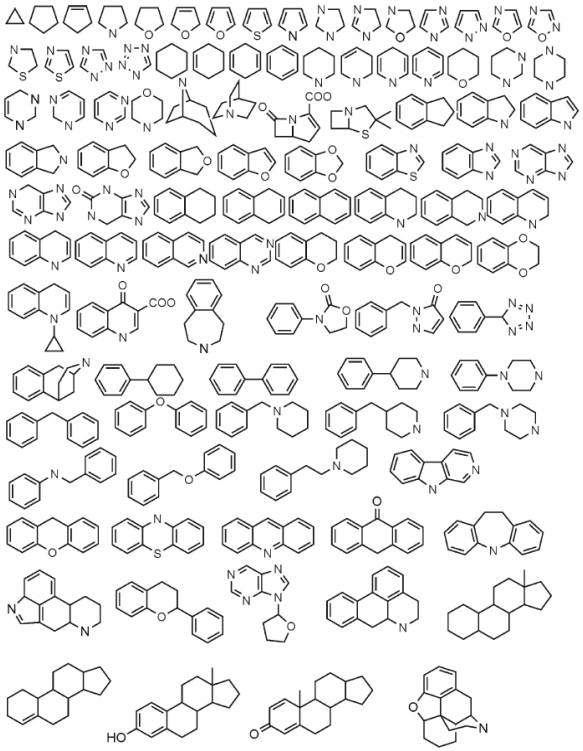
Создание самолёта

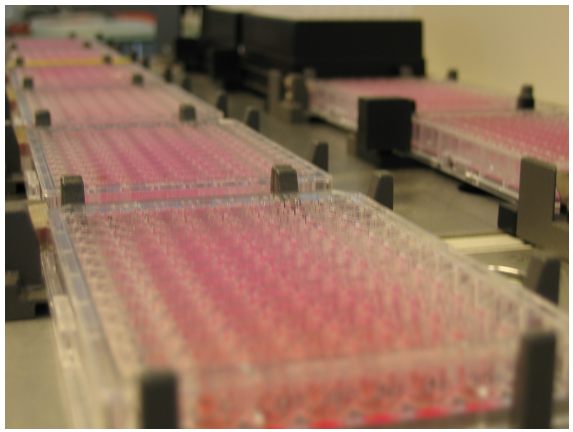


W. S. Woltosz, JCAMD, 2012, 26, 159-163

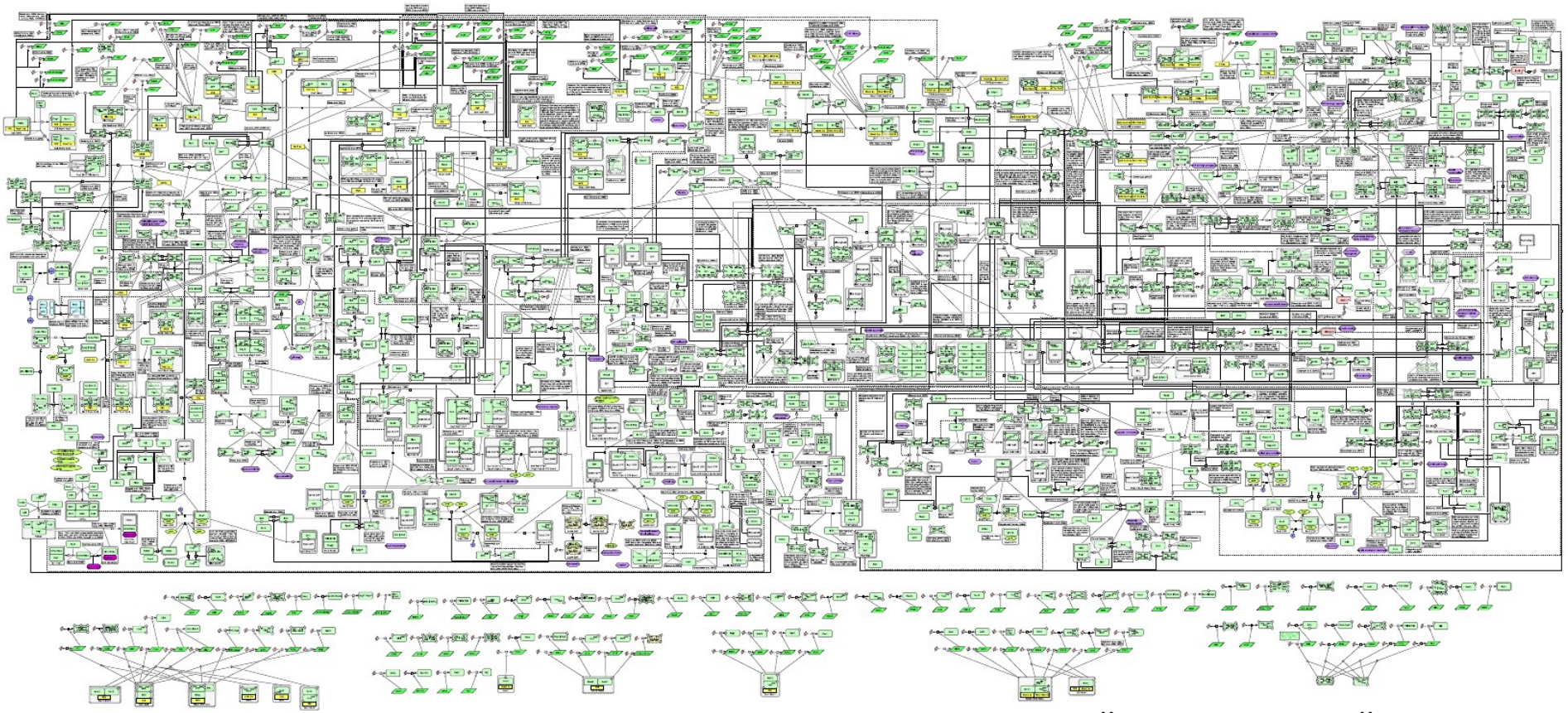
Создание лекарства без использования моделирования

- **Синтез** тысяч соединений
- **Высокопроизводительный скрининг** — поиск *хита*
- **Оптимизация хита** — классическая медицинская химия, синтез сотен соединений
- **Тестирование в клеточных системах,** определение цитотоксичности
- **Тестирование *in vivo*** — часто с использованием нескольких видов животных
- **Клинические испытания** — многолетний процесс с привлечением здоровых и больных добровольцев



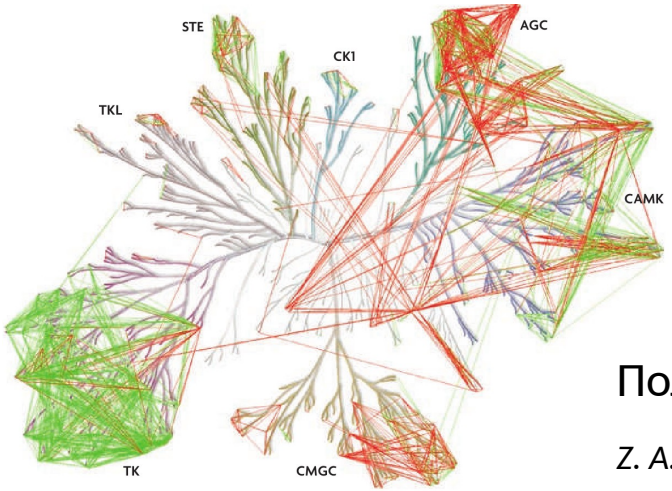


Почему так?



Клеточный цикл дрожжей

K. Kaizu et al., Molecular Systems Biology, 2010, 6, 415

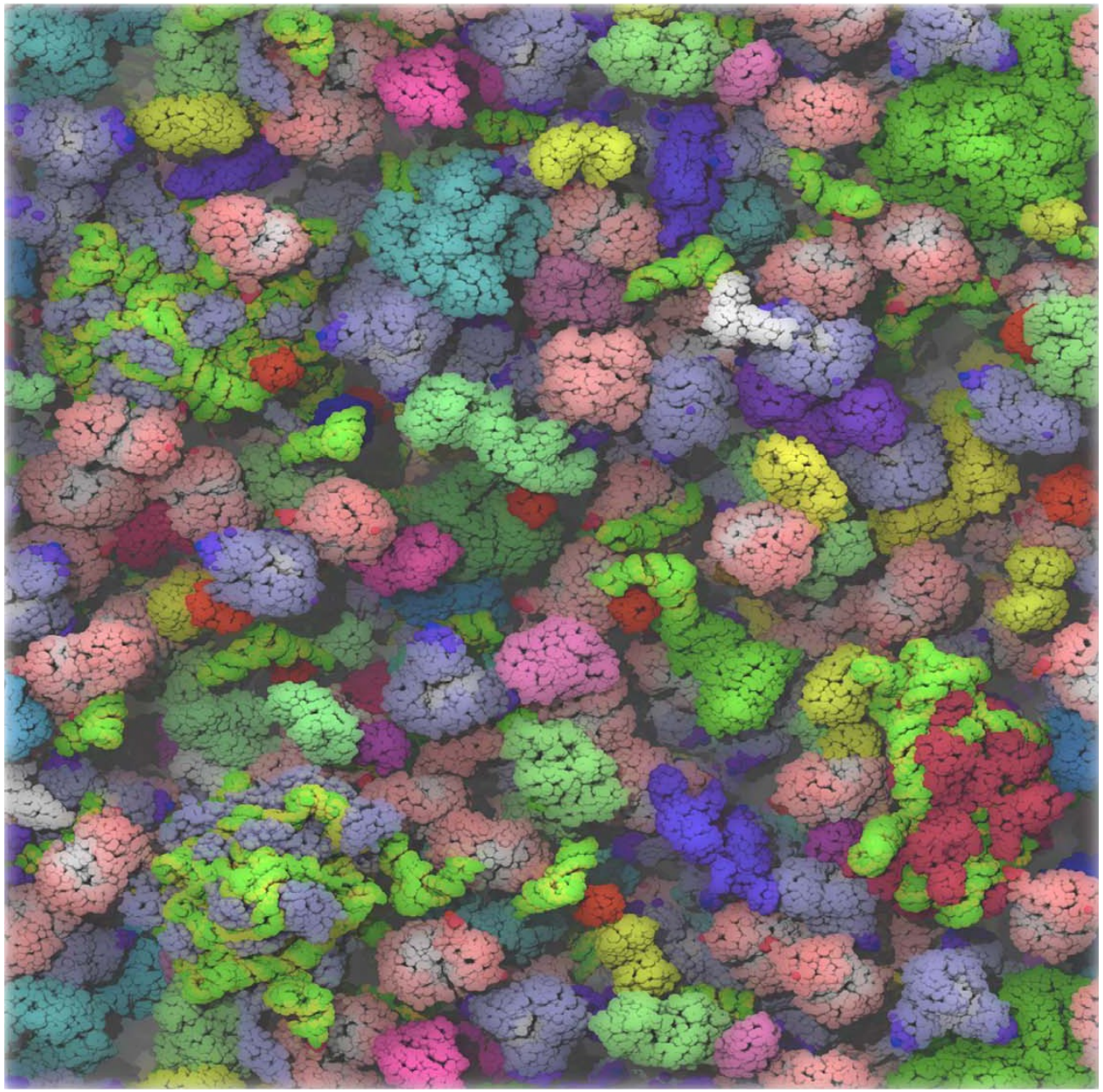


Кином — совокупность всех киназ организма

Полифармакология кинома человека

Z. A. Knight et al., Nat. Rev. Cancer, 2010, 10, 130-137

Модель цитоплазмы *E. coli*



PLoS Comp. Biol., 2010, 6, e1000694

Суть термина

Расчётные методы, применяемые для моделирования поведения молекул, и их теоретическая основа

Область применения

- Структура малых молекул
- Структура макромолекул — белков и нуклеиновых кислот
- Изучение межмолекулярных взаимодействий
- Изучение динамического поведения молекул
- Изучение реакционной способности
- Молекулярный дизайн

The Nobel Prize in Chemistry 2013

© Nobel Media AB

Martin Karplus

Photo: Keilana via Wikimedia Commons

Michael Levitt

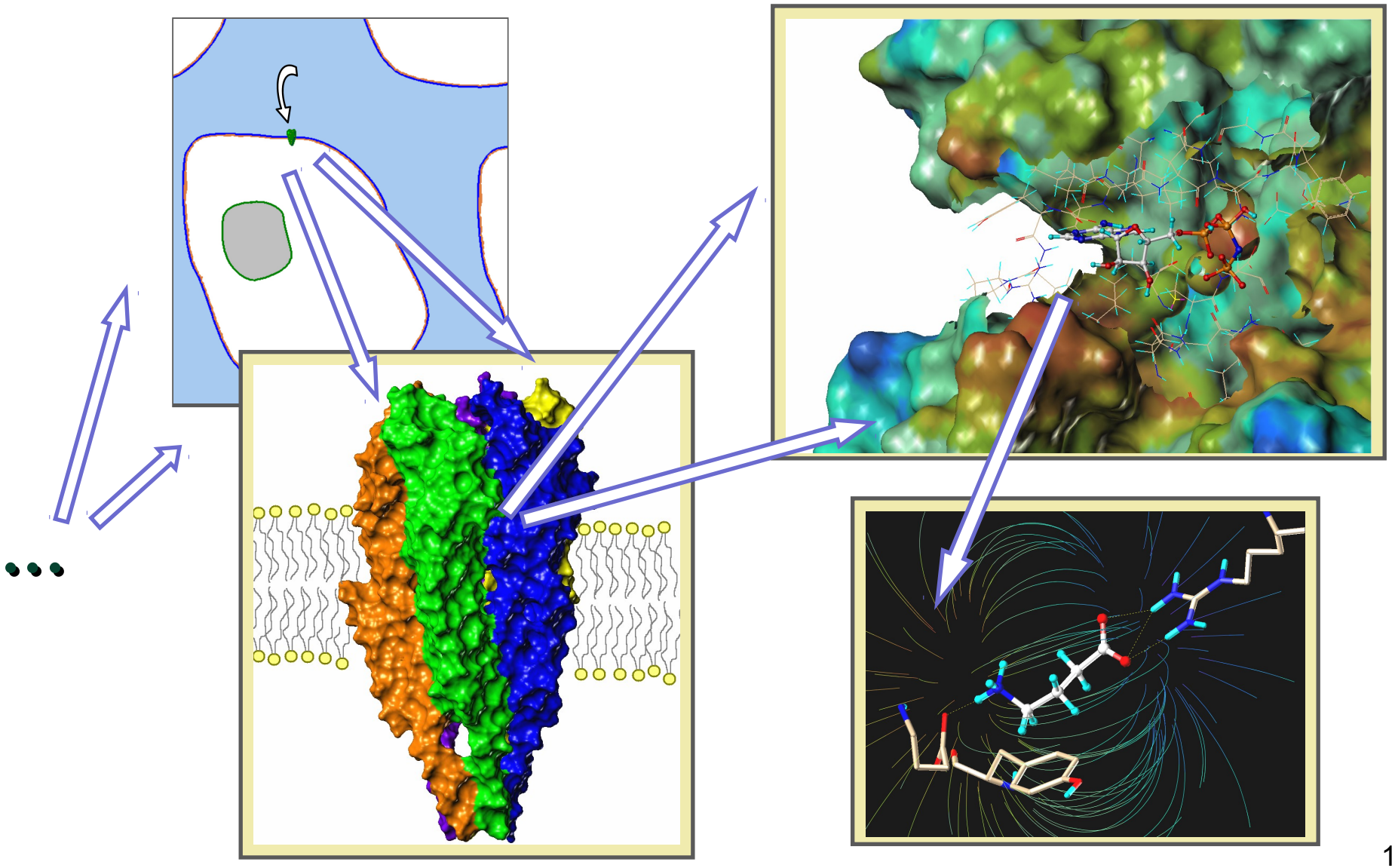
Photo: Wikimedia Commons

Arieh Warshel

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel ***"for the development of multiscale models for complex chemical systems".***

Share | Tell a Friend | Comments

Иерархия характерных времён и масштабов на примере живых систем



1. Введение

2. Теоретические основы

3. Моделирование малых молекул

4. Моделирование больших молекул

5. Моделирование комплексов

6. Дизайн комплексов

Молекулярная механика

Приближение Борна — Оппенгеймера:
пренебрегаем электронами, изучаем лишь движение атомов

$$V = V(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N) = V(x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_N, y_N, z_N)$$

Ключевые предположения:

- Аддитивность
- Переносимость

Определение **силового поля** включает:

- Набор формул, связывающих компоненты потенциальной энергии молекулы с её геометрическими параметрами;
- Числовые значения параметров, входящих в эти формулы

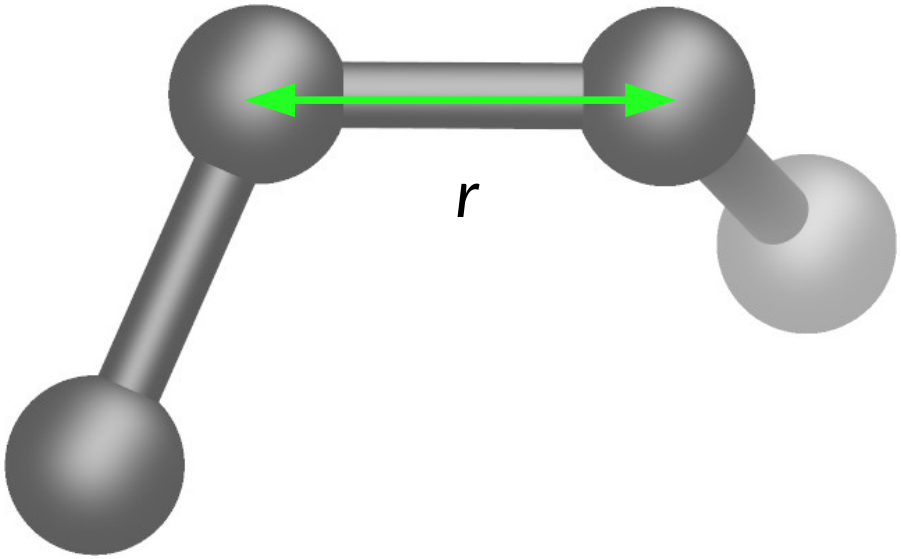
Компоненты потенциальной энергии молекулы

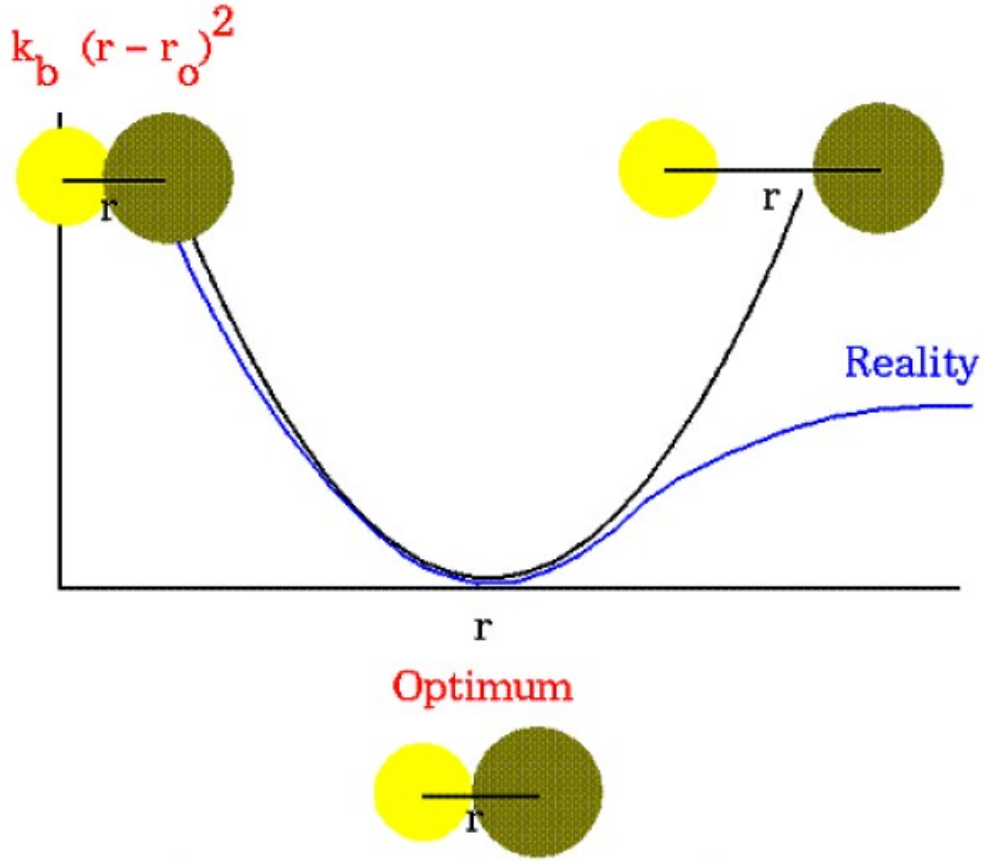
$$V = V_b + V_\theta + V_\tau + V_{vdw} + V_{el} + \ldots$$

- V_b — энергия растяжения связей (закон Гука)
- V_θ — энергия изгиба валентных углов
- V_τ — торсионный потенциал
- V_{vdw} — энергия взаимодействия между несвязанными атомами (ван-дер-Ваальсово взаимодействие)
- V_{el} — энергия электростатического взаимодействия (закон Кулона)

Энергия растяжения связей

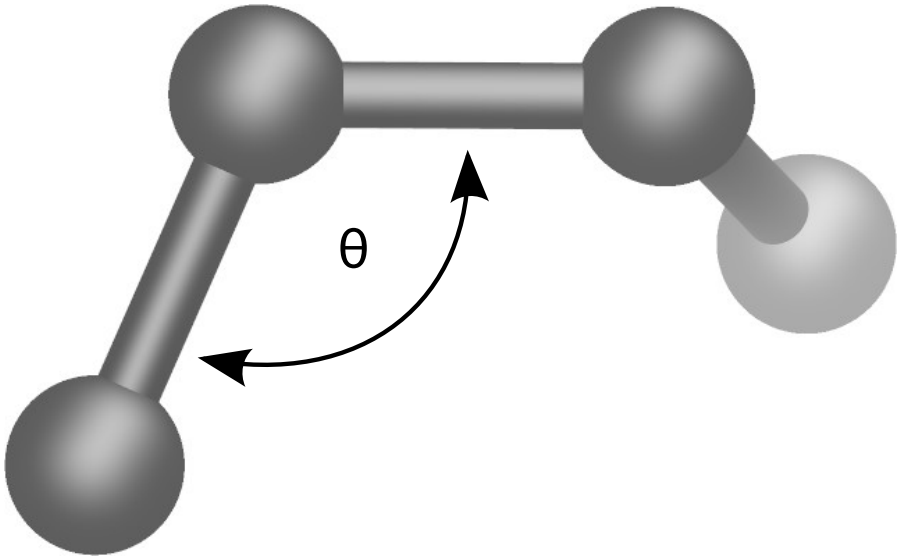
Закон Гука: $$V_b = \sum_{bonds} k_b (r - r_0)^2$$

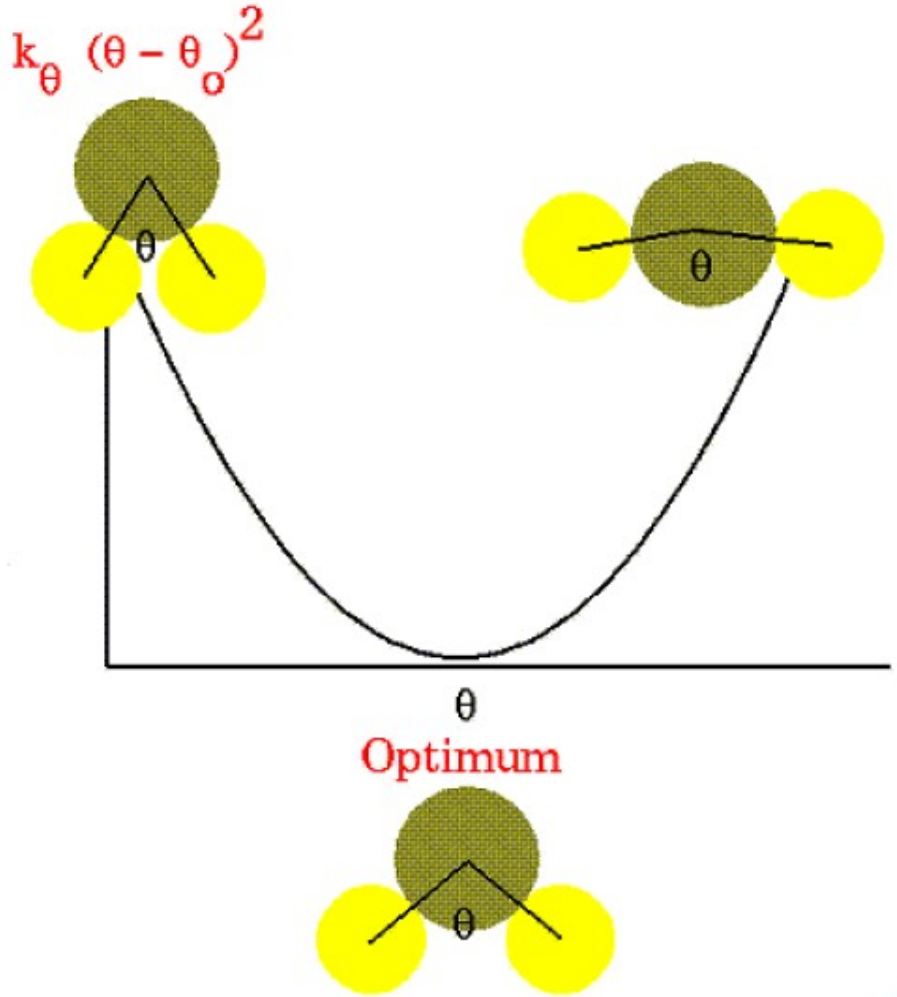




Энергия изгиба валентных углов

Закон Гука: $$V_\theta = \sum_{angles} k_\theta (\theta - \theta_0)^2$$

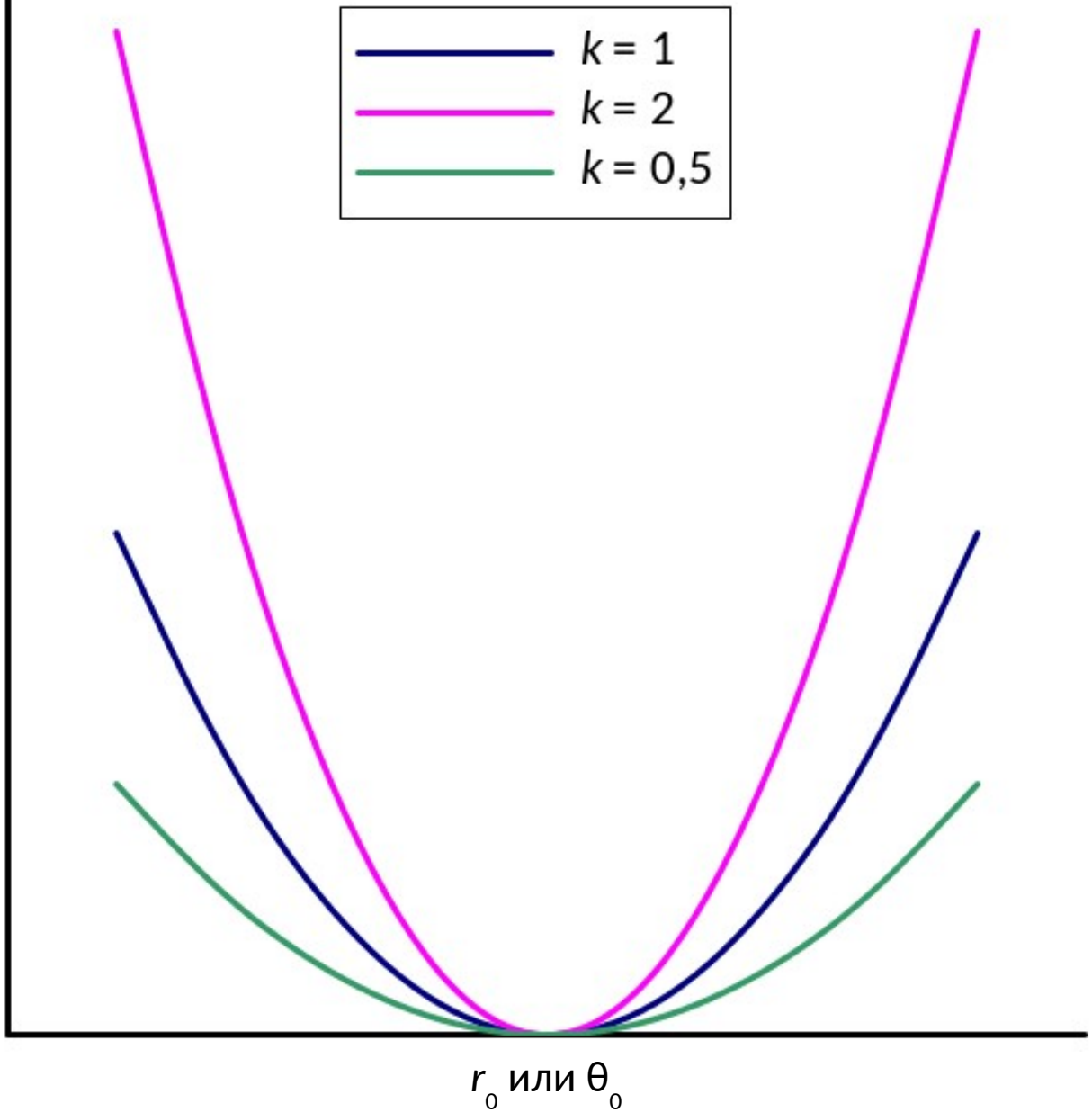




Потенциал Гука: зависимость от параметров

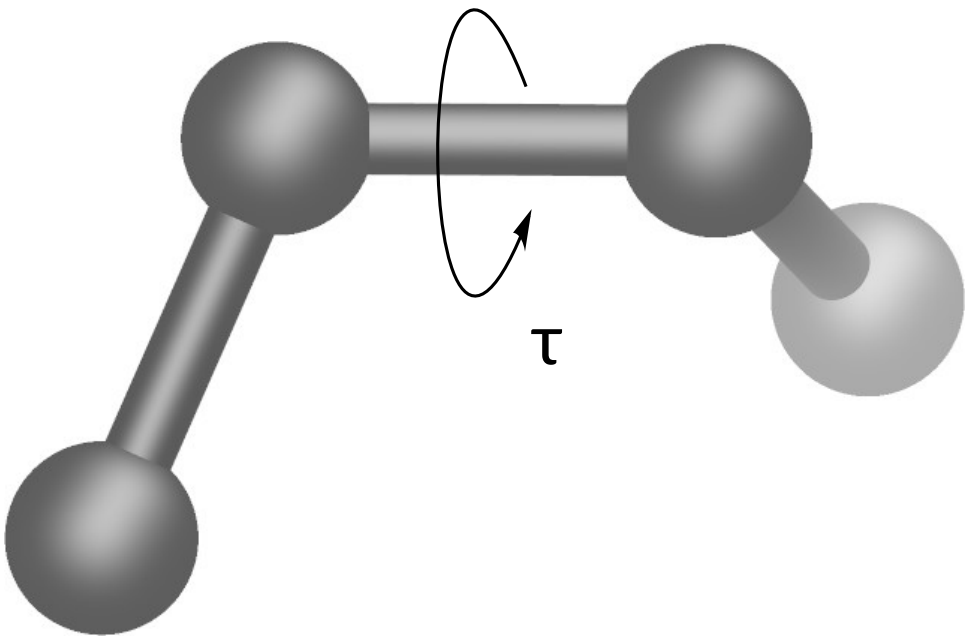
$$k_\theta(\theta-\theta_0)^2$$

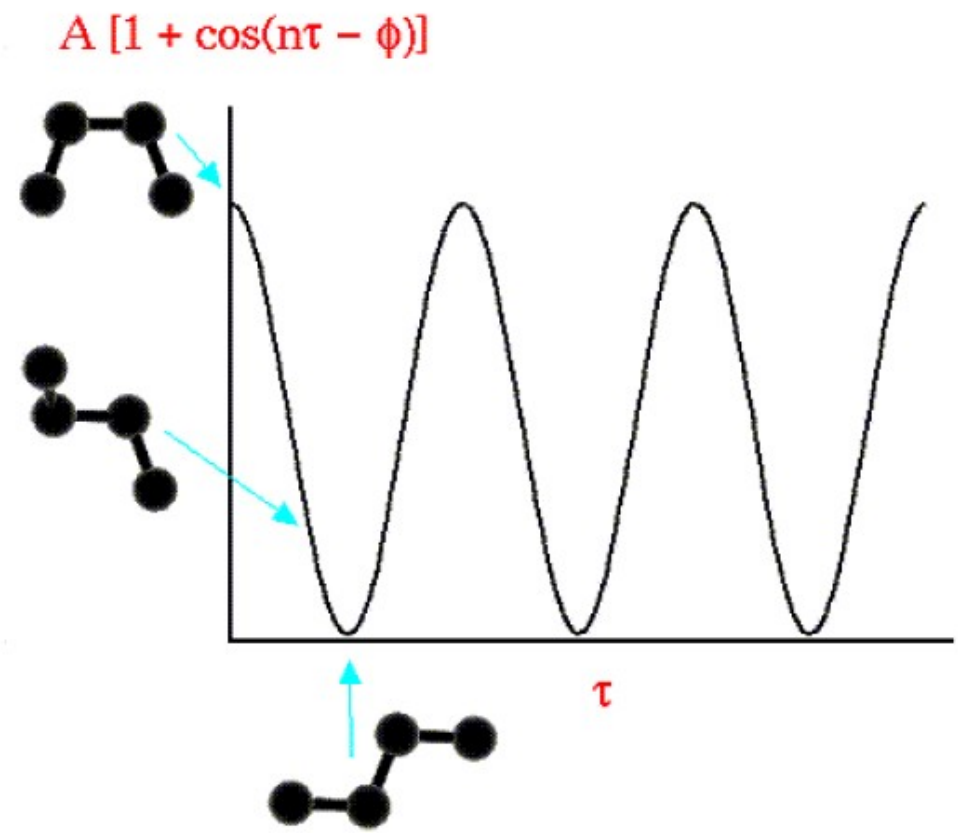
$$k_b(r-r_0)^2$$



Торсионная энергия

$$V_\tau = \sum_{torsions} A\left[1 + \cos\left(n\tau - \varphi\right)\right]$$





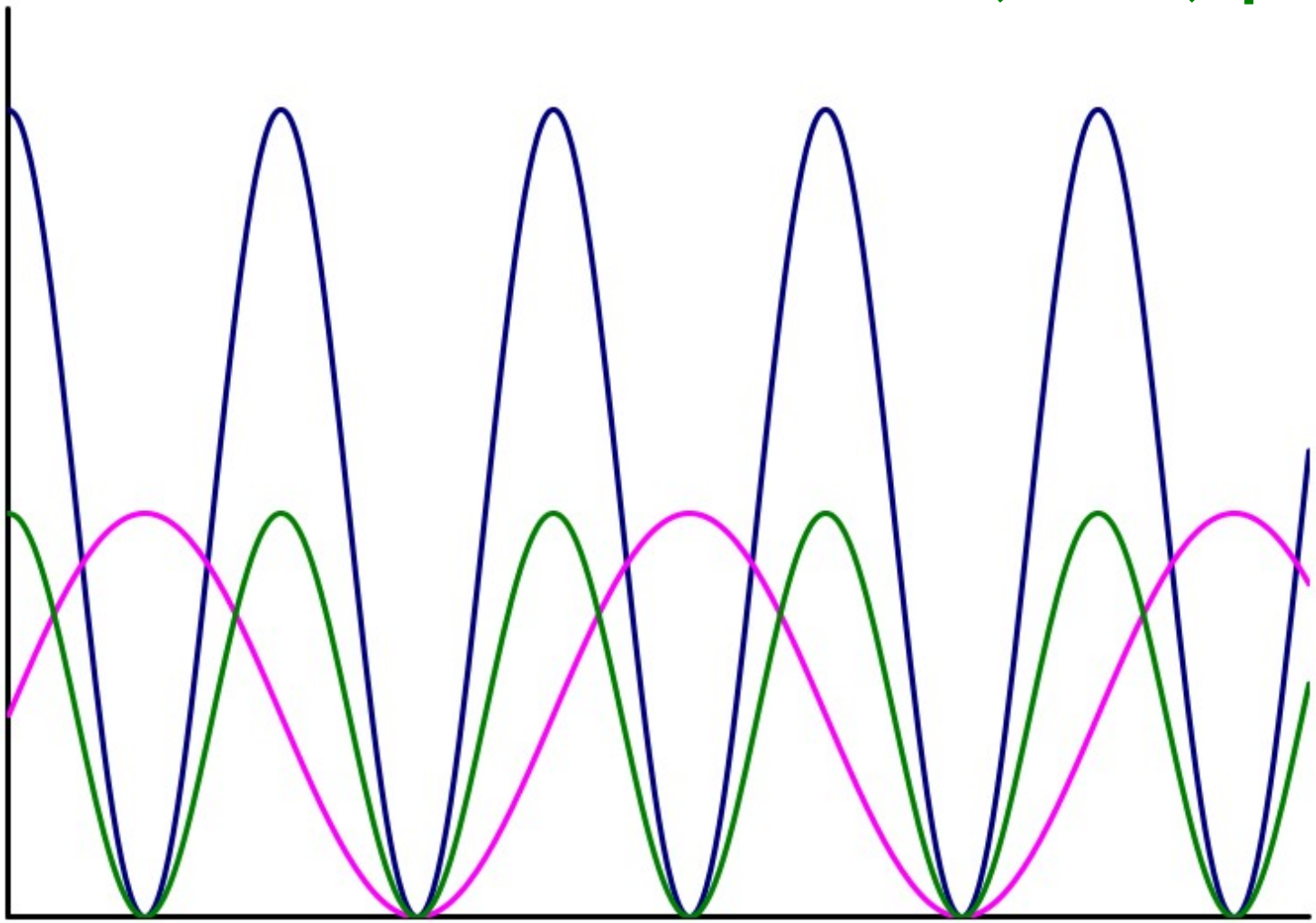
Зависимость торсионного потенциала от параметров

$$A\left[1+\cos\left(n\tau-\varphi\right)\right]$$

***A* = 2, *n* = 2, φ = 0**

***A* = 1, *n* = 1, φ = π/2**

***A* = 1, *n* = 2, φ = 0**

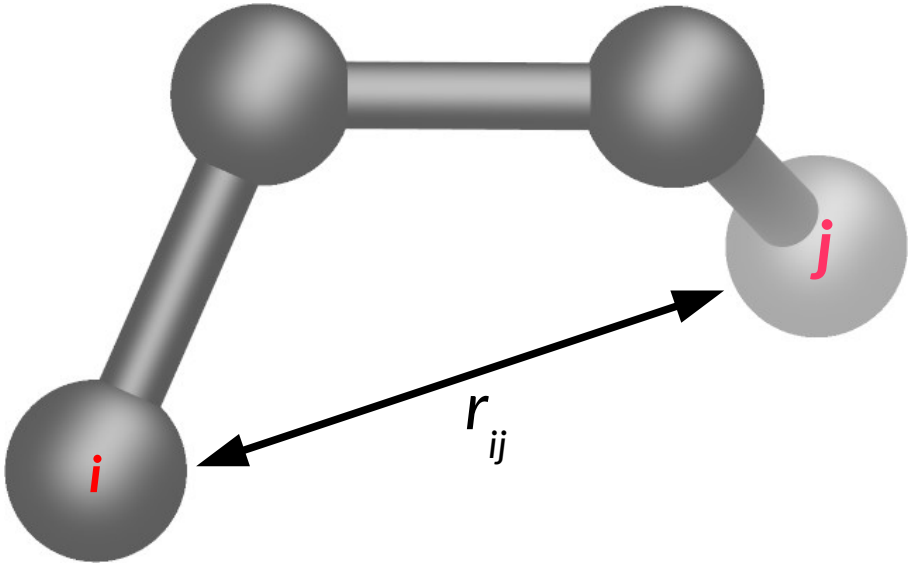


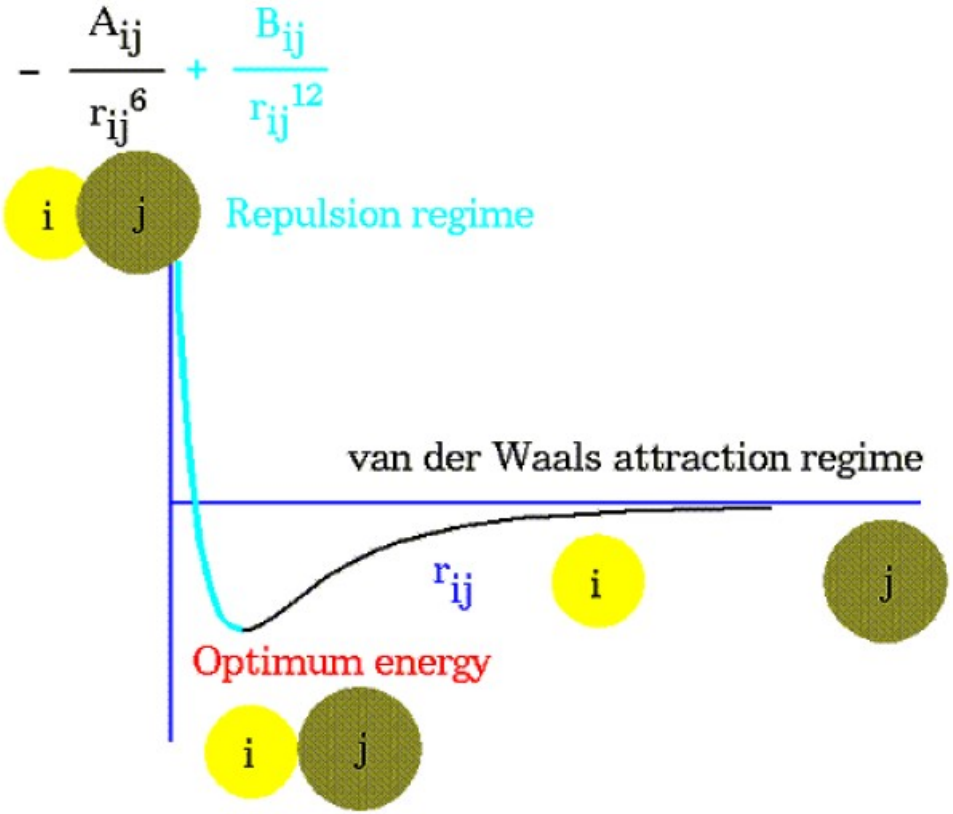
Энергия взаимодействия между несвязанными атомами

$$V_{nb} = \sum_{i,j} \left(-\frac{A_{ij}}{r_{ij}^{6}} + \frac{B_{ij}}{r_{ij}^{12}} \right) + \sum_{i,j} \frac{q_i q_j}{r_{ij}}$$

Потенциал Леннарда-Джонса

Электростатический потенциал



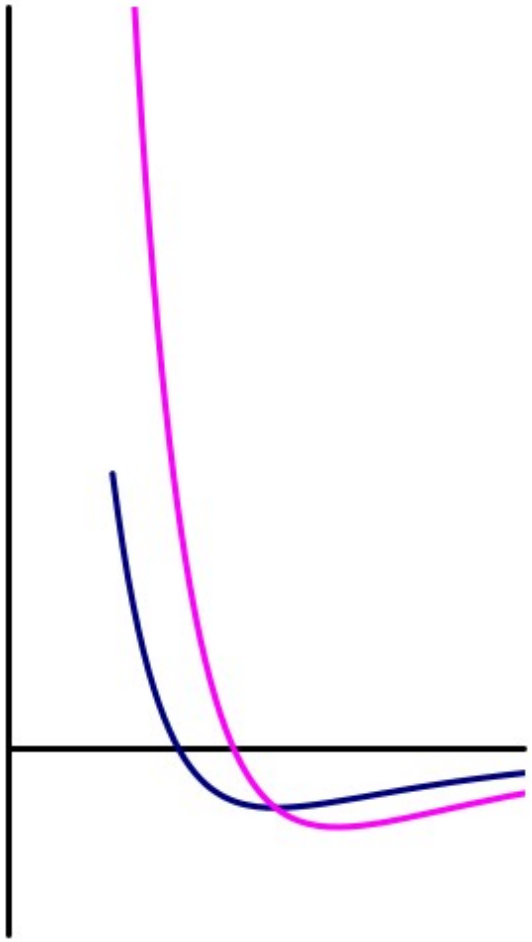


Зависимость потенциала ван-дер-Ваальсова взаимодействия от параметров

$$-\frac{A_{ij}}{r_{ij}^{6}}+\frac{B_{ij}}{r_{ij}^{12}}$$

A **= 0,5;** ***B*** **= 0,1**

A **= 1;** ***B*** **= 0,3**

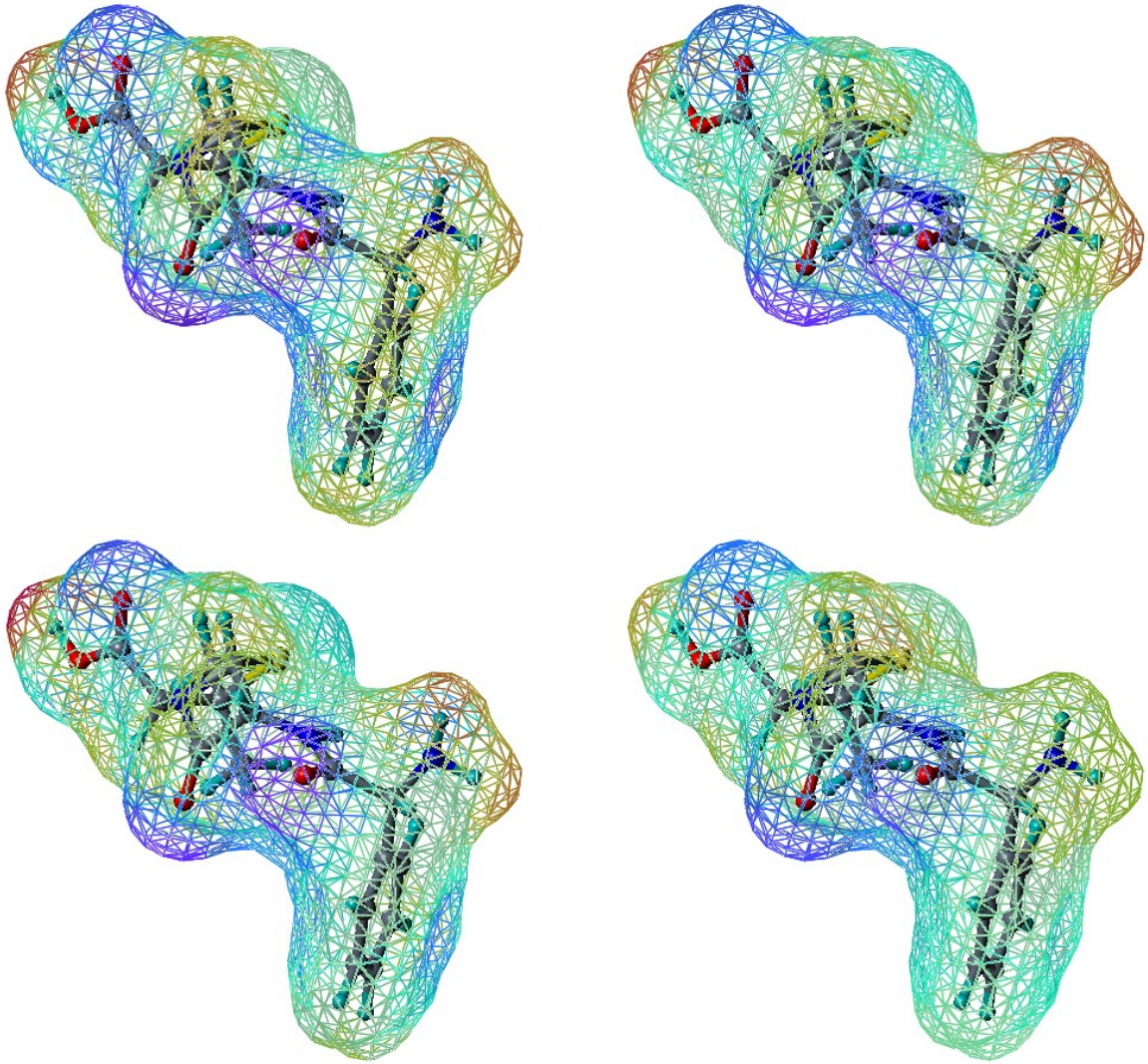


Леннард-Джонс

Электростатический потенциал

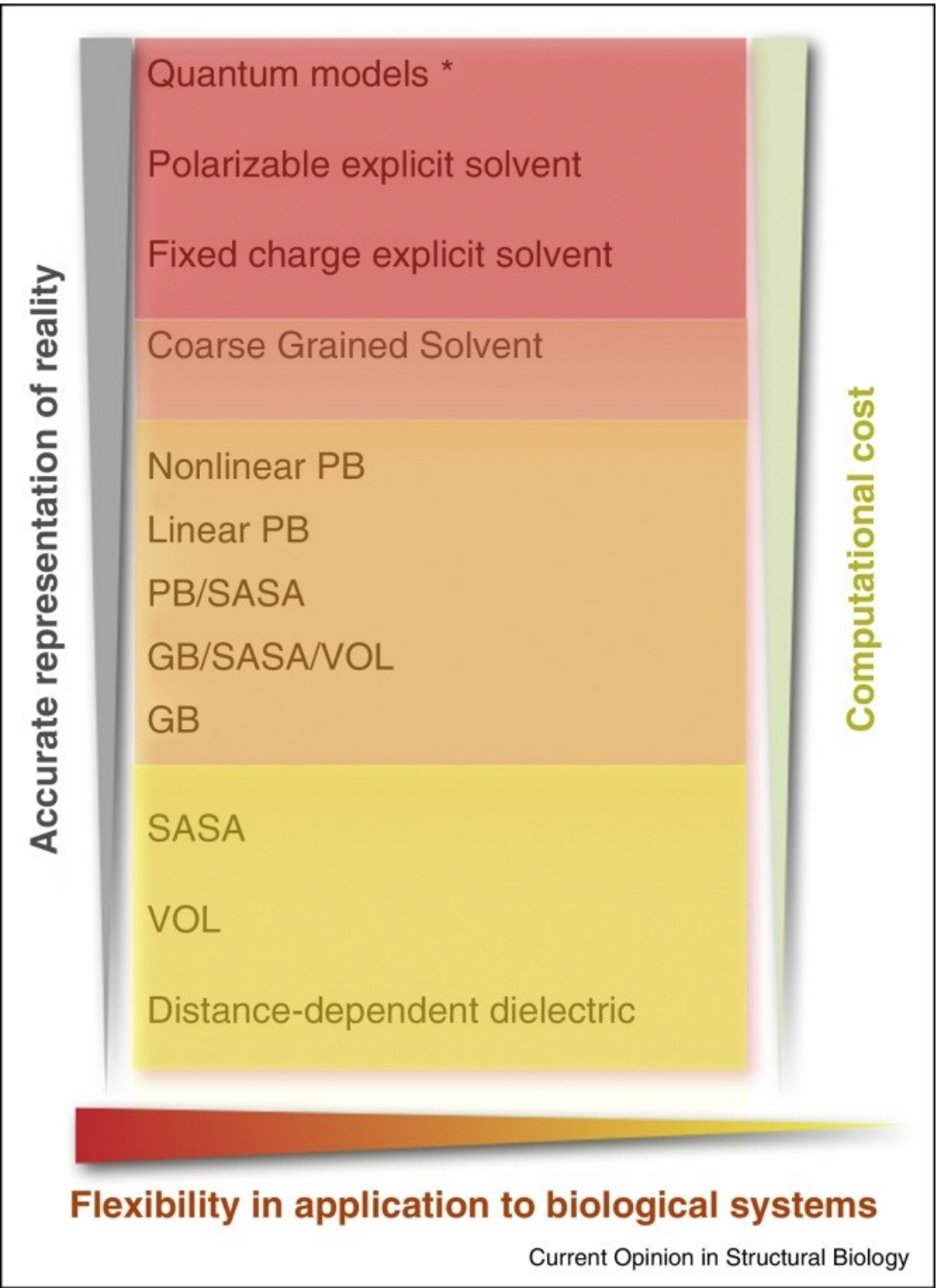
Закон Кулона:

$$V_{el} = \sum_{ij} \frac{q_i q_j}{\varepsilon_{ij} r_{ij}}$$



Кулон

Учёт влияния растворителя



- Зависящая от расстояния диэлектрическая проницаемость

$$\varepsilon(r)=r \qquad \varepsilon(r)=A+\frac{B}{1+ke^{-\lambda Br}}$$

- Численное решение нелинейного дифференциального уравнения Пуассона — Больцмана (**PB**)

$$\vec{\nabla}\left[\epsilon(\vec{r})\vec{\nabla}\psi(\vec{r})\right]=-4\pi\rho^{f}(\vec{r})-4\pi\sum_{i}c_i^{\infty}z_i q\, e^{\frac{-z_i q\,\psi(\vec{r})}{kT}}\lambda(\vec{r})$$

- *Приближённый подход:* обобщённая модель Борна (**GBSA**)

- *Специальные подходы:* COSMO-RS, 3D-RISM

- Явно заданный растворитель

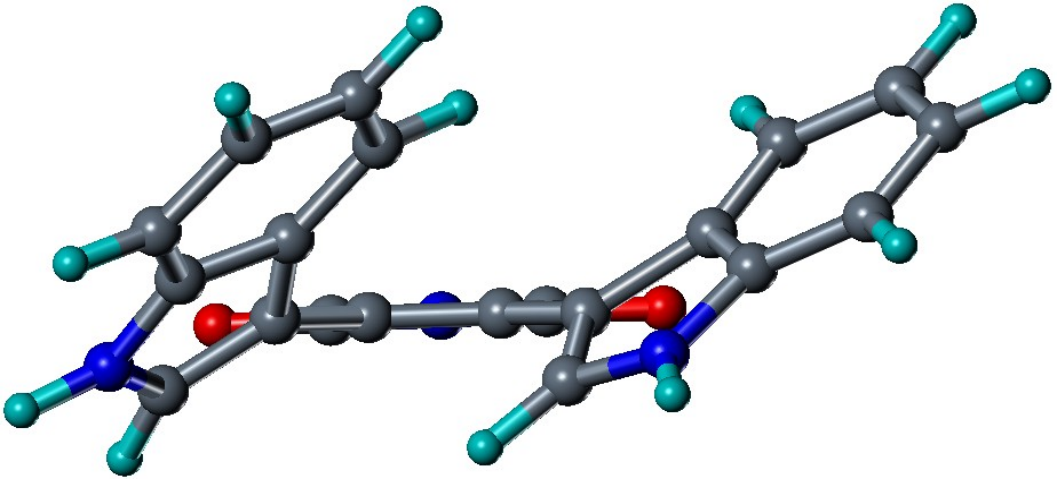
Kleinjung & Fraternalli, *Curr. Opin. Struct. Biol.*, **2014,** *25*, 126-134

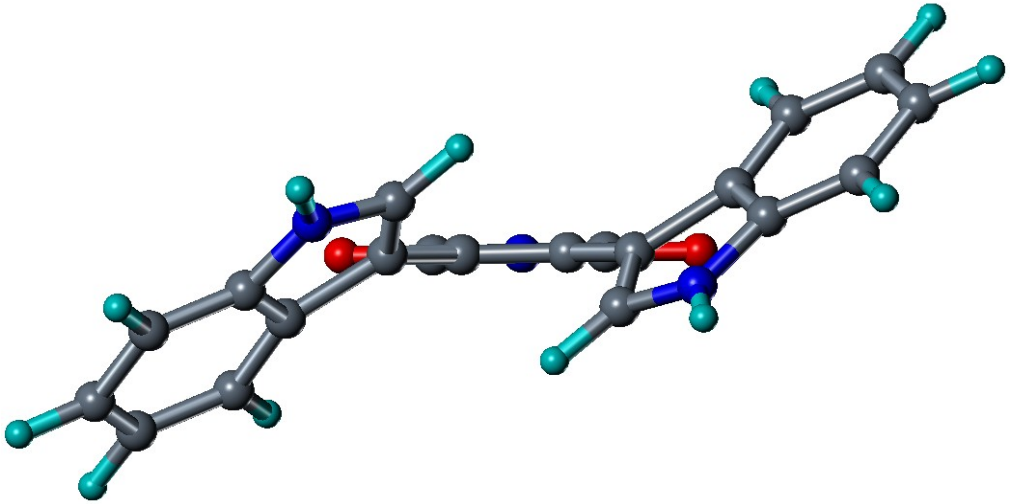
Типы силовых полей

- **Для малых молекул** (MM2/MM3/MM4, Tinker, Momec, Cosmos, GAFF, ...)
- **Для биологических макромолекул** (AMBER, GROMOS, CHARMM, OPLS, ECEPP, CVFF/CFF, ...)
- **Универсальные** (Tripos, MMFF, UFF)
- **«Крупнозернистые»** (MARTINI)

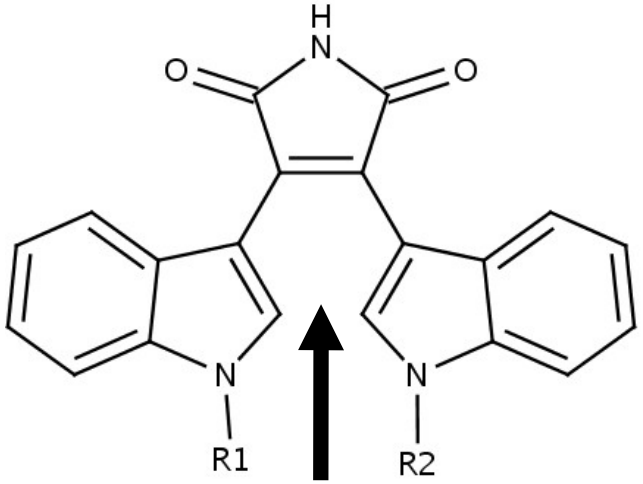
1. Введение
2. Теоретические основы
3. Моделирование малых молекул
4. Моделирование больших молекул
5. Моделирование комплексов
6. Дизайн комплексов

Конформационный анализ

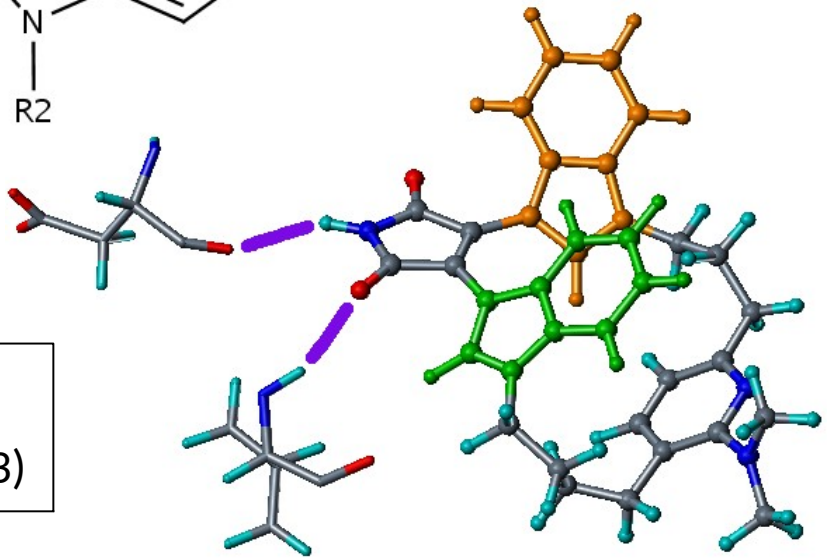




Свёрнутая конформация
E = 34,8 *ккал/моль*



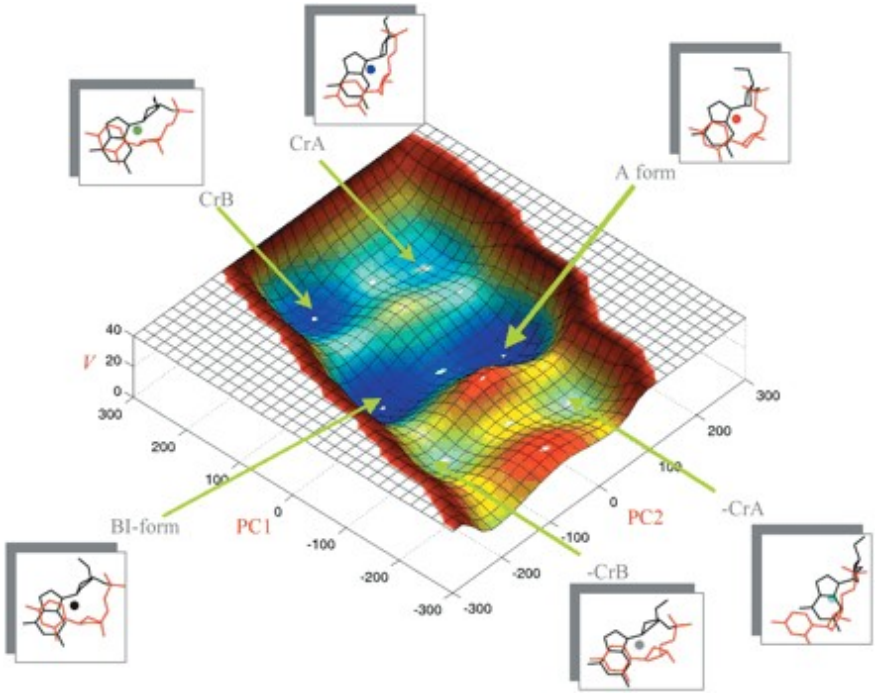
Развёрнутая конформация
E = 36,9 *ккал/моль*

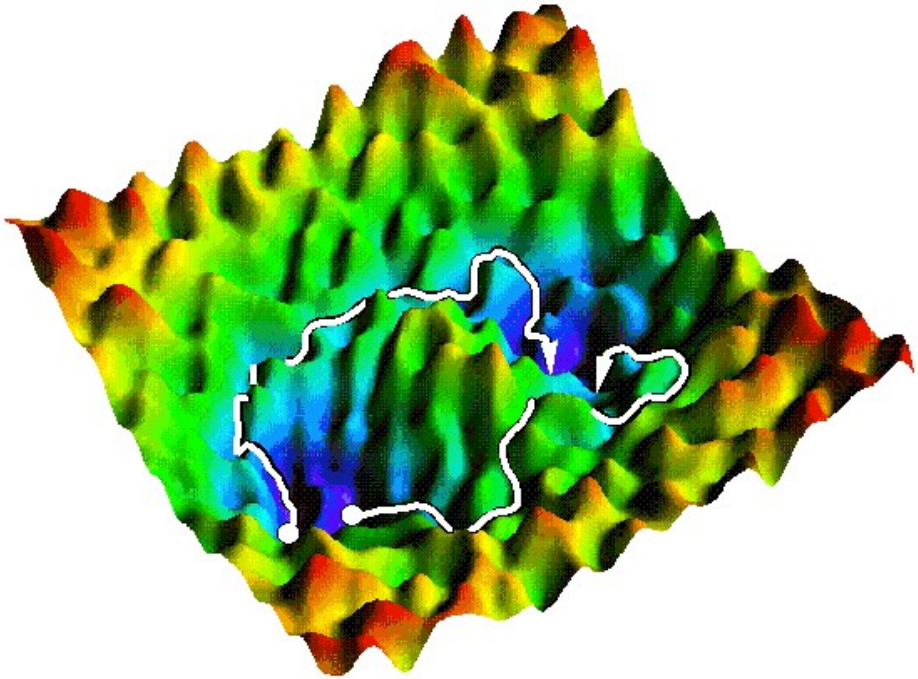


Подтверждено кристаллической структурой (2OW3)

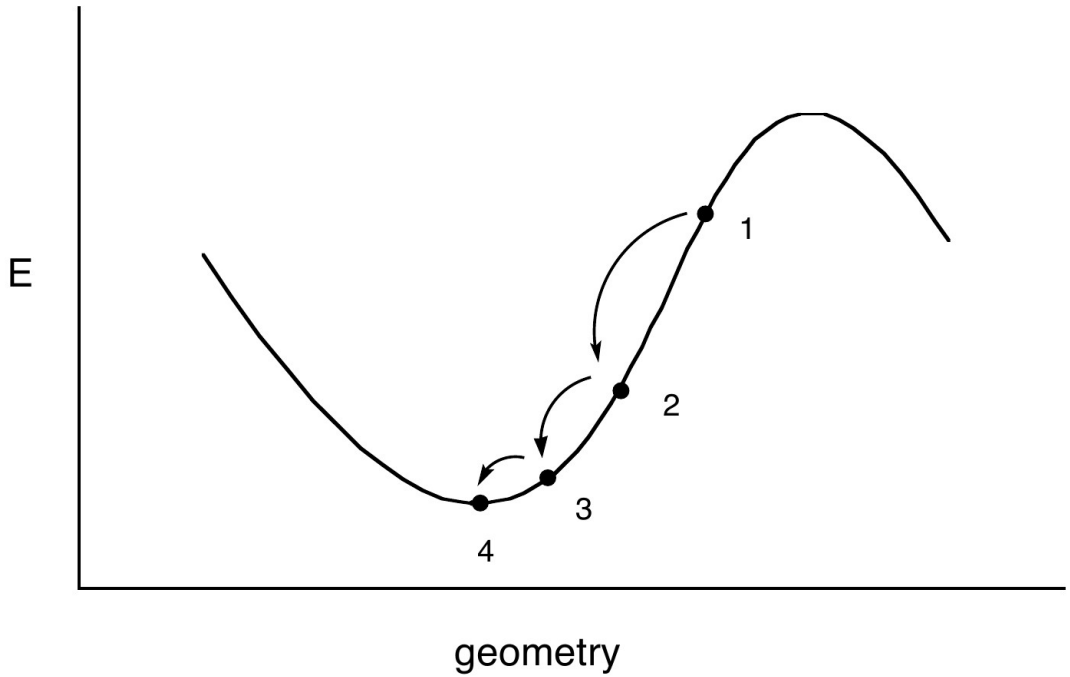
Поверхность потенциальной энергии (ППЭ)

Локальные минимумы – конформации,
седловые точки – переходные состояния





Минимизация энергии (Оптимизация геометрии)



Идеология:

$x_{new} = x_{old} + \text{correction}$

Метод Ньютона — Рафсона:

$$x_{new} = x_{old} - \frac{E'(x_{old})}{E''(x_{old})}$$

очень медленно сходится

Метод наискорейшего спуска:

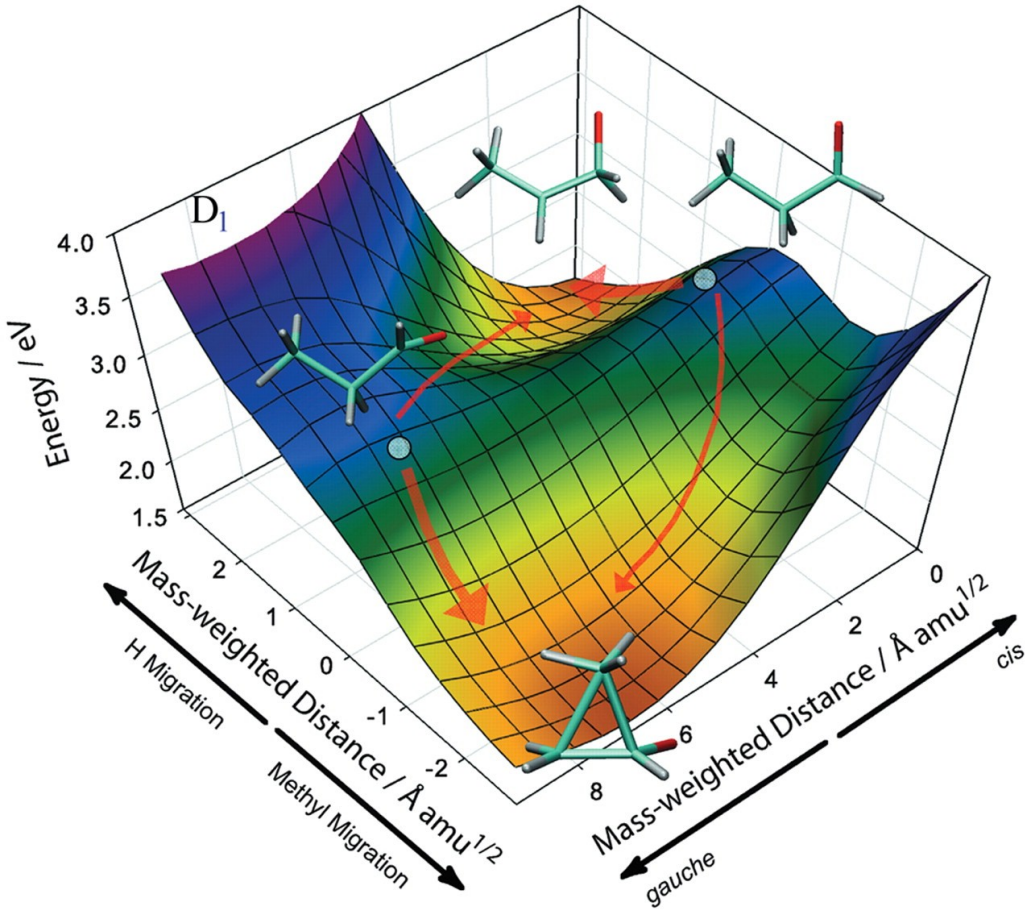
$$x_{new} = x_{old} - \gamma E'(x_{old})$$

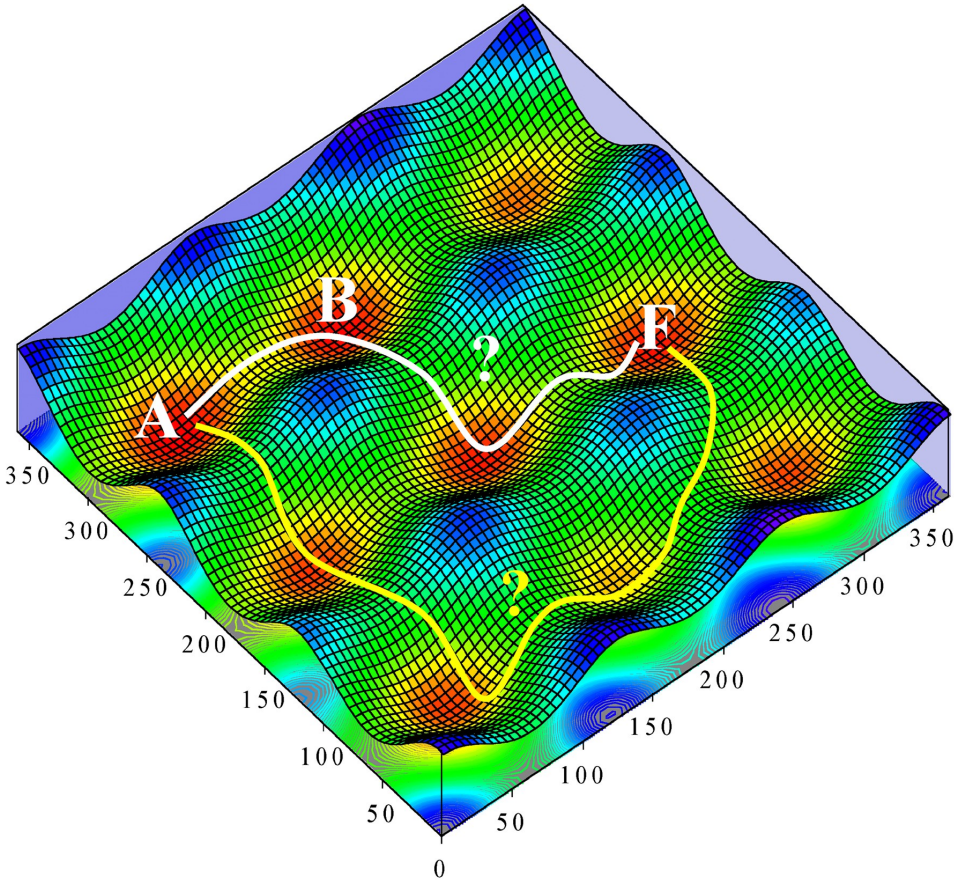
медленно сходится
в окрестности
минимума

Метод сопряжённых градиентов:

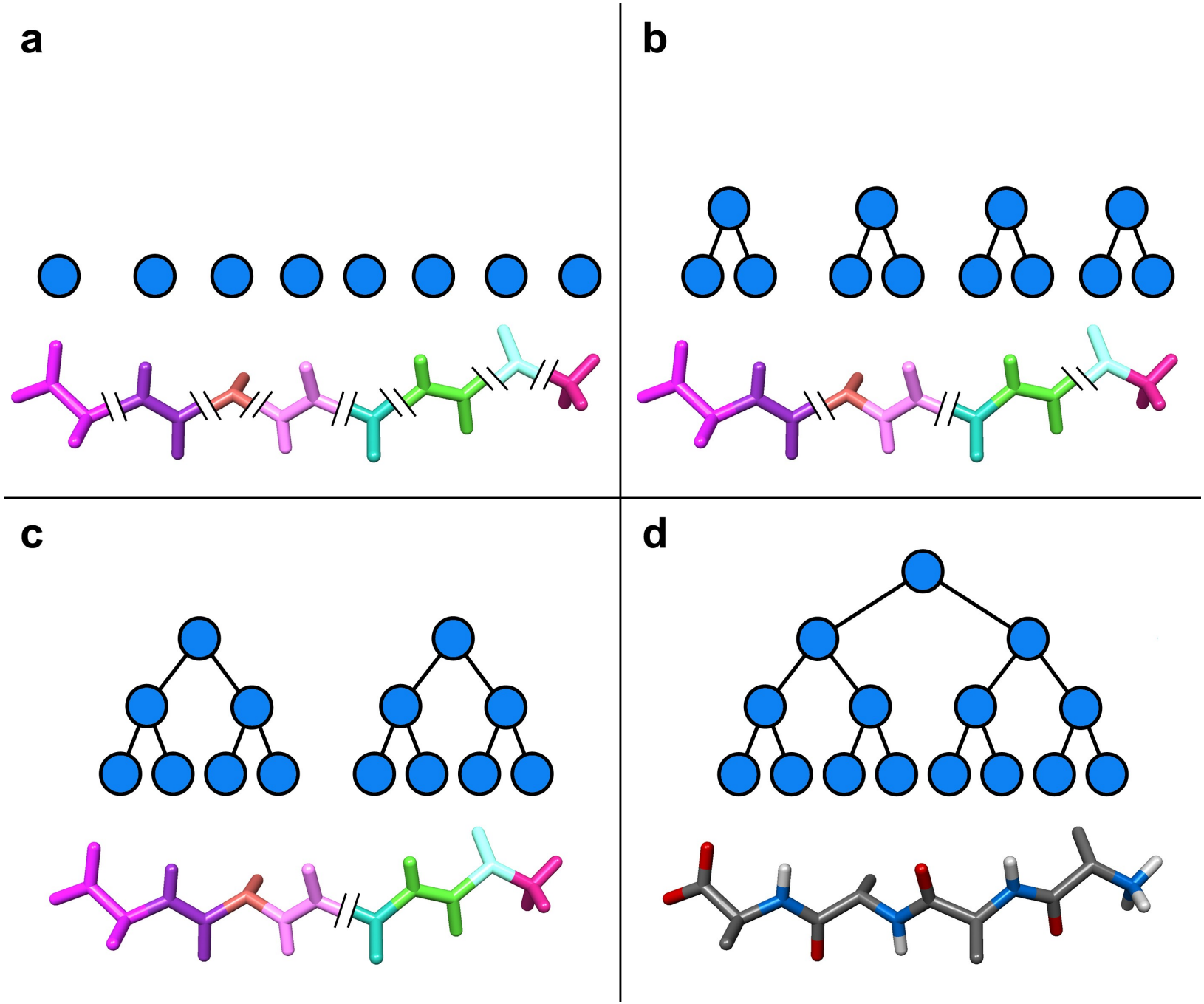
быстрее сходится
в окрестности минимума
и применяется
в паре с методом
наискорейшего спуска

Поиск полного набора локальных минимумов на ППЭ
Поиск по решётке





Поиск полного набора локальных минимумов на ППЭ
Поиск по дереву



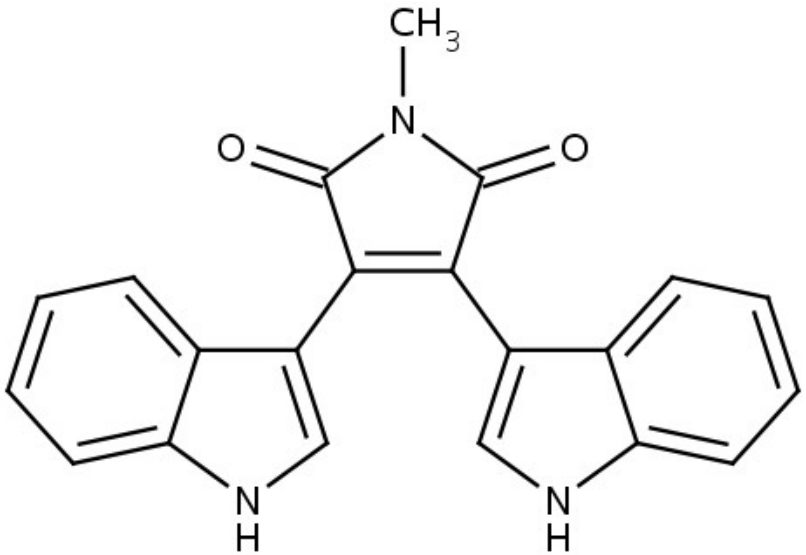
Поиск полного набора локальных минимумов на ППЭ
Поиск методом Монте-Карло

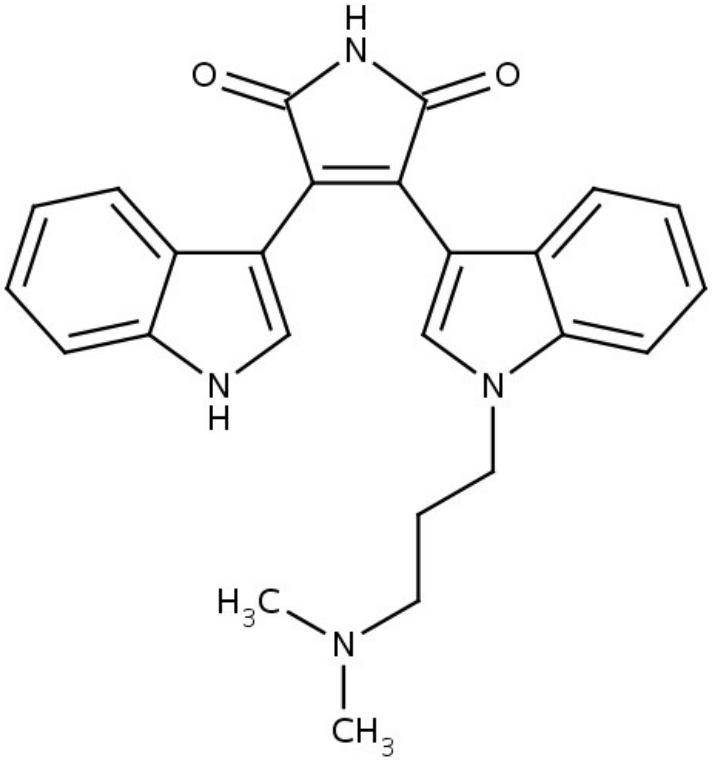
Генерация структур
со случайными
параметрами
и оптимизация;

позволяет добиться
максимального
разнообразия
конформеров

Перечисление конформаций

Проблема №1: комбинационный взрыв





BIS-V

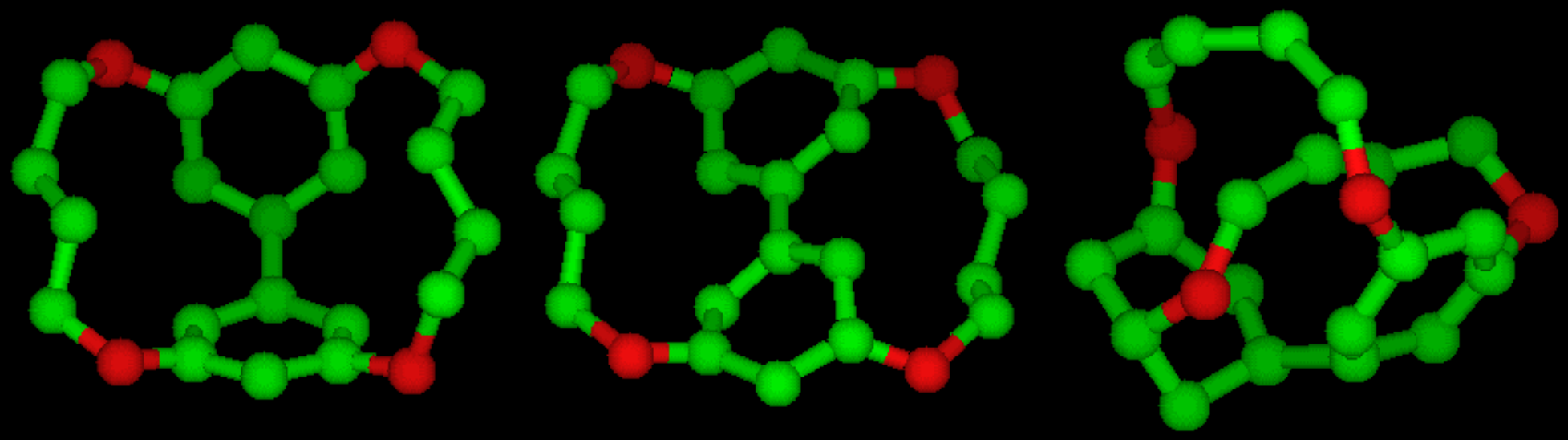
- 2 степени свободы
- 14 конформаций

BIS-I

- 6 степеней свободы
- 2123 конформации

Перечисление конформаций (2)

Проблема №2: качество автоматической генерации



Требуется вдумчивый анализ результатов

1. Введение
2. Теоретические основы
3. Моделирование малых молекул
4. Моделирование больших молекул
5. Моделирование комплексов
6. Дизайн комплексов

Получение информации о структуре биомишени

- Эксперимент (РСА, ЯМР, крио-ЭМ) (Но: в PDB сегодня всего лишь 117651 структура — http://pdb.org)
- Моделирование:
  - по гомологии (Modeller)
  - *de novo* (Rosetta)

Предпосылка метода моделирования по гомологии:

Белки с похожими последовательностями будут иметь похожую структуру

Структурные данные

Экспериментальные методы:

- **Рентгеноструктурный анализ** (конформация зафиксирована, структуры зачастую фрагментарны, необходимы кристаллы)
- **ЯМР** (набор конформаций, необходим раствор)
- **Криоэлектронная микроскопия** (низкое разрешение)
- **Малоугловое рентгеновское рассеяние** (низкое разрешение, сложность интерпретации)

- **Качество структур зачастую неприемлемо для изучения межмолекулярных взаимодействий**
- **Встречаются случаи неправильной интерпретации данных авторами структуры**

Моделирование структуры белка

Процедура построения модели

1. Выравнивание аминокислотных последовательностей

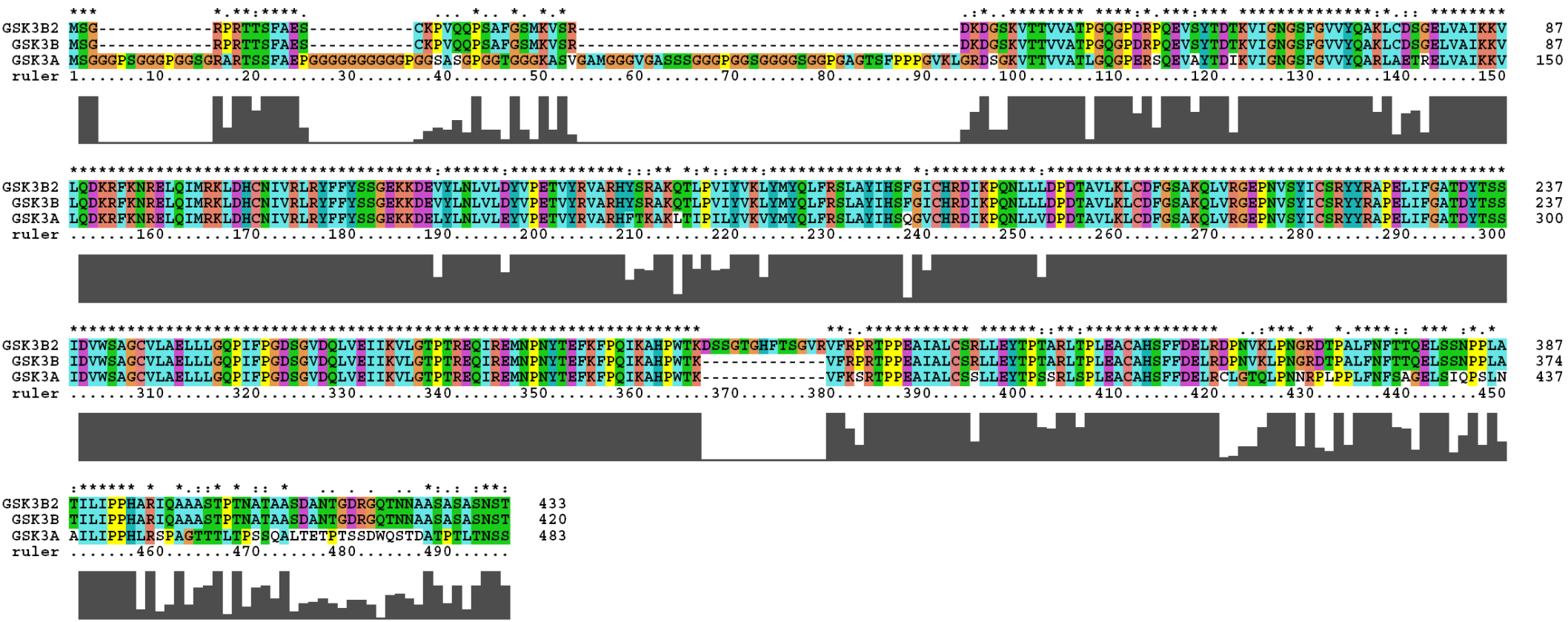
2. Поиск шаблонной структуры (PDB)

3. Моделирование по гомологии

4. Оценка качества построенной модели структуры белка

5. Минимизация энергии модели, добавление атомов водорода и зарядов

Выравнивание аминокислотных последовательностей

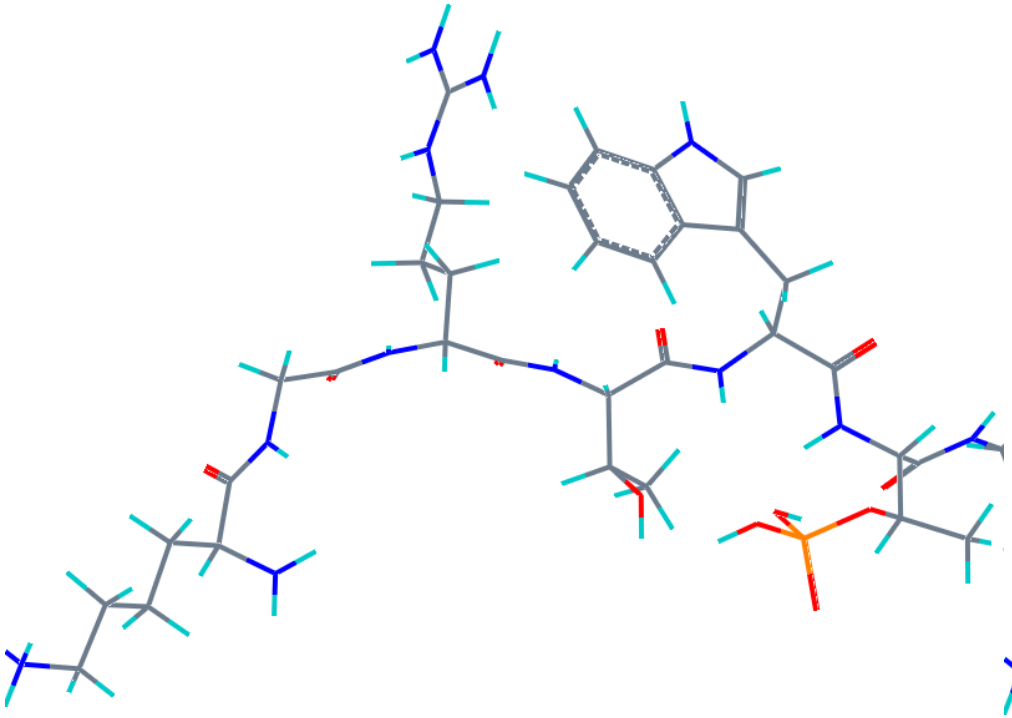


Необходимое условие возможности моделирования:

идентичность последовательностей не меньше 20%.

Моделирование по гомологии

Метод точечных мутаций

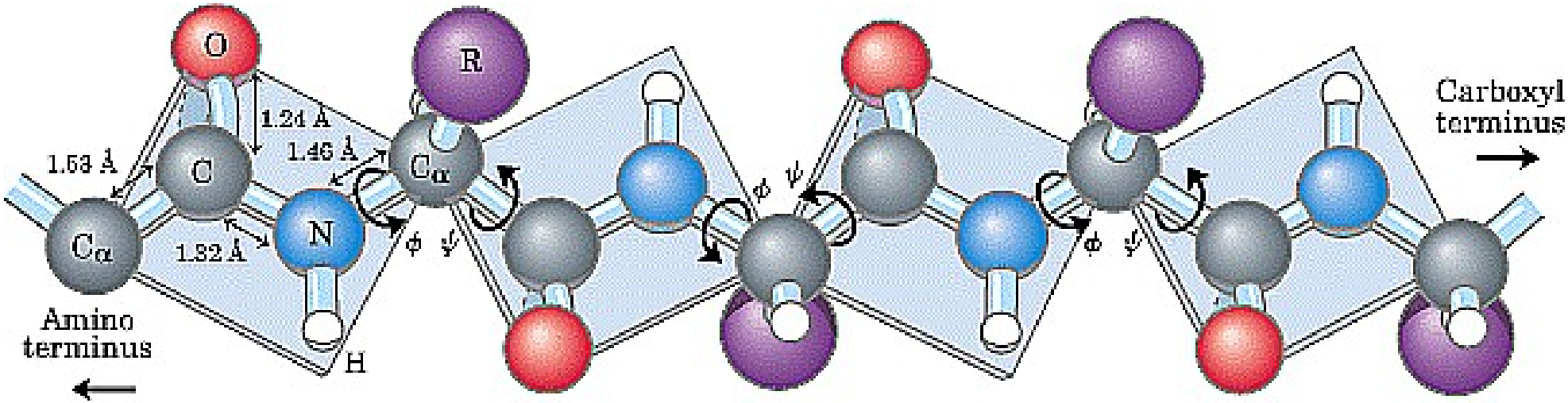


KGRTWTL **KKMTRTL**

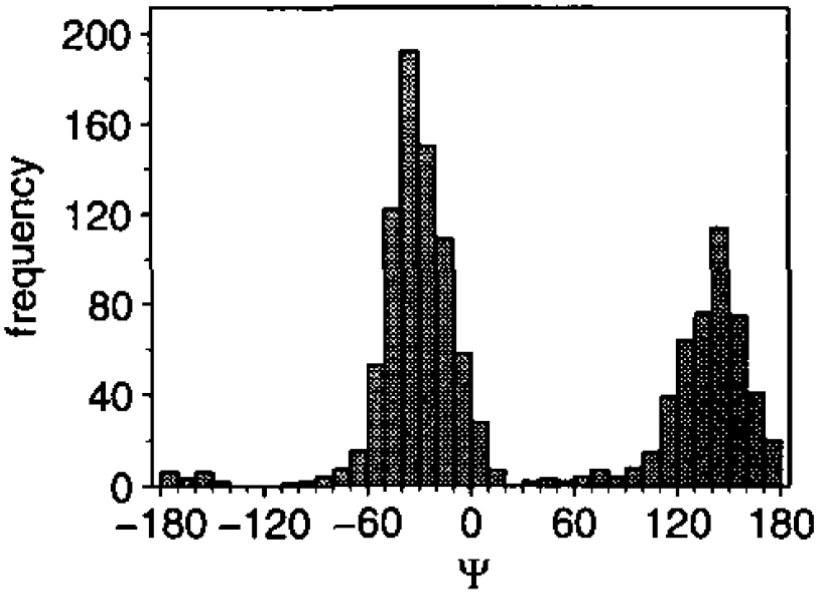
- Необходима вдумчивая оптимизация
- Проблемы с обработкой инсерций и делеций

Моделирование по гомологии

Метод Шали — Бланделла (Modeller)



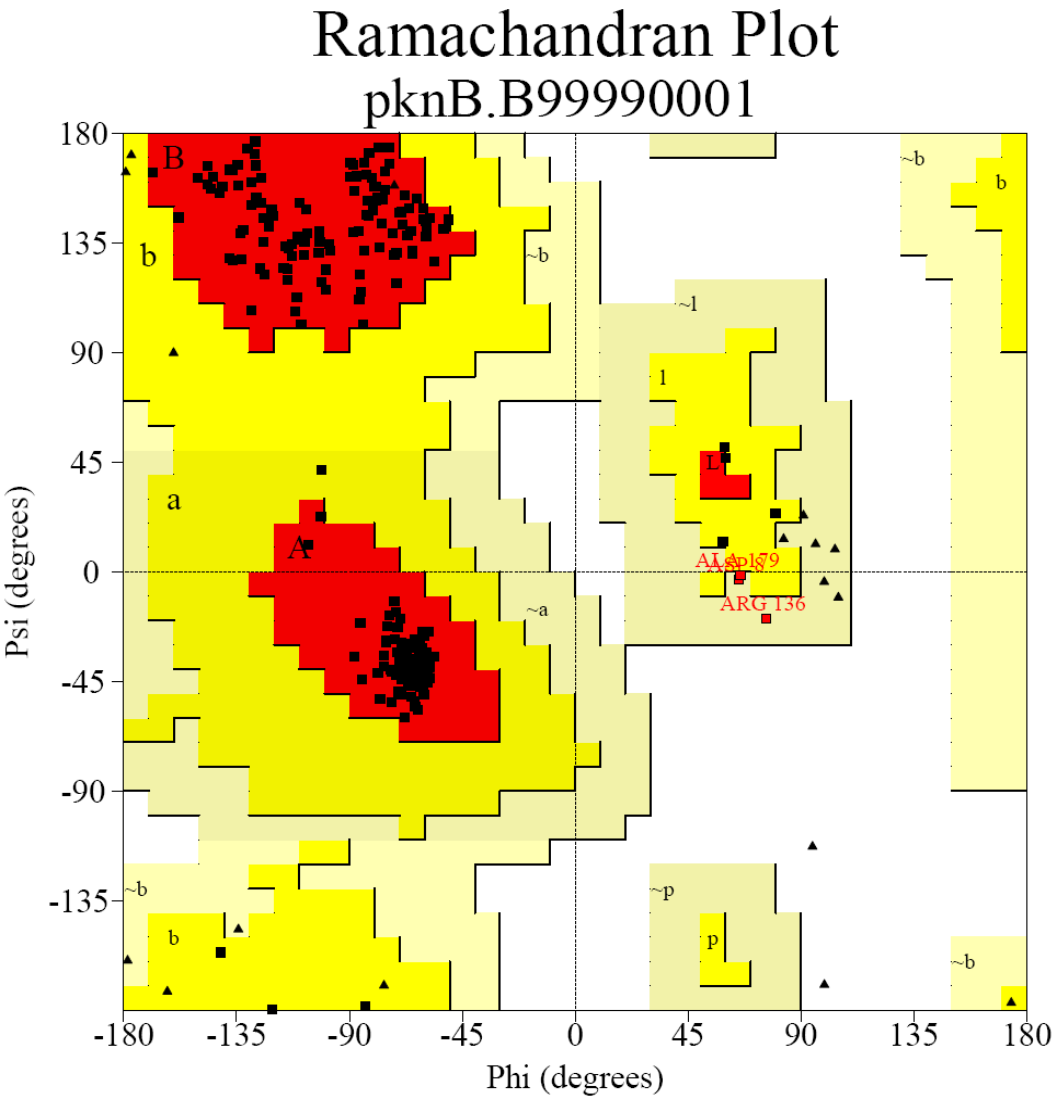
- Длины связей
- Углы между связями
- Торсионные углы
- Другие ограничения (ЯМР, водородные связи)



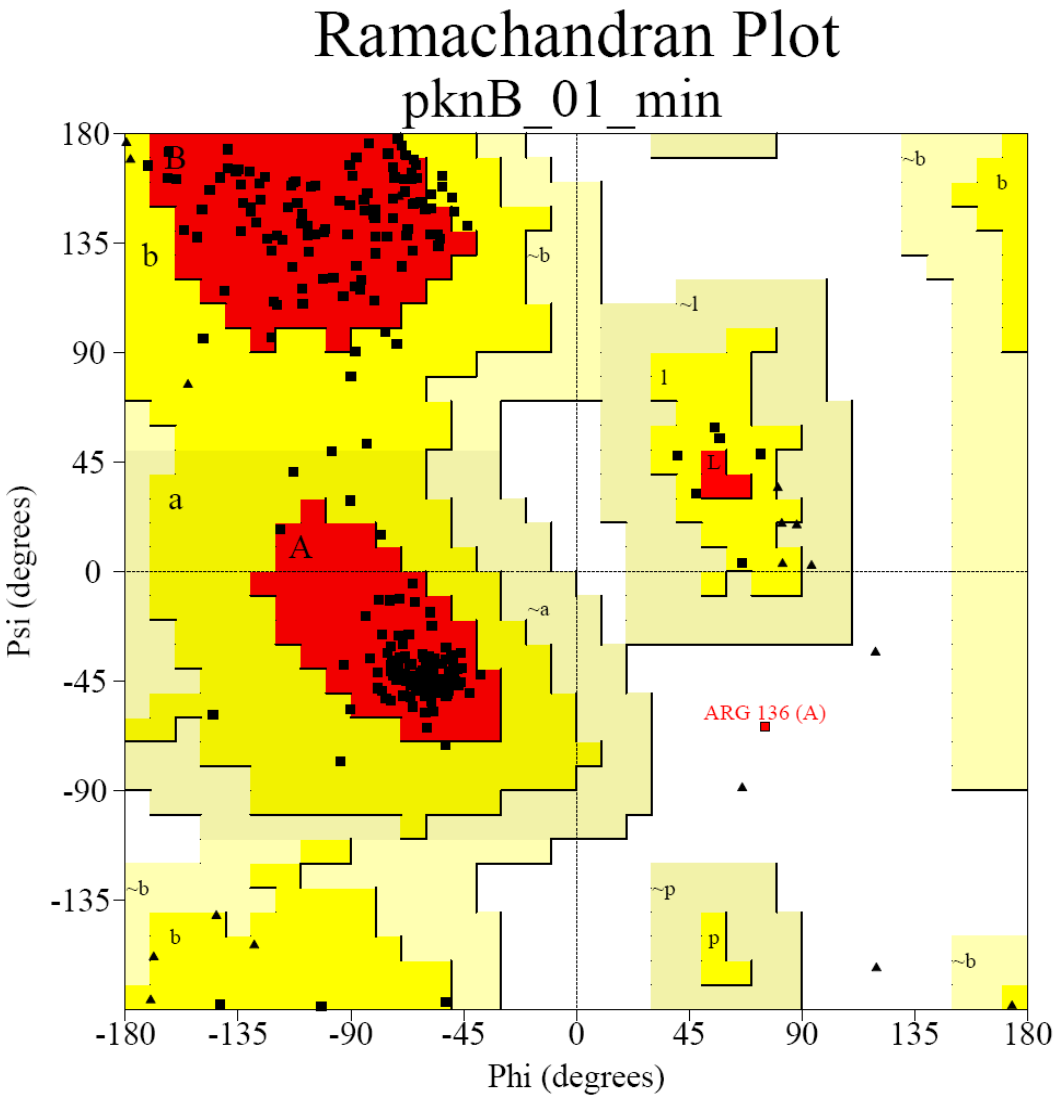
Šali & Blundell, *J. Mol. Biol.*, **1993,** *234*, 779-815

Figure 6. Distribution of the Ψ main-chain dihedral angle at $\Phi = [-80°, -70°]$. The table $W'(\Phi, \Psi)$ was obtained from the local database of protein structures.

Контроль качества: Карта Рамачандрана



До минимизации



После минимизации

Laskowski *et al.*, *J. Appl. Cryst.*, **1993,** *26*, 283-291

Контроль качества: ProSA

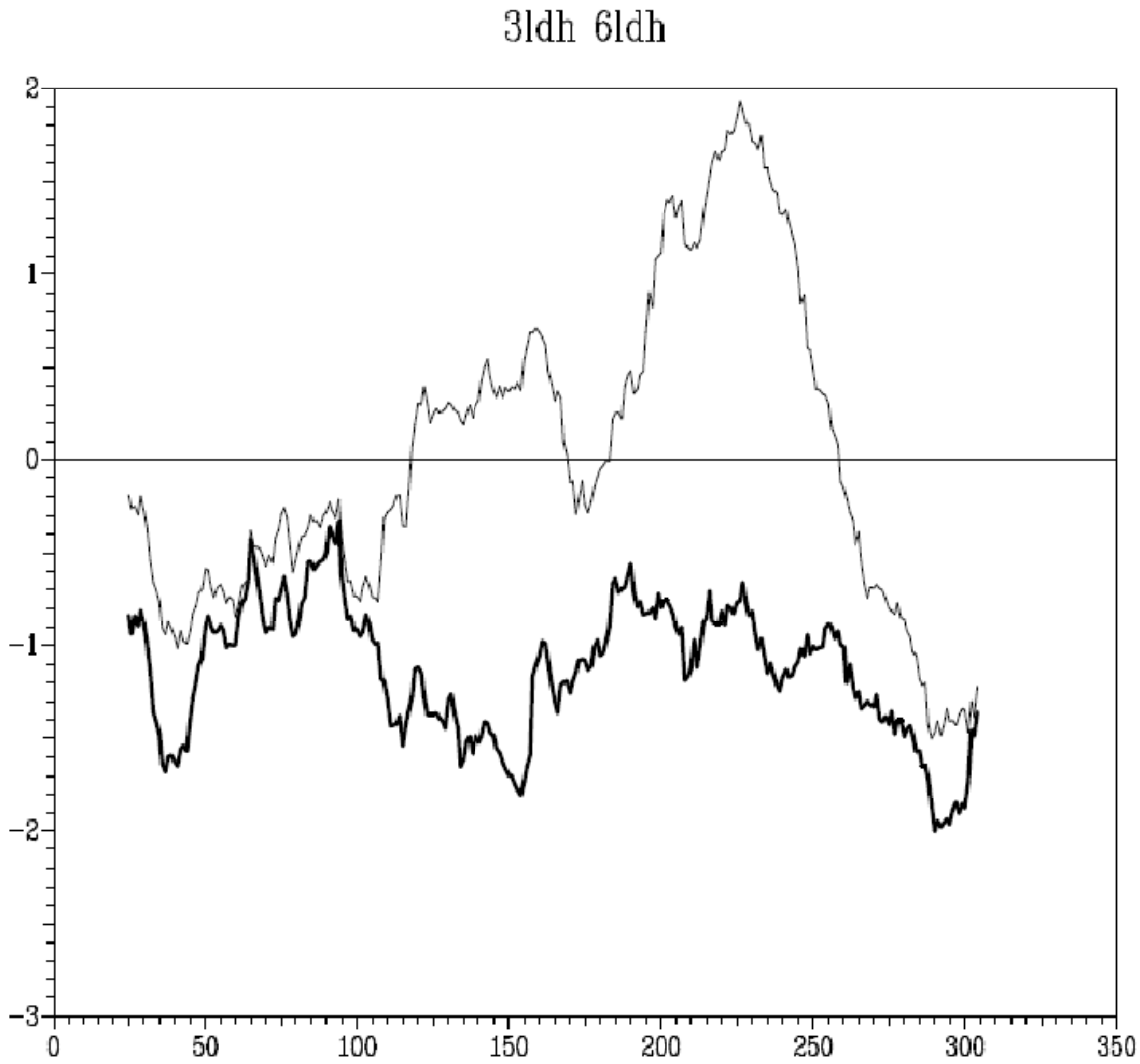


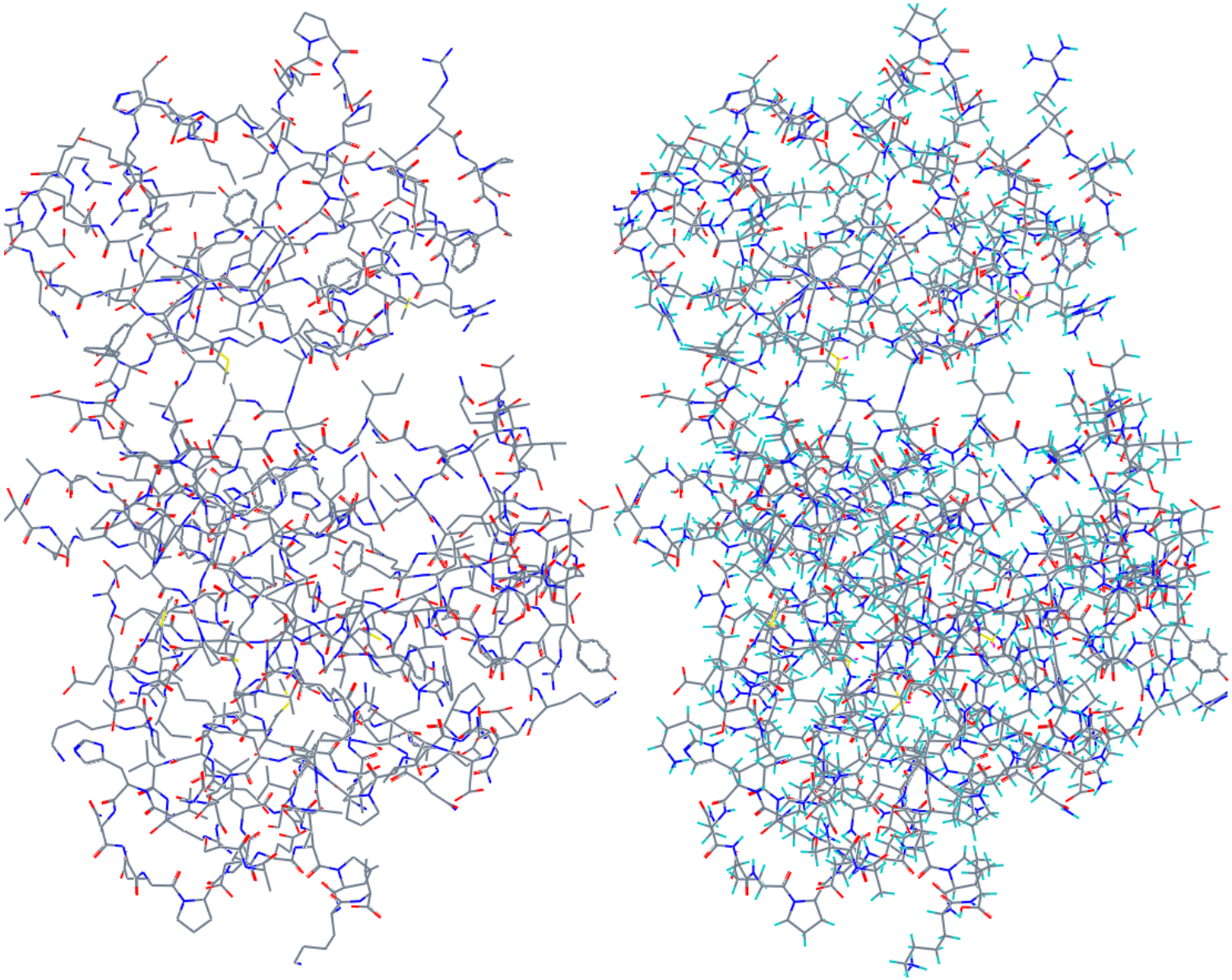
Figure 6: The result of the command file of sample session 3. Combined energies are shown. Window size is 50. Thin line: 3ldh; bold line: 6ldh.

Оценка качества локальной укладки белка: профиль энергии вдоль последовательности.

Общая оценка — *z*-score.

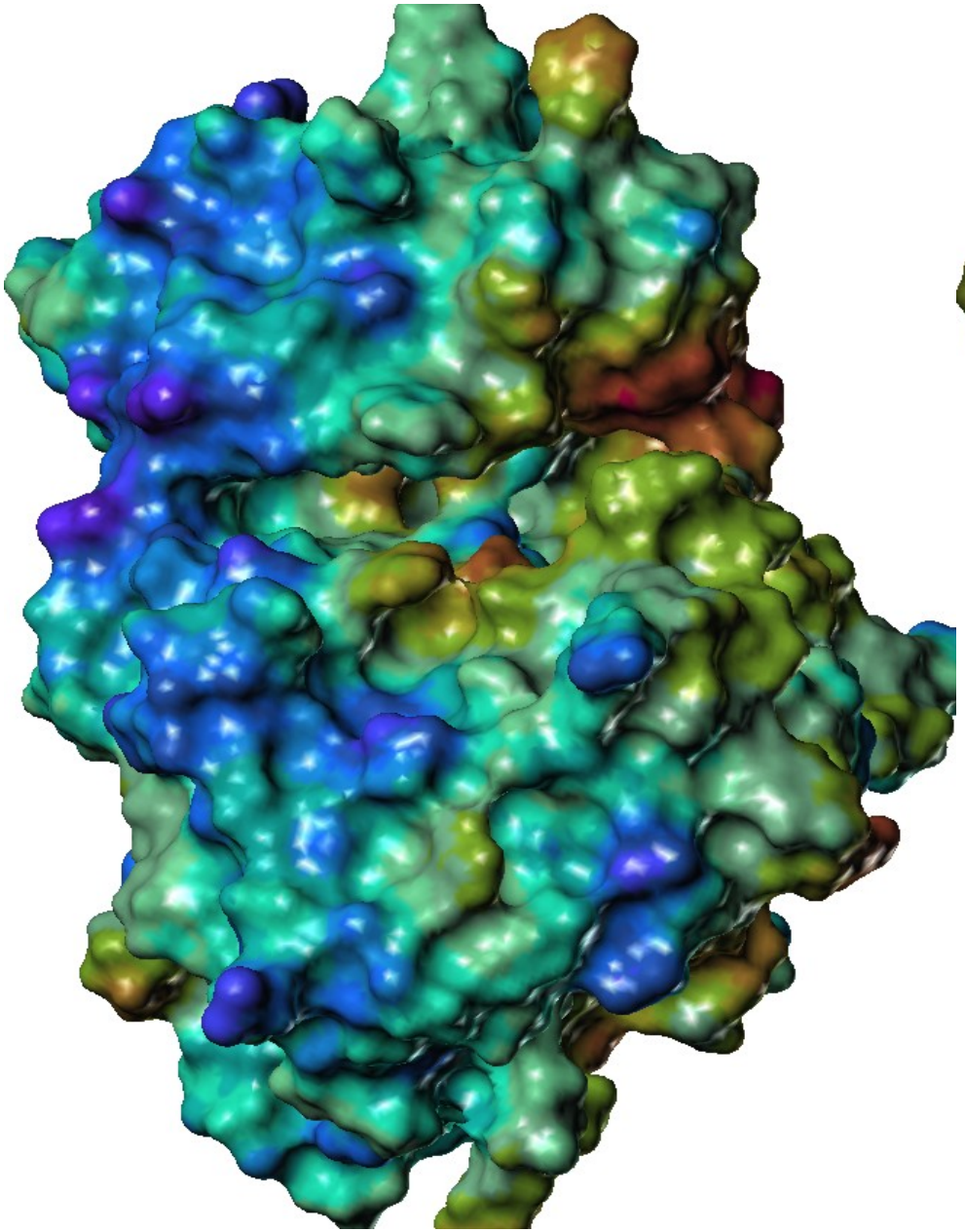
Sippl, *Proteins*, **1993,** *17*, 355-362

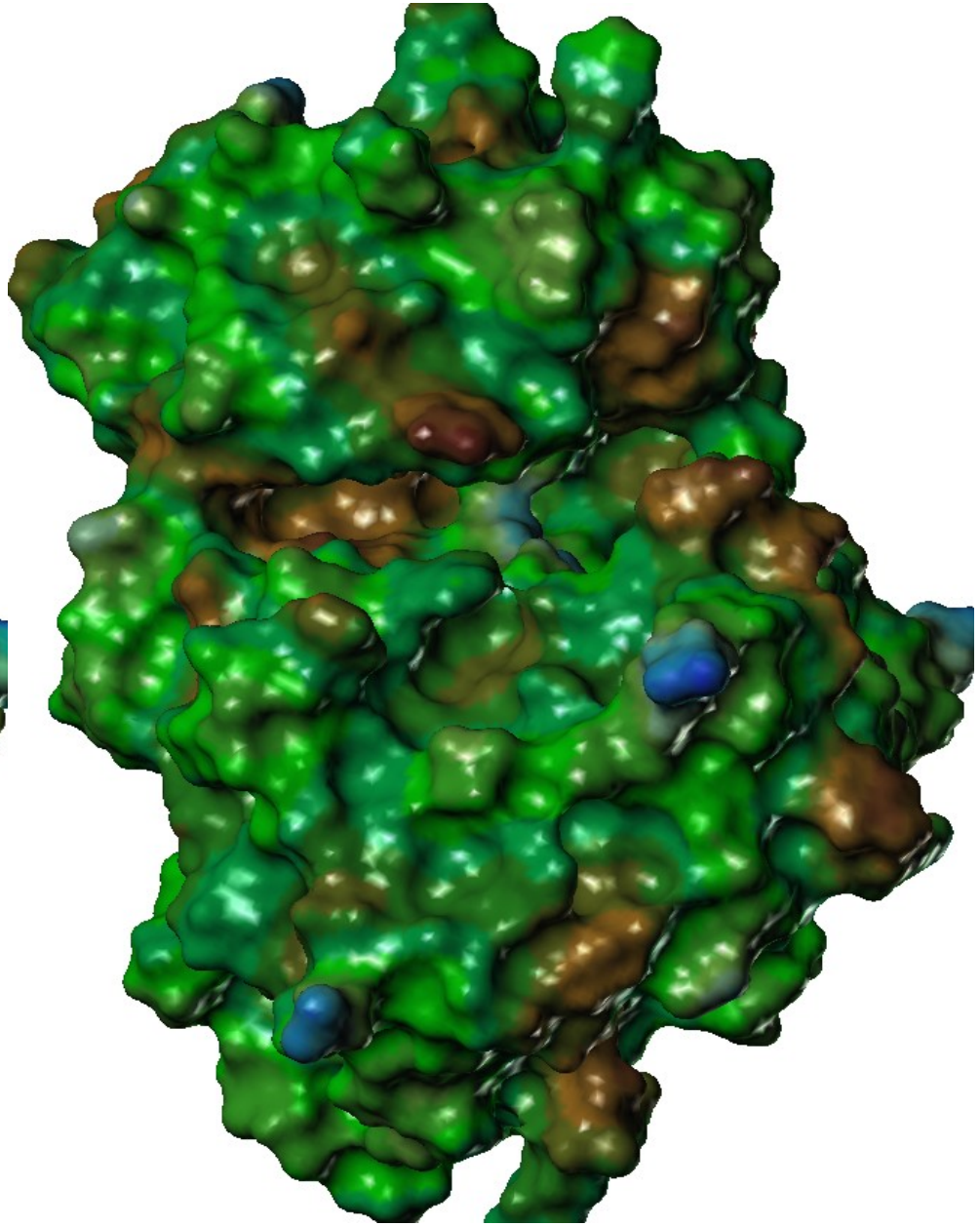
Атомы водорода и др.



Задача моделирования №1: приведение рентгеновской структуры в удобоваримый вид

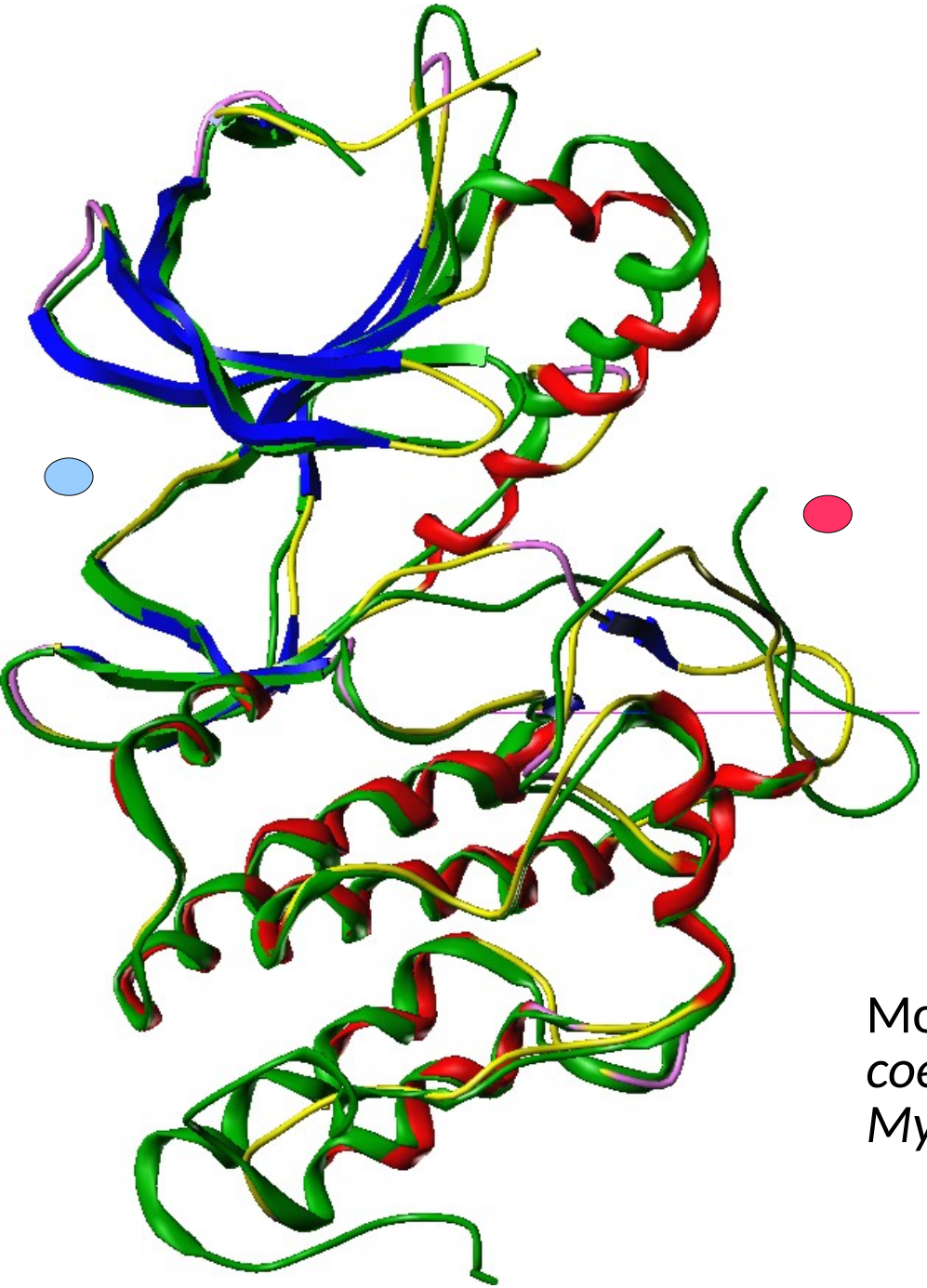
Электростатика **Липофильность**



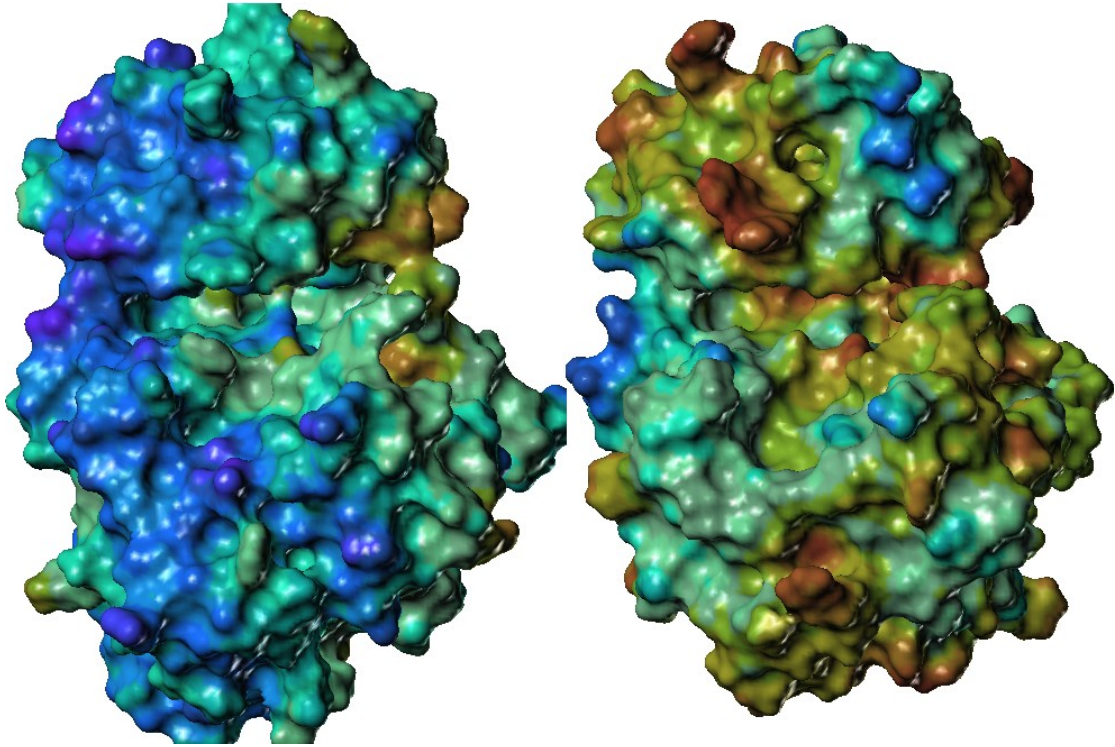


Киназа PknB *Mycobacterium tuberculosis*

Модель и шаблон



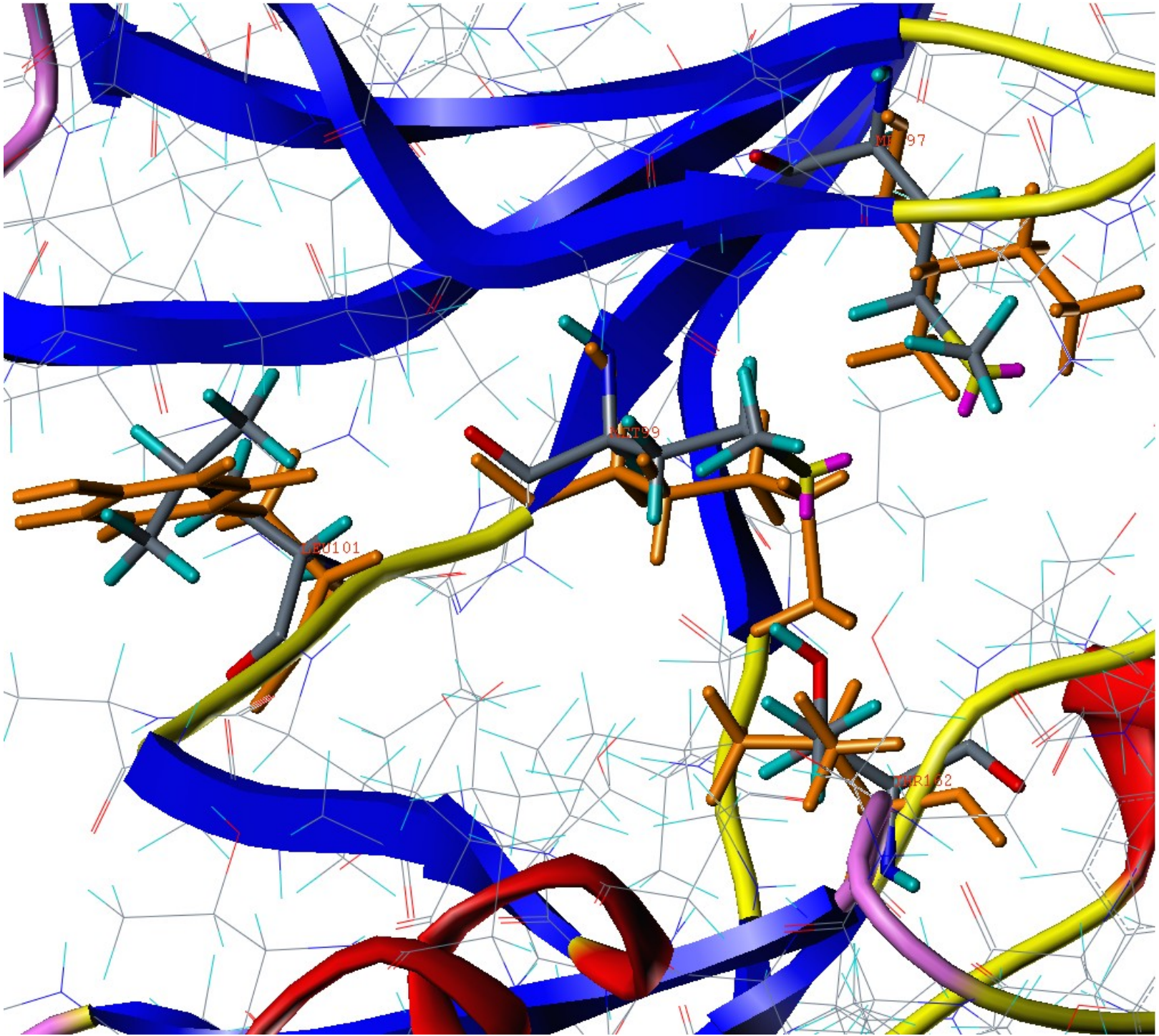
Задача моделирования №2:
сравнение структур



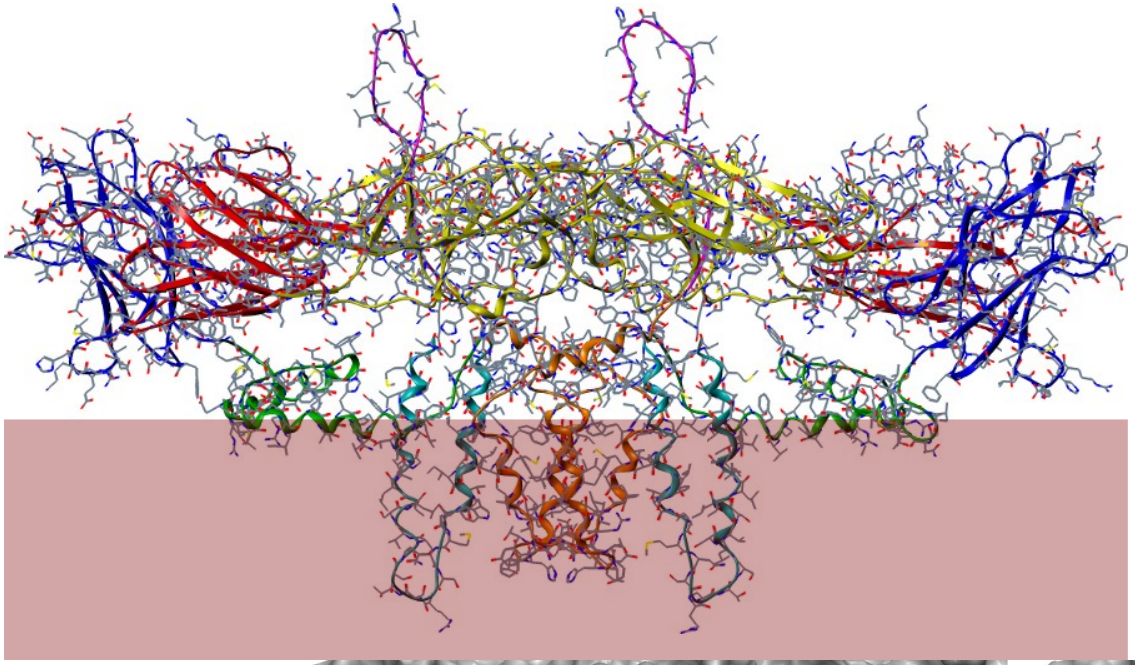
Модель структуры киназы Pk25 *Streptomyces coelicolor* (справа) по структуре киназы PknB *Mycobacterium tuberculosis*

Bekker *et al.*, *Acta Naturae*, **2010,** *2 (3)*, 110-121

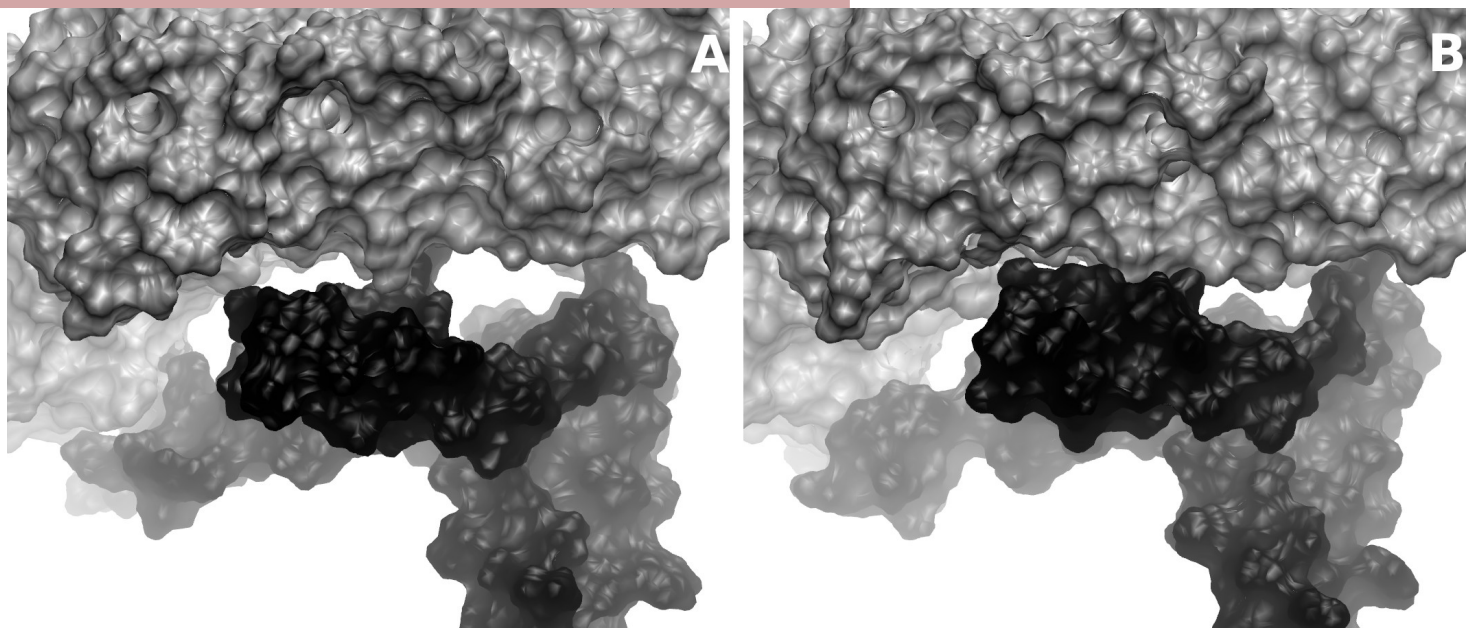
Модель и шаблон (2)



Белки оболочки флавивирусов: моделирование Монте-Карло



Model of DENV envelope proteins was derived from low-resolution cryo-EM structure (1P58) lacking M ectodomain, which should penetrate the holes to contact with *pr*.



Задача моделирования №3: исследование структур, не доступных экспериментально

Osolodkin *et al.*, *Biochem. Biophys. Res. Commun.*, **2012,** *425*, 207-211

Молекулярная динамика

Уравнение движения Ньютона

$$-\frac{dE}{dr_i}=m_i\frac{d^2r_i}{dt^2}$$

Ускорение атома

$$a_i=\frac{dv_i}{dt}$$

Скорость атома

$$v_i=\frac{dr_i}{dt}$$

Численное интегрирование уравнений Ньютона (“прыжок лягушки”)

```
10 SOLVE FOR a_i
20 SOLVE FOR VELOCITY @ (t + DELTAt/2)
30 SOLVE FOR POSITIONS @ (t + DELTAt)
40 GOTO 10
```

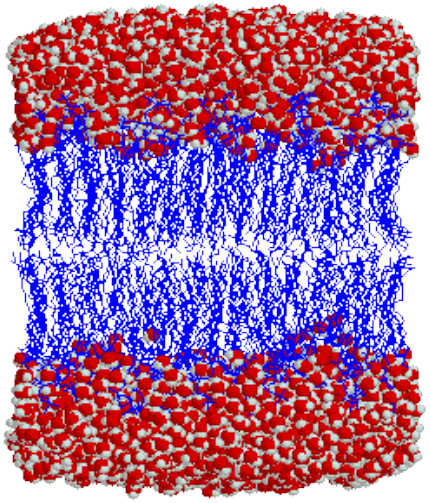
$$-\frac{dE}{dr_i}=F_i=m_i a_i(t)$$

$$v_{ix}\left(t+\frac{\Delta t}{2}\right)=v_{ix}\left(t-\frac{\Delta t}{2}\right)+\Delta t a_{ix}(t)$$

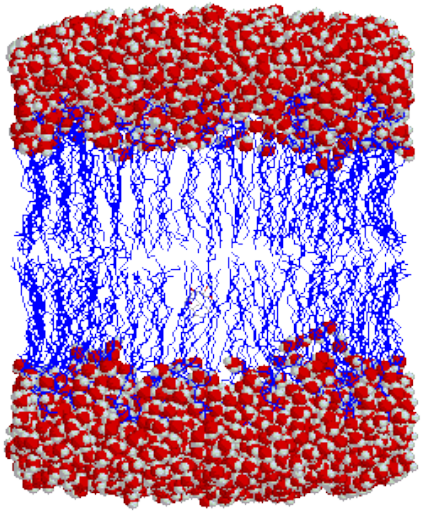
$$r_{ix}(t+\Delta t)=r_{ix}(t)+\Delta t v_{ix}\left(t+\frac{\Delta t}{2}\right)$$

$\mathbf{r}_i(t)$ – молекулярно-динамическая траектория атома *i*

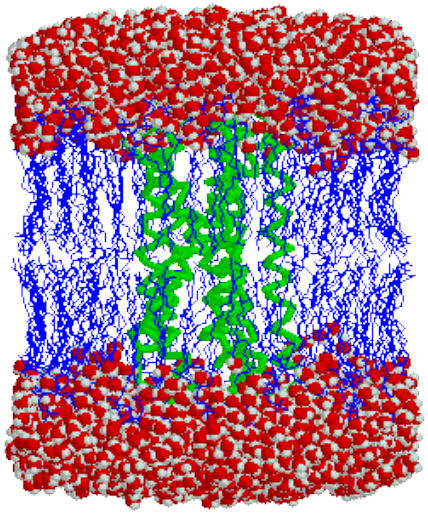
Внедрение модели белка в среду (воду или мембрану)



Мембрана

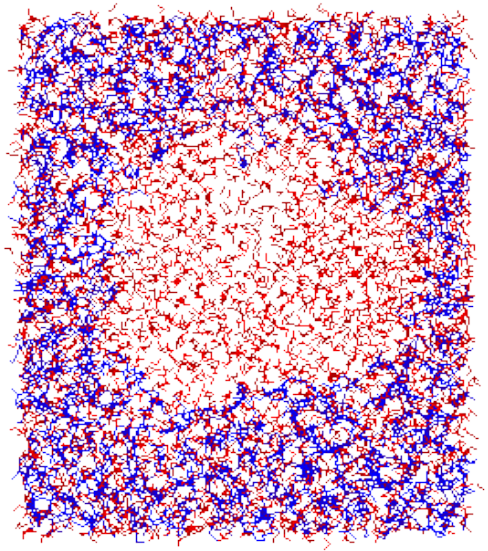


Дыра в мембране

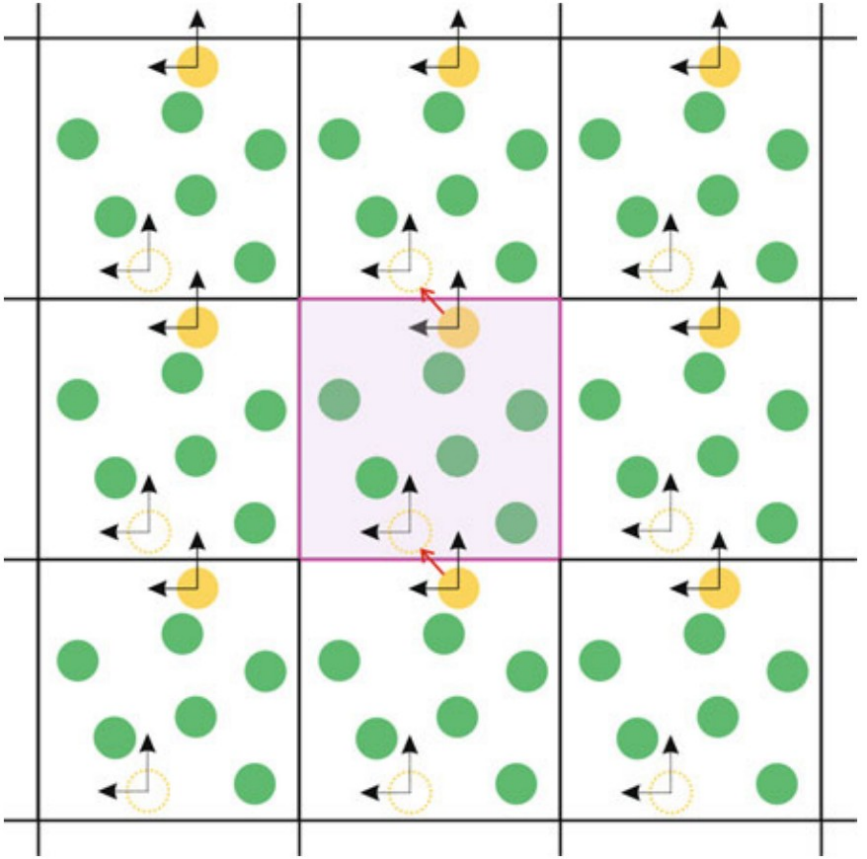


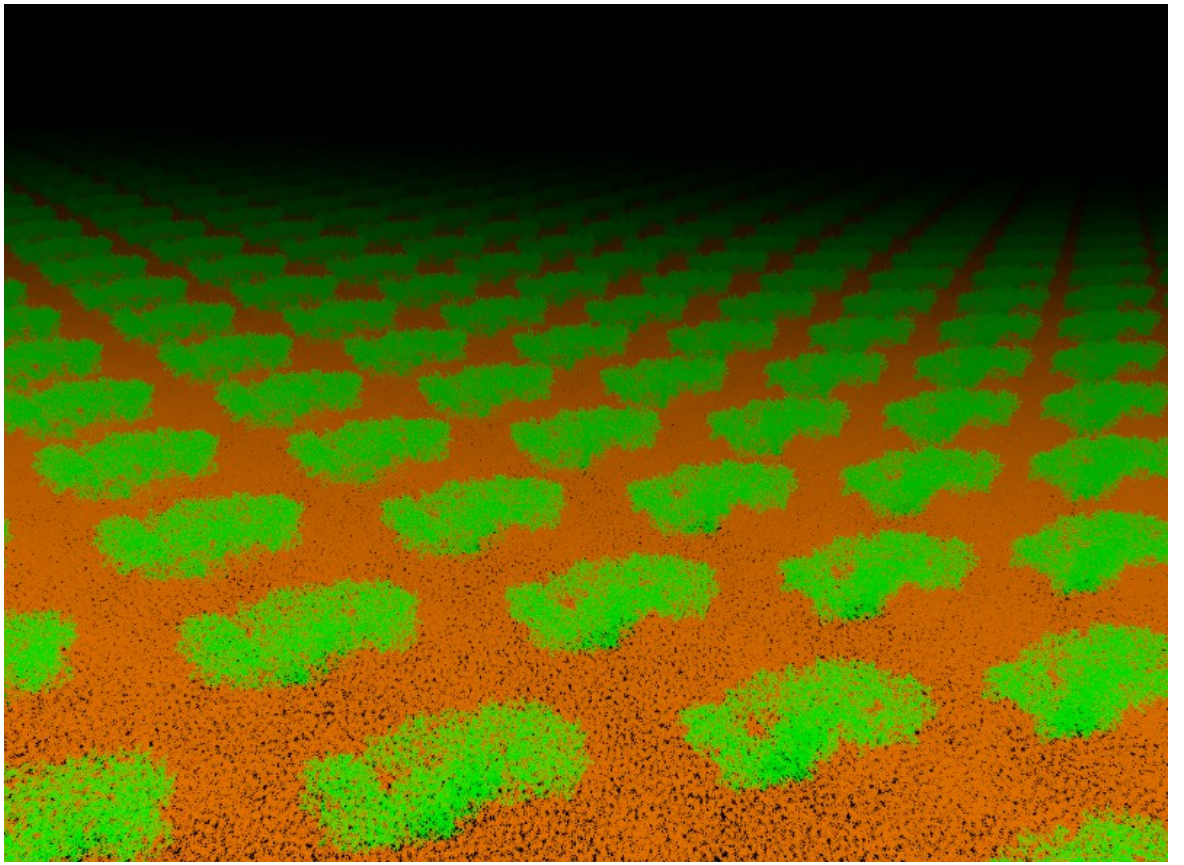
Белок
в мембране

Модель белка

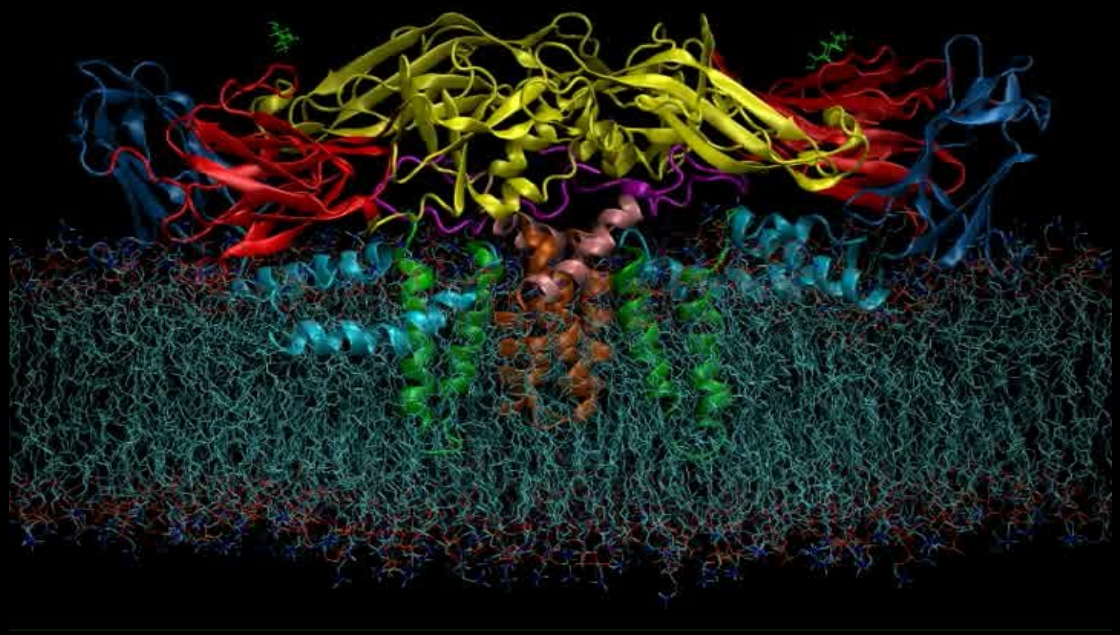


Периодические граничные условия





Моделирование динамики модели белка в среде



Modelling:

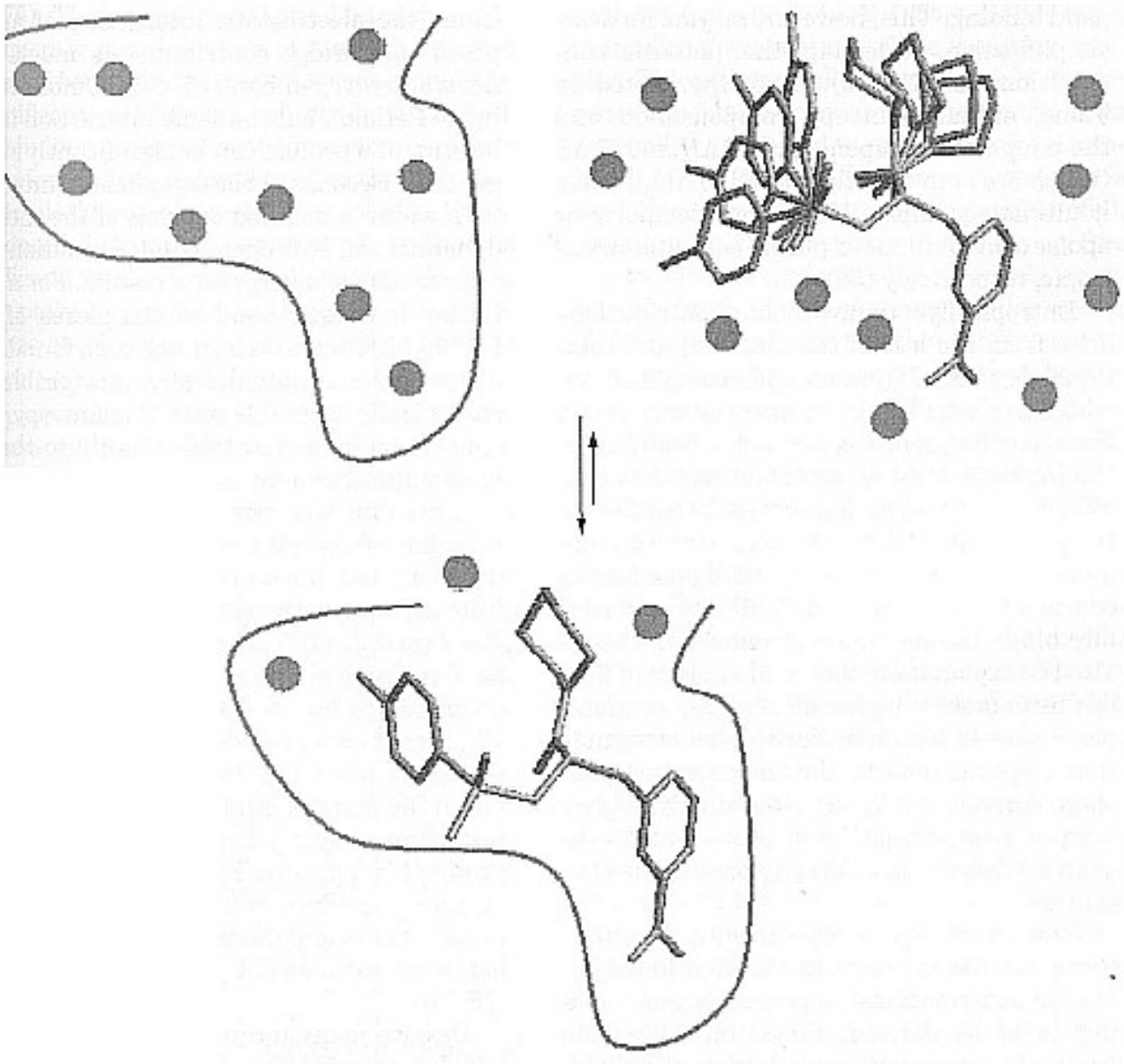
- TBEV (strain EK-328) E with NAGs and M
- POPC membrane
- NaCl 0.15 M
- 500 ns of production MD
- NAMD
- 248 435 atoms

Есть структура мишени. Что дальше?

- Докинг
- Дизайн *de novo*
- Фрагментный дизайн

1. Введение
2. Теоретические основы
3. Моделирование малых молекул
4. Моделирование больших молекул
5. Моделирование комплексов
6. Дизайн комплексов

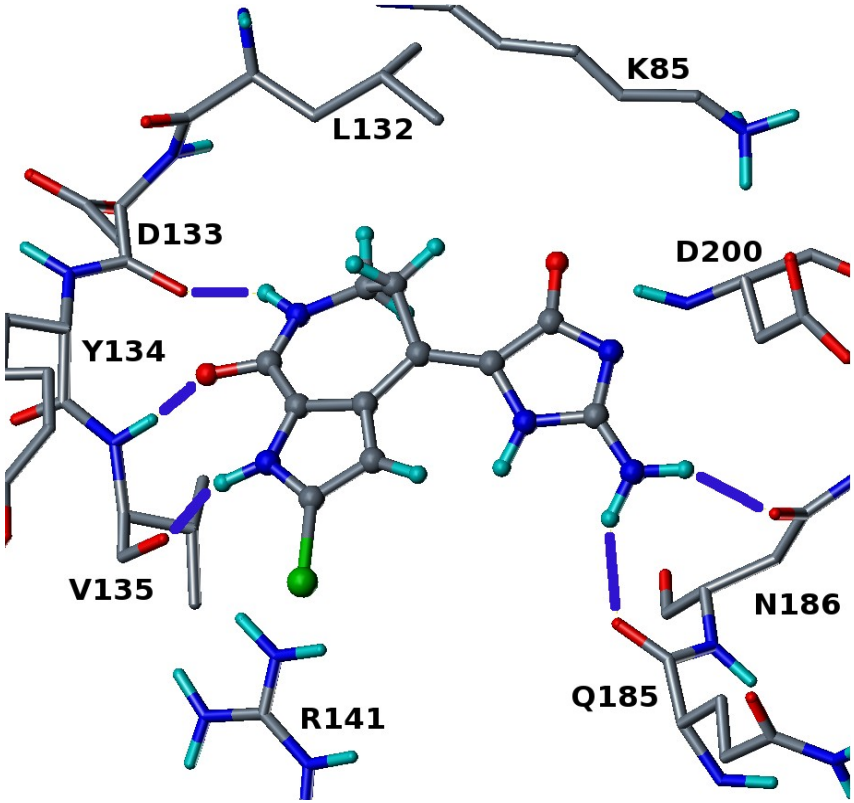
Процесс связывания белка с лигандом



Сложности: необходимость рассмотрения сольватации и энтропийных эффектов

Докинг

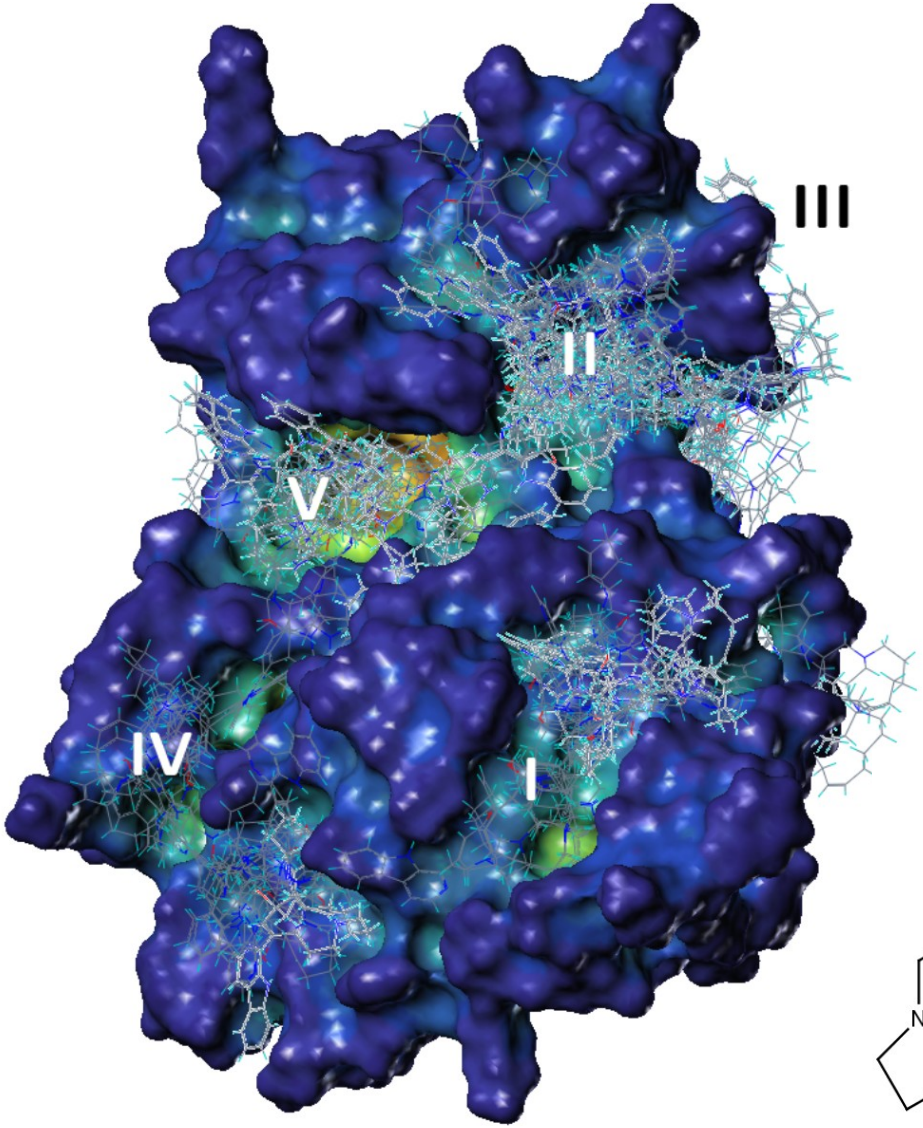
- Располагание молекулы лиганда в полости рецептора
- Универсальный метод со множеством применений
- Может быть ориентирован на скорость либо на точность прогноза
- Принципиально различные алгоритмы (методы разрабатываются с 1970-х годов)
- Варианты:
  - Жёсткий лиганд, жёсткий рецептор
  - Гибкий лиганд, жёсткий рецептор
  - Гибкий лиганд, гибкий рецептор



Докинг в медицинской химии

- Поиск первого лиганда нового рецептора
- Поиск потенциальных лигандов в библиотеке
- Поиск места связывания лиганда

Поиск потенциальных мест связывания



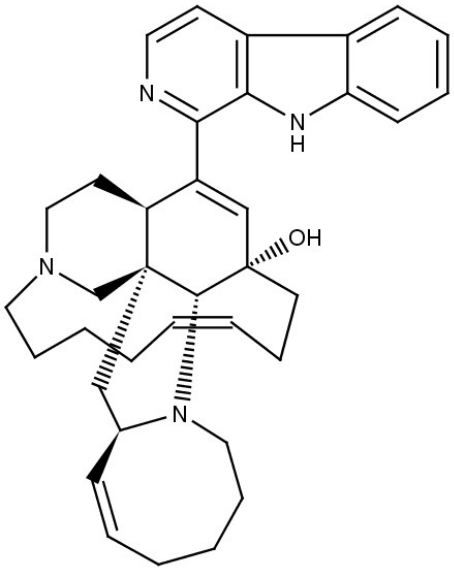
I — место связывания FRAT;

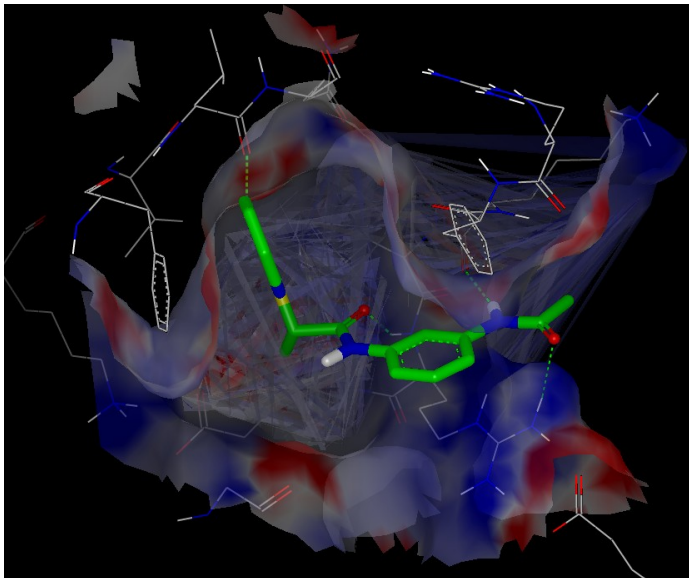
II — предполагаемое место связывания неконкурентных ингибиторов;

III — предполагаемое место связывания манзамина А;

IV — возможное место связывания аллостерических ингибиторов;

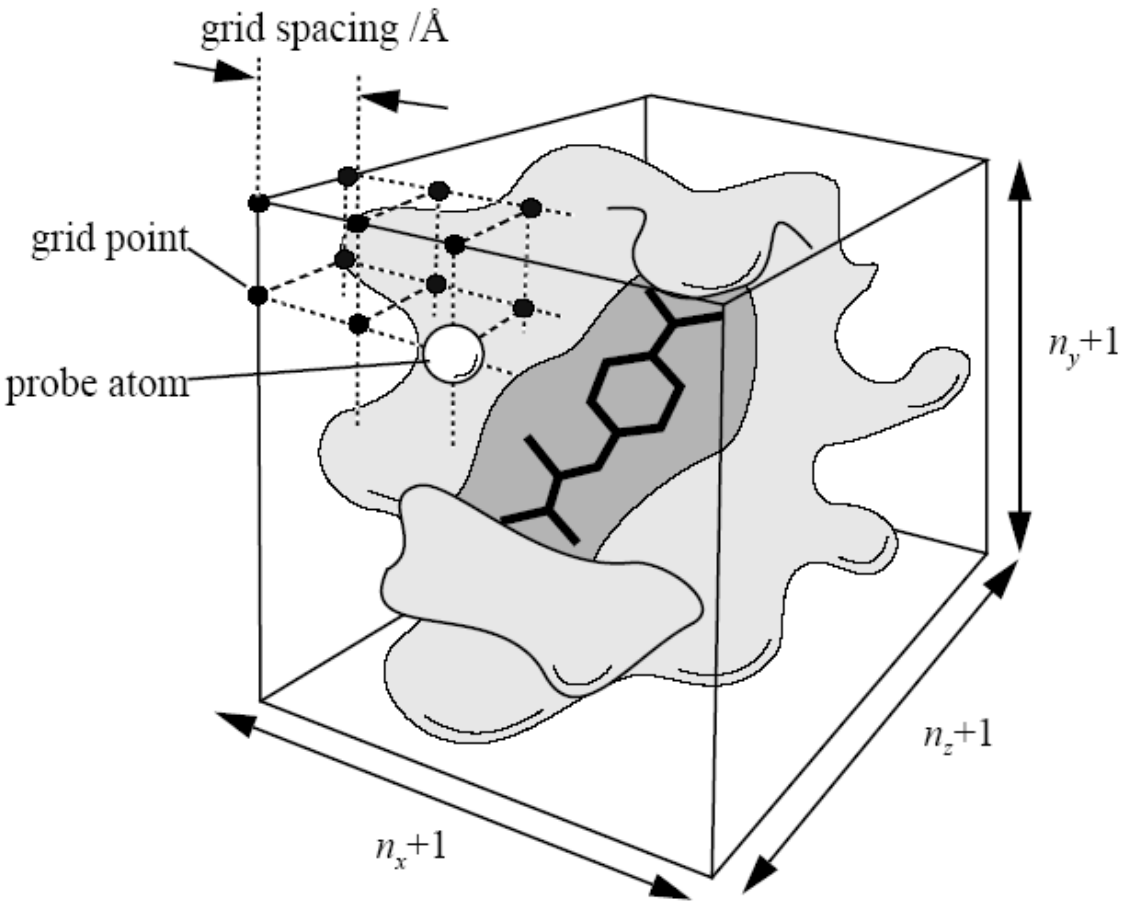
V — место связывания АТФ.

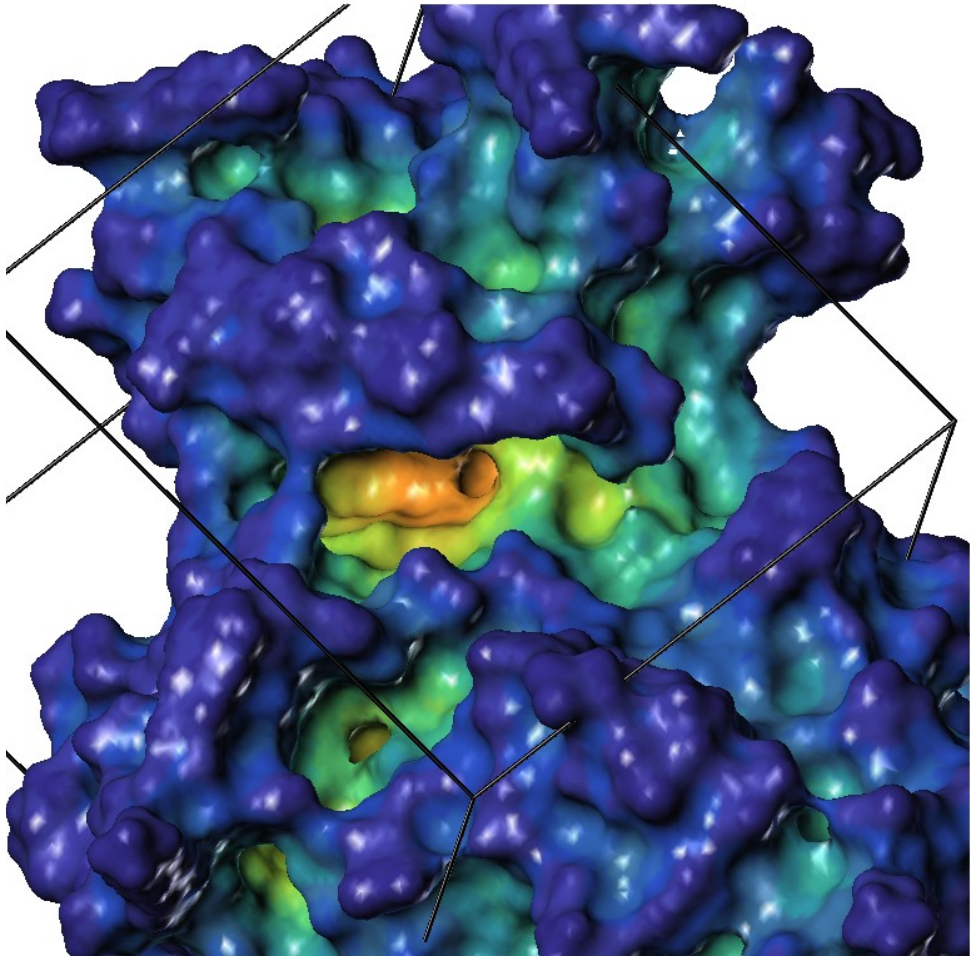




Osolodkin *et al.*, *Russ. Chem. Bull.*, **2010,** *59*, 1983-1993

Сеточные поля





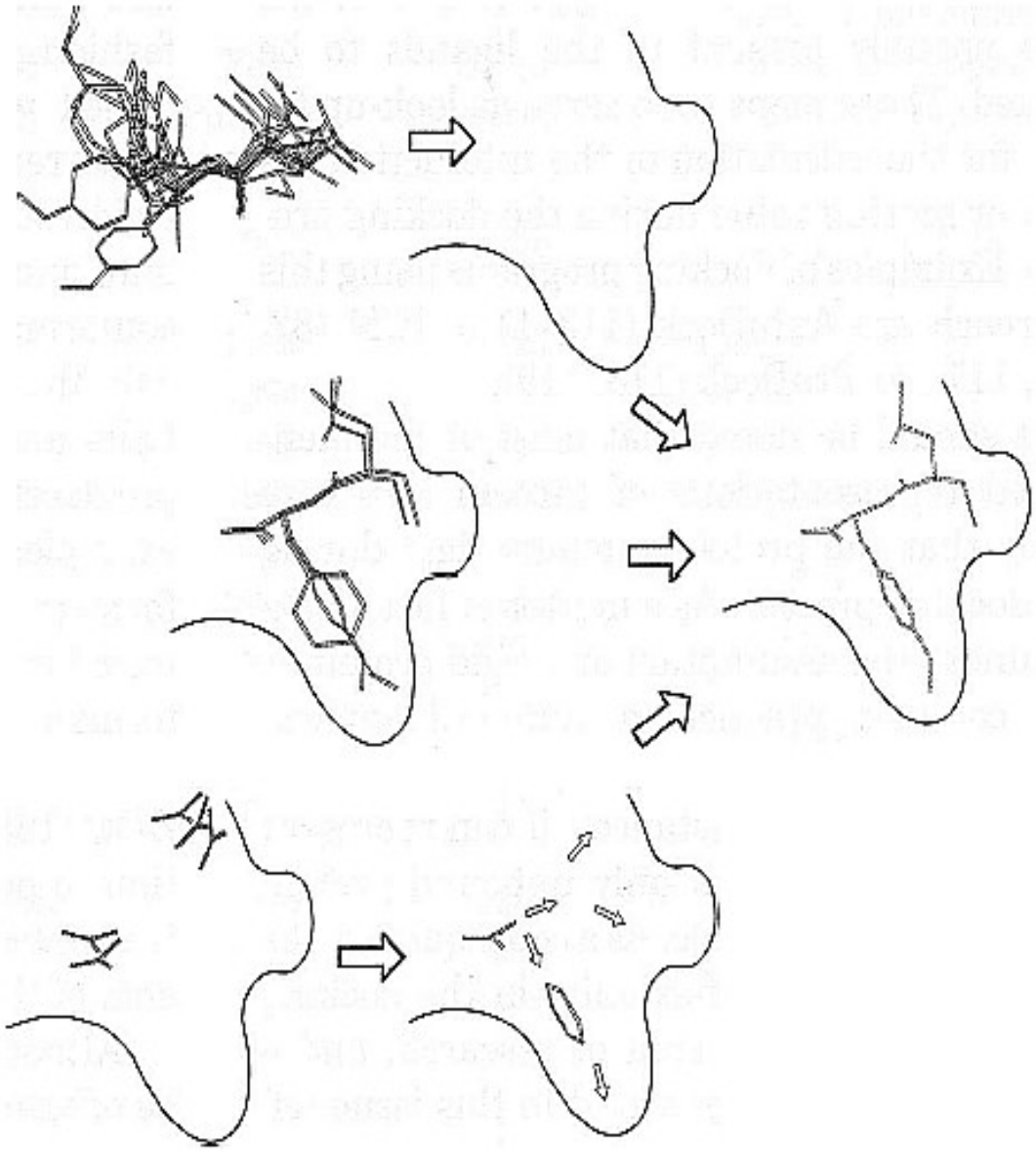
Энергия взаимодействия атомного зонда
с белком в каждой точке решётки

Стратегии докинга гибких лигандов

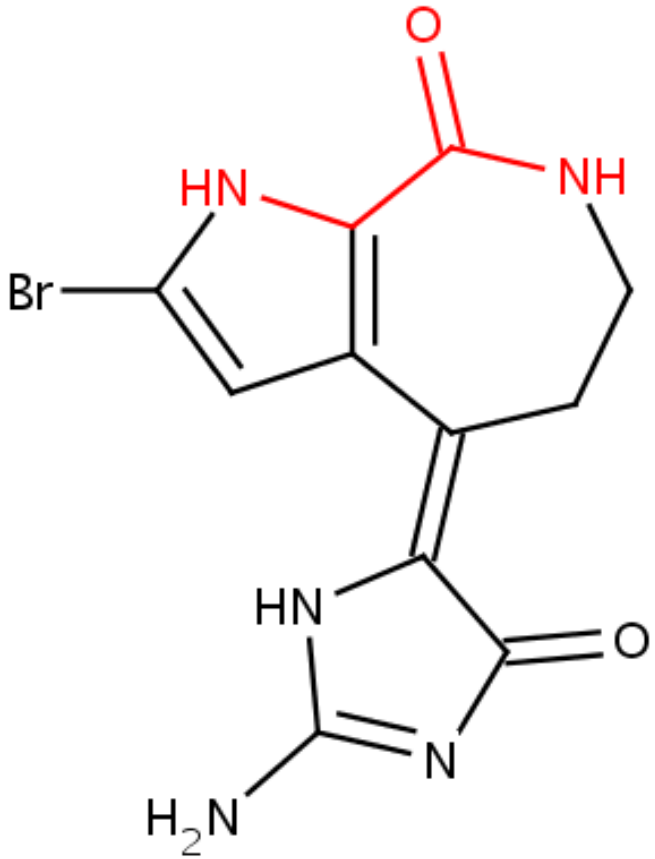
1. «Жёсткий» докинг предварительно построенного набора конформаций лиганда (FRED)

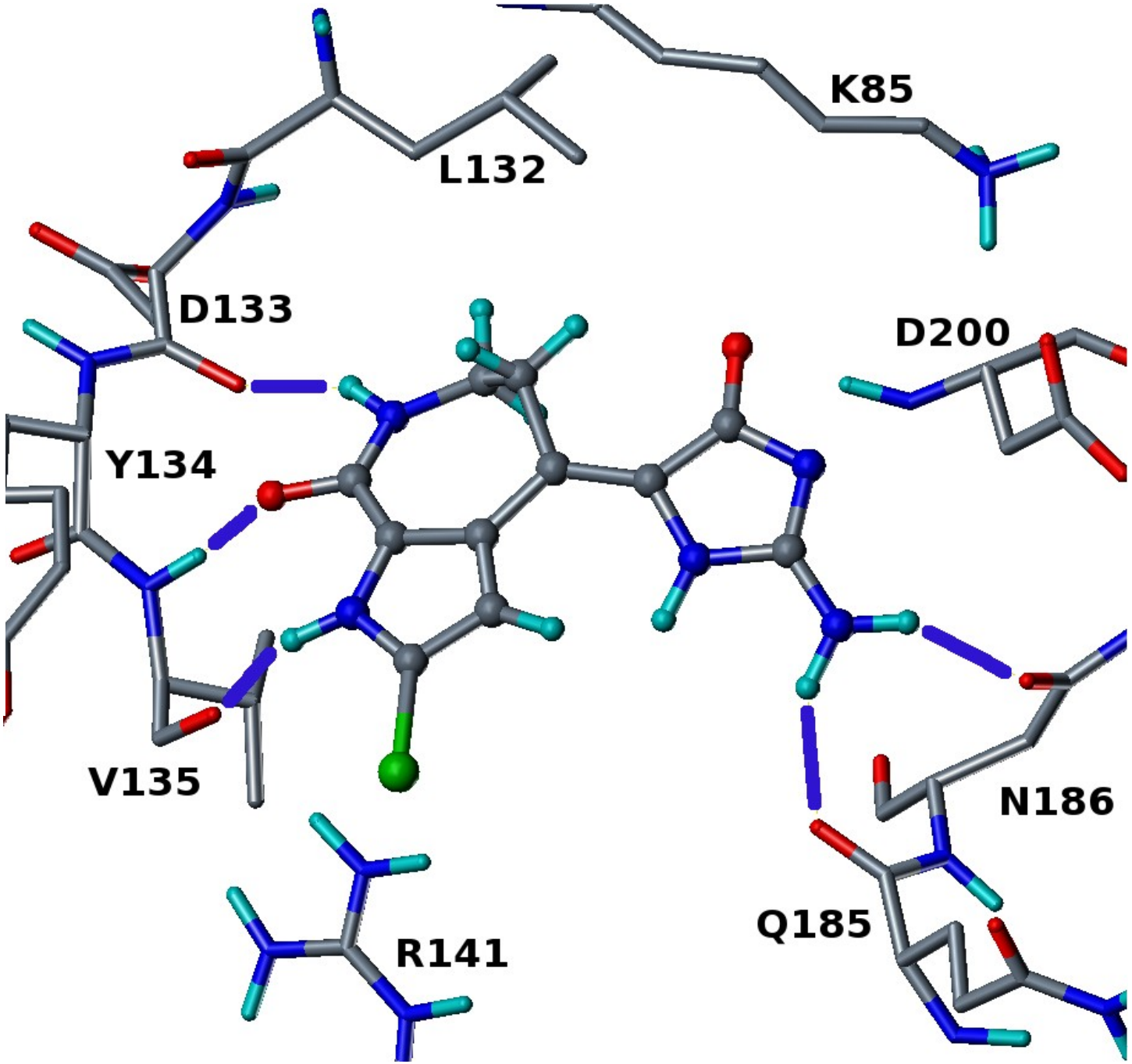
2. Одновременная оптимизация ориентации и конформации лиганда (AUTODOCK, GOLD)

3. Наращивание прикреплённых к белку якорных фрагментов (FLEXX, DOCK)



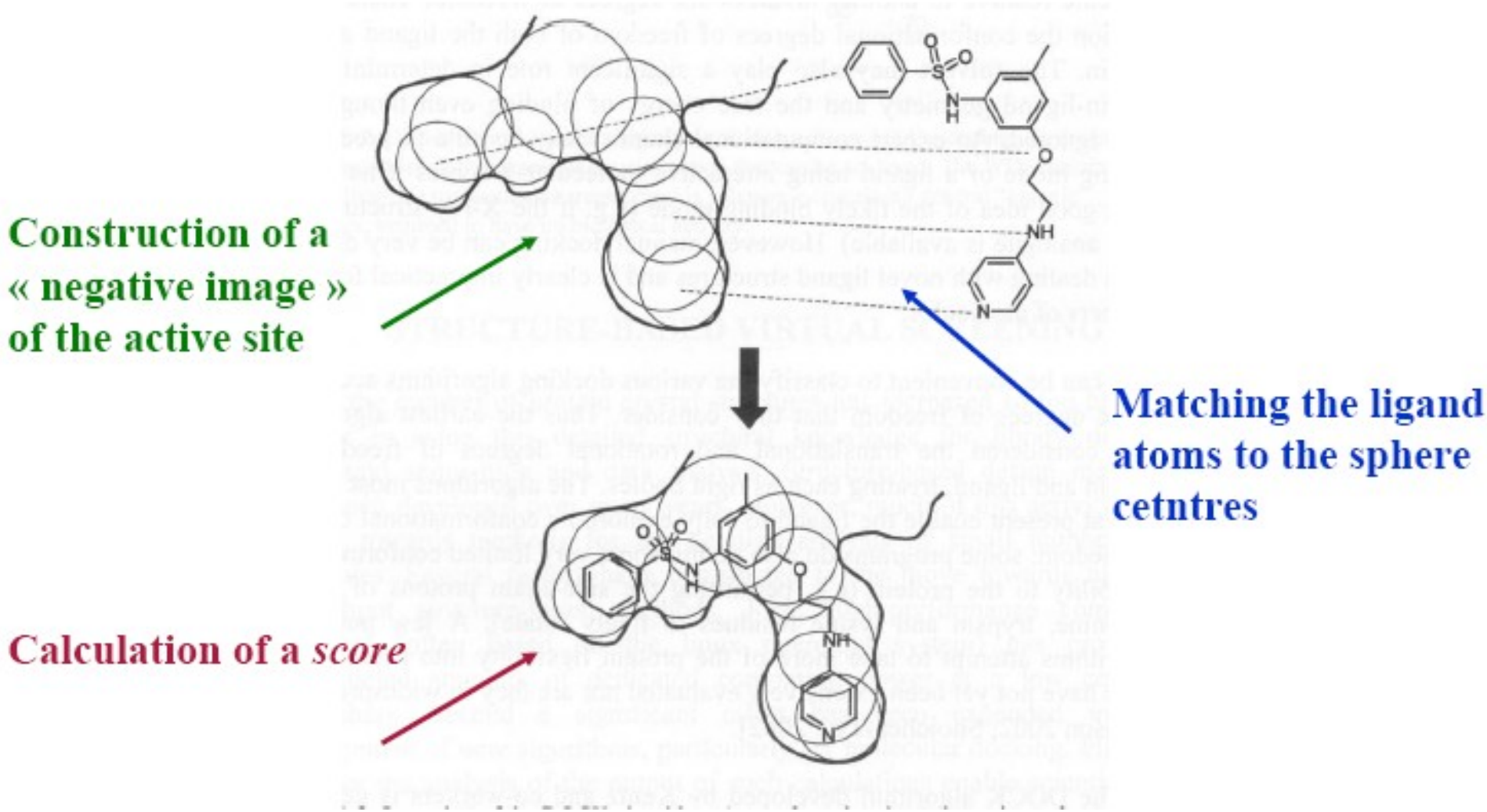
Химениальдизин: ингибитор киназы GSK-3



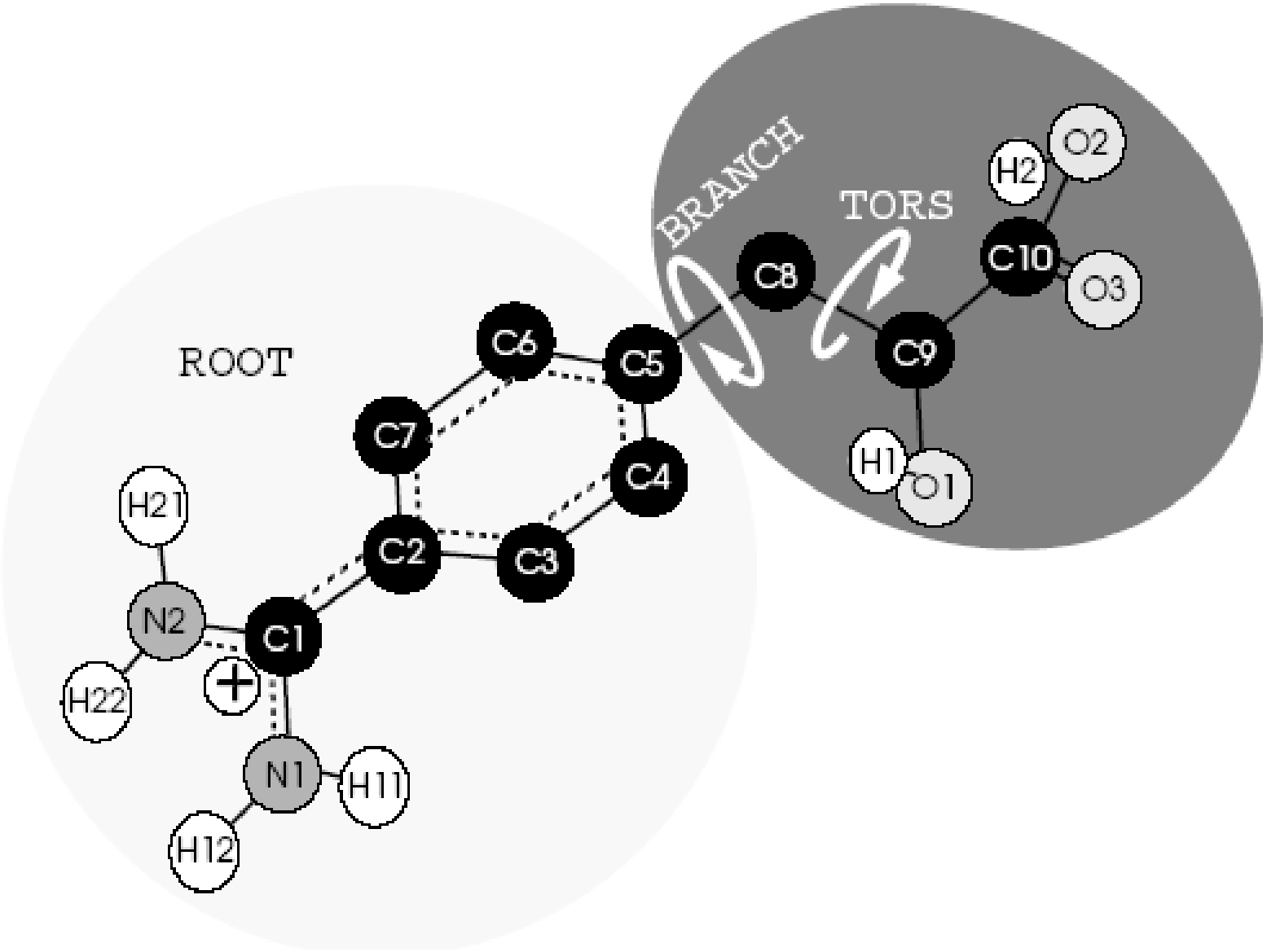


Пример жёсткого докинга

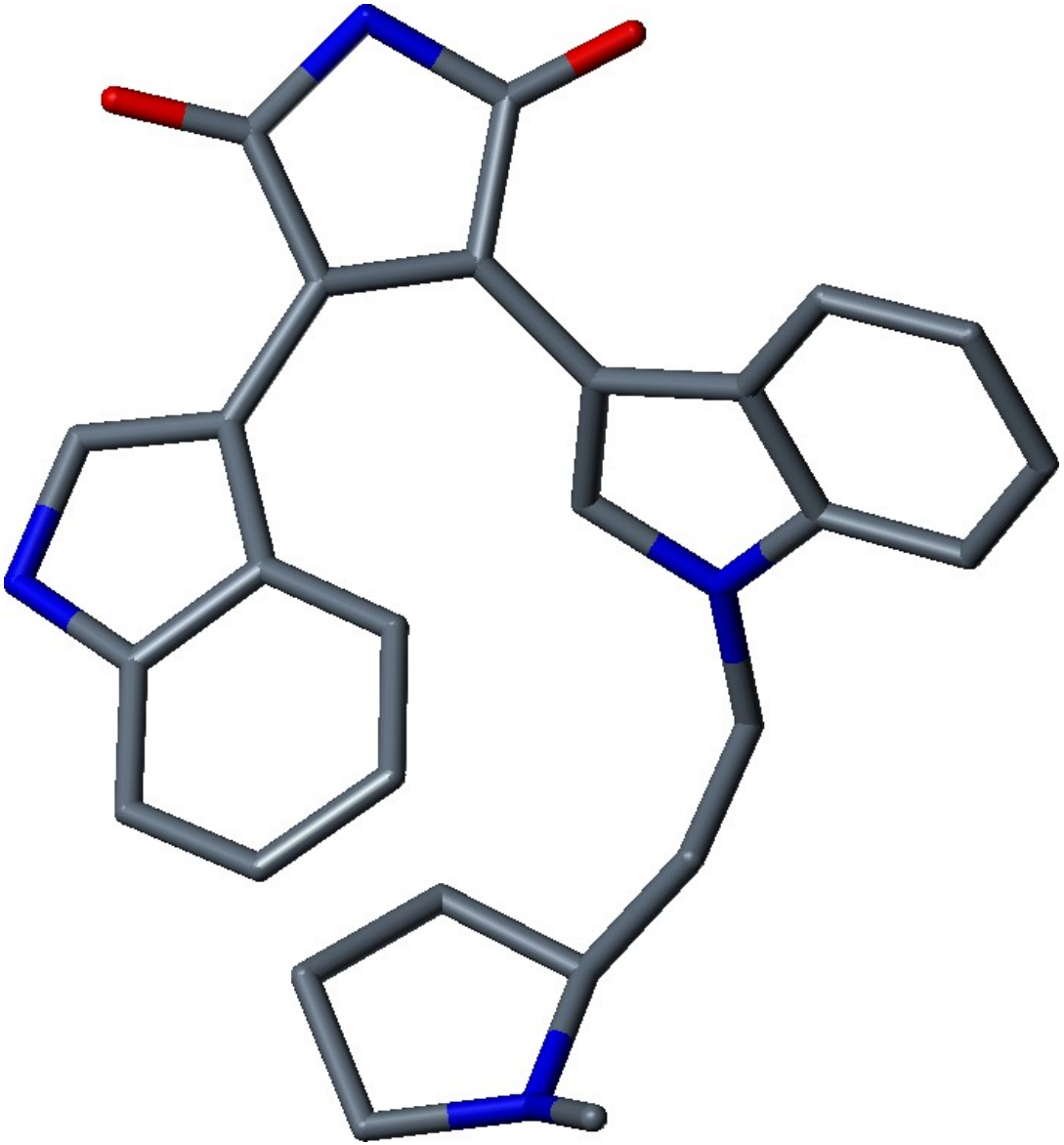
Докинг. Метод обратного образа сайта связывания (программа DOCK)



Докинг гибких лигандов при помощи AUTODOCK



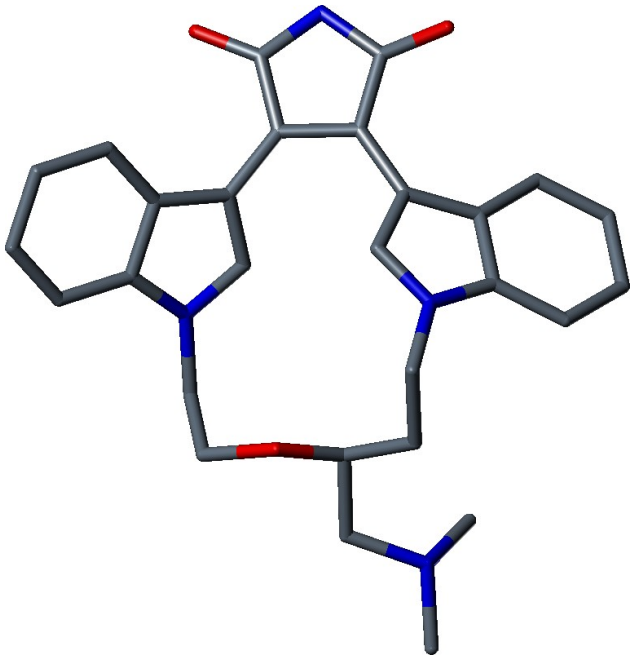
Биологически активная конформация



1SZM

свёрнутая конформация

- Характерна для всех несимметричных *бис*-арилмалеимидов
- Близка к локальному минимуму

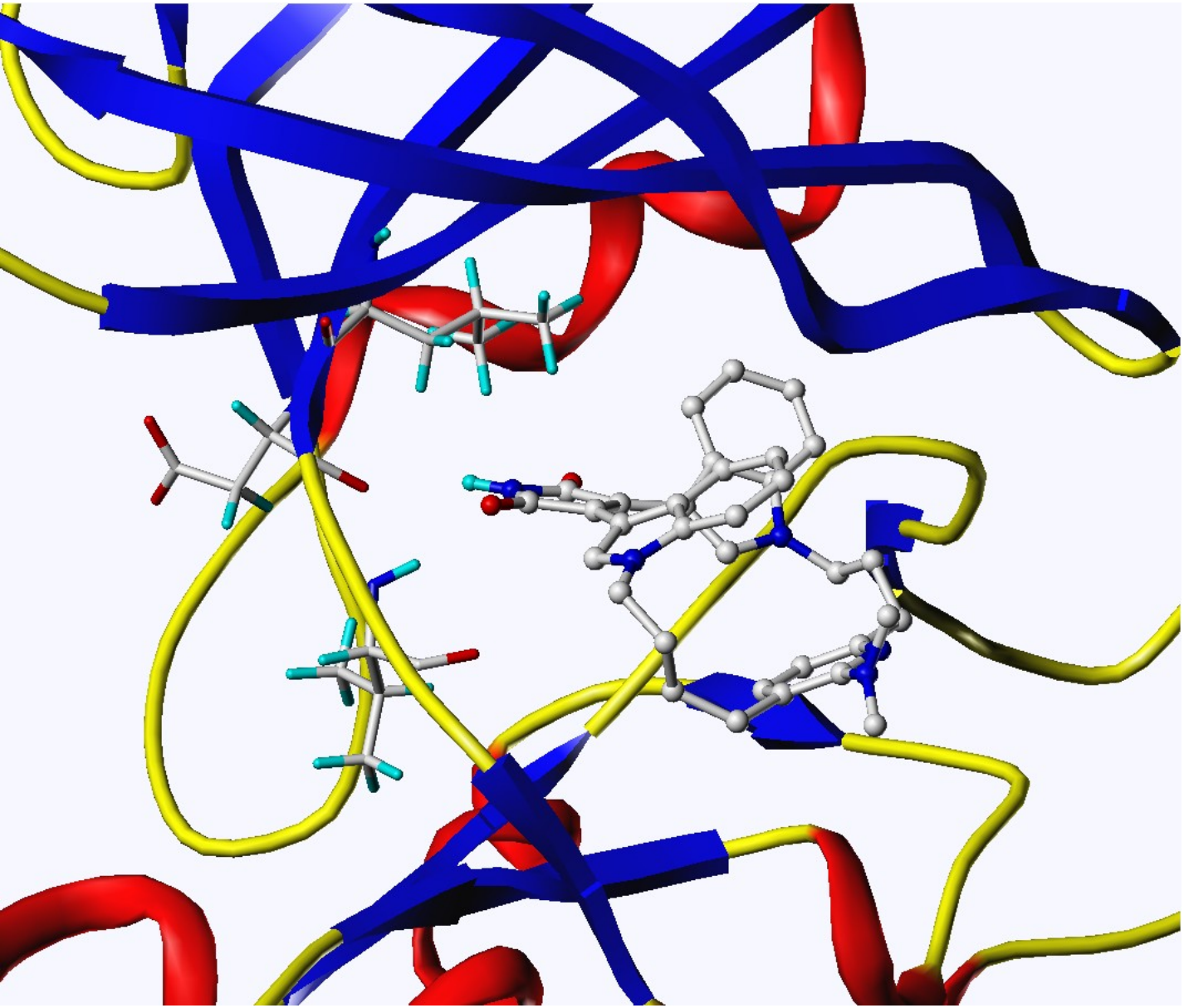


2J2I

развёрнутая конформация *зафиксирована в цикле*

Критерии отбора

Пример (2OW3)



- Свёрнутая конформация
- Консервативные водородные связи

Улучшение результатов докинга

- Учёт подвижности белка непосредственно при докинге — использование гибких боковых цепей;

- Молекулярная динамика комплексов и/или *апо*-формы белка;

- Идеально — рентген лиганд-рецепторных комплексов для новых классов лигандов.

Теоретические методы расчёта свободной энергии взаимодействия лиганда с белком

- Строгие на основе уравнений статистической физики (требуют много данных и очень много расчётов)
- Приближённые на основе анализа траекторий молекулярной динамики (требуют много расчётов)
- **Оценочные функции на основе знания геометрии комплексов лиганд-белок (быстрые, но неточные)**

Метод термодинамического интегрирования
Метод возмущения свободной энергии

A→B λ: 0→1 $k_{A\to B}=(1-\lambda)k_A+\lambda k_B$

$$\Delta G=\int_{\lambda=0}^{1}\left\langle\frac{\partial E}{\partial\lambda}\right\rangle_{\lambda}d\lambda$$

+ Очень точные результаты

– Большая часть расчётного времени тратится на состояния, не имеющие физического смысла

Метод линейной энергии взаимодействия
(Linear Interaction Energy — LIE)

$$\Delta G_{bind} = \alpha\left(\langle V^{LJ}_{l-s}\rangle_{bound} - \langle V^{LJ}_{l-s}\rangle_{free}\right) + \beta\left(\langle V^{el}_{l-s}\rangle_{bound} - \langle V^{el}_{l-s}\rangle_{free}\right)$$

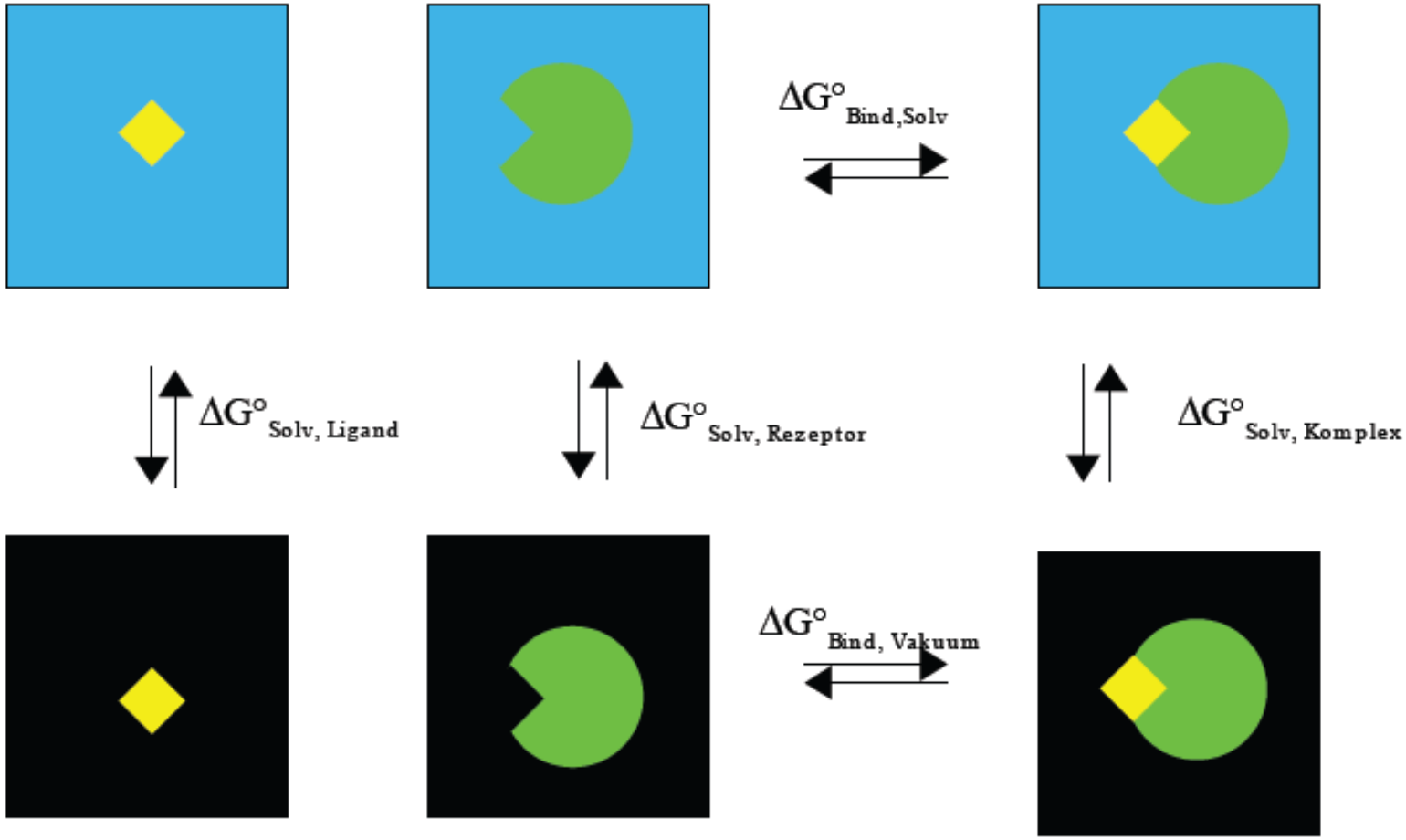
$\langle V_{l\text{-}s}\rangle_{bound}$ — средние значения вкладов ван-дер-Ваальсова (*LJ*) и электростатического (*el*) взаимодействия лиганда с белком и сольватным окружением

$\langle V_{l\text{-}s}\rangle_{free}$ — то же для взаимодействия с растворителем

α, β — эмпирические коэффициенты

Åqvist *et al.*, *Protein Eng.*, **1994,** *7*, 385-391

Вычислительная алхимия: расчёт энергии связывания белка и лиганда



$$\Delta G^0_{Bind,Solv} = \Delta G^0_{Bind,Vakuum} + \Delta G^0_{Solv,Komplex} - \left(\Delta G^0_{Solv,Ligand} + \Delta G^0_{Solv,Rezeptor}\right)$$

$$\Delta G^0_{Solv} = G^0_{elektrostatisch,\epsilon=80} - G^0_{elektrostatisch,\epsilon=1} + \Delta G^0_{Hydrophob}$$

$$\Delta G^0_{Vakuum} = \Delta E^0_{Molekülmechanik} - T \cdot \Delta S^0_{Normalmodenanalyse}$$

Эмпирические оценочные функции

— Энергия гидрофобного взаимодействия (lipo)

— Вклад подвижных связей в энтропию (rot)

— Энергия водородных связей (hb)

— Энергия солевых мостиков (io)

— Энергия ароматических взаимодействий (ar)

$$\Delta G = \Delta G_0 + \Delta G_{rot} N_{rot} + \Delta G_{hb} \Sigma f(\Delta R, \Delta\alpha) + \Delta G_{io} \Sigma f(\Delta R, \Delta\alpha) + \Delta G_{ar} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} f^*(\Delta R)$$

ChemScore, ScreenScore, FlexX, PLP, GlideScore, Ludi

Оценочные функции, основанные на знании

Параметризуются на основе частоты встречаемости определённых межатомных взаимодействий в известных базах данных (PDB, CSD, etc.)

$$\mathrm{PMF_score} = \sum_{\substack{kl \\ r < r^{ij}_{\mathrm{cut-off}}}} A_{ij}(r)$$

$$A_{ij}(r) = -k_{\mathrm{B}} T \ln\left[f^{j}_{\mathrm{Vol_corr}}(r) \, \frac{\rho^{ij}_{\mathrm{seg}}(r)}{\rho^{ij}_{\mathrm{bulk}}} \right]$$

SMoG, PMF, DrugScore, ...

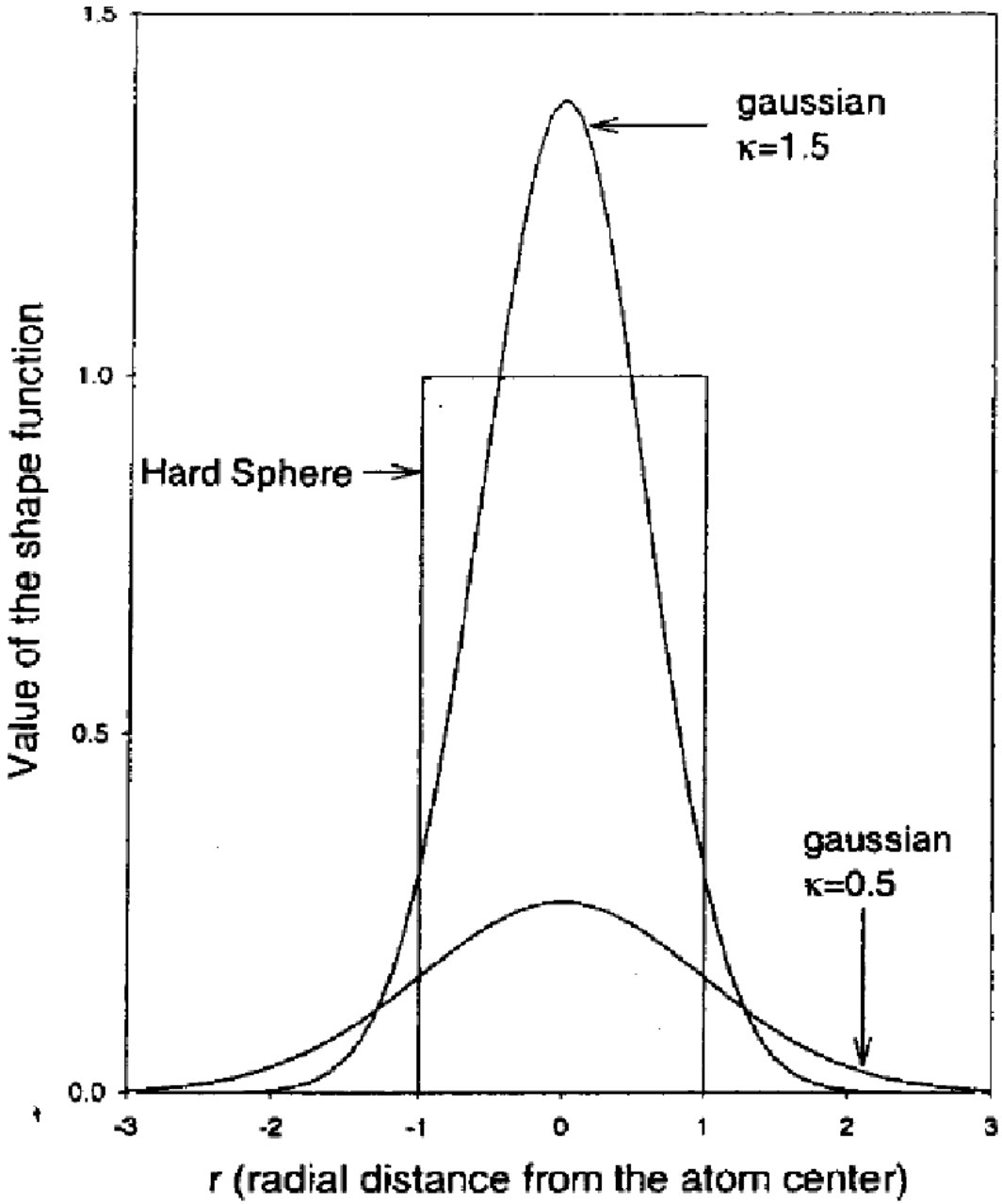
Оценочные функции на основе силовых полей

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332 \frac{q_i q_j}{D r_{ij}} \right]$$

Энергия нековалентных взаимодействий — потенциалы Леннарда-Джонса и Кулона

AutoDock, Dock, Gold...

Гауссовы оценочные функции



$$\kappa = \alpha R^2$$

$$g(r) = \frac{4}{3}\left(\frac{\kappa^3}{\pi}\right)^{1/2} \exp\left(-\frac{\kappa}{R^2} r^2\right)$$

Форма молекулы представляется в виде суперпозиции гауссовых функций, центрированных на атомах. Энергия взаимодействия оценивается на основе перекрывания объёмов лиганда и рецептора

1. Введение
2. Теоретические основы
3. Моделирование малых молекул
4. Моделирование больших молекул
5. Моделирование комплексов
6. Дизайн комплексов

Дизайн лигандов белков *de novo*

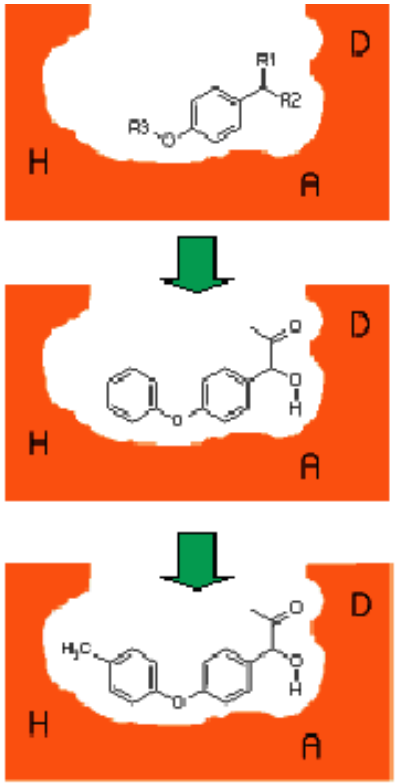
LUDI, LeapFrog, LigBuilder, etc.

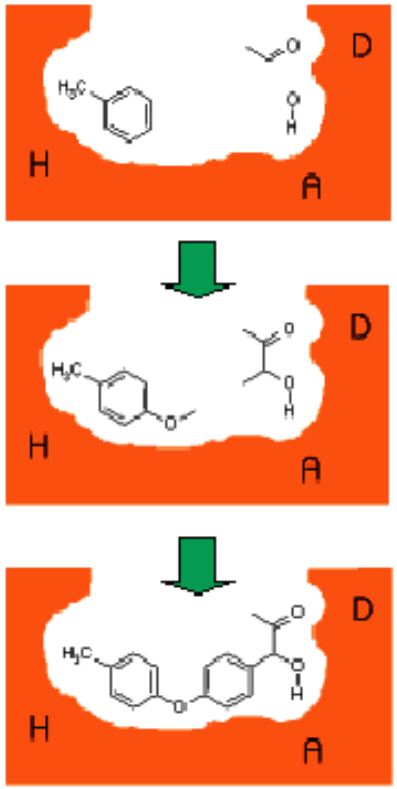
Grow

(метод наращивания затравки)

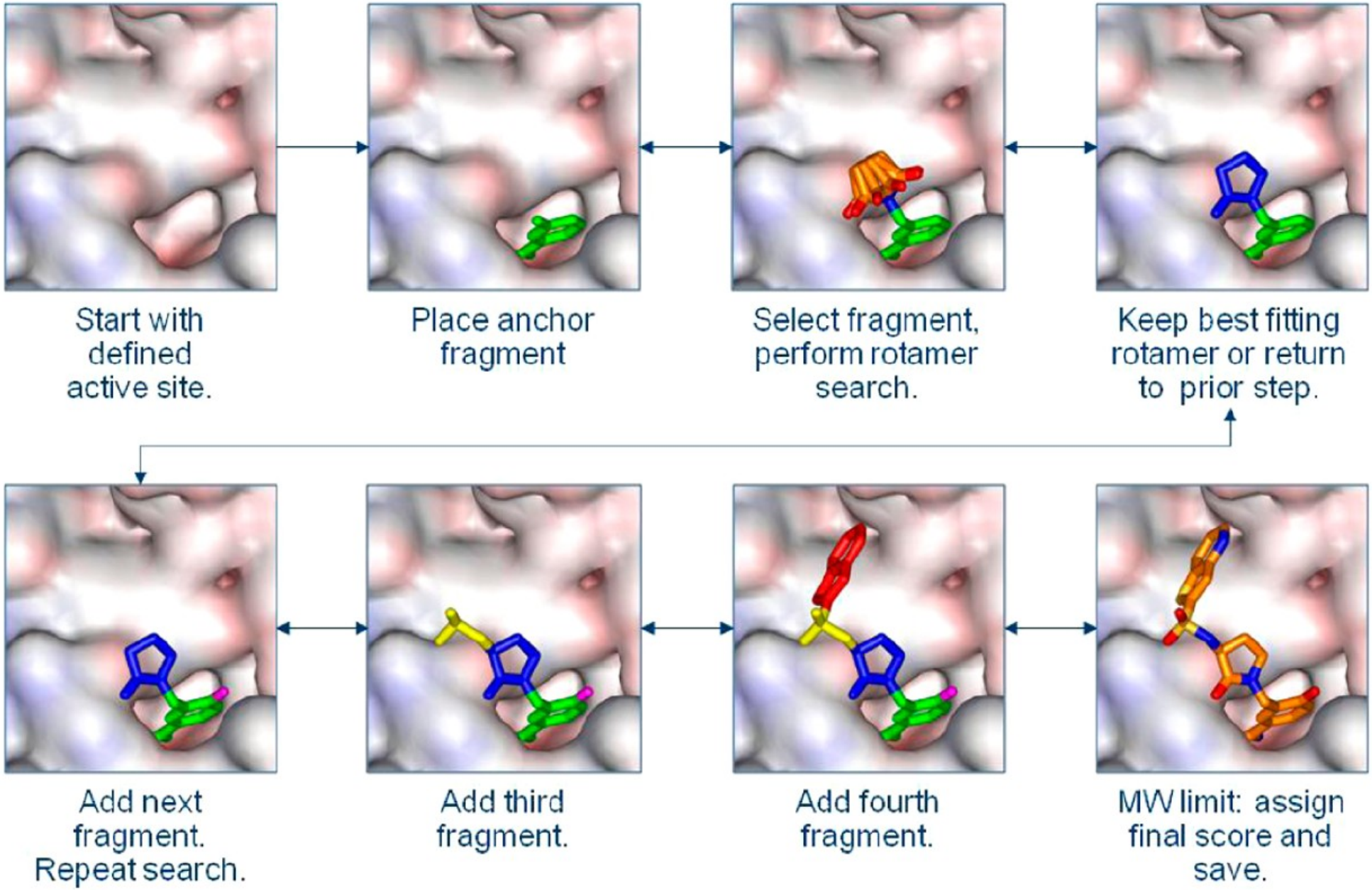
Link

(объединение важных
для связывания фрагментов)





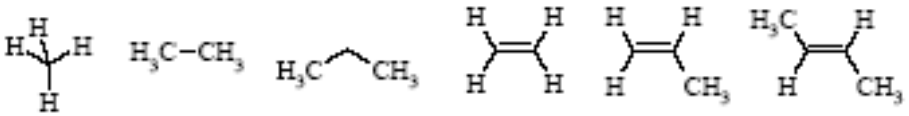
Методика дизайна *de novo* (Contour)



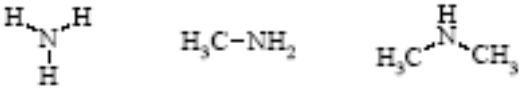
Ishchenko *et al.*, *J. Chem. Inf. Model.*, ***2012,*** *52*, 2089-2097

Дизайн *de novo*. Строительные блоки

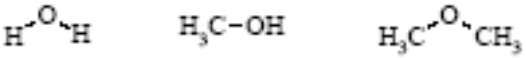
углеводороды



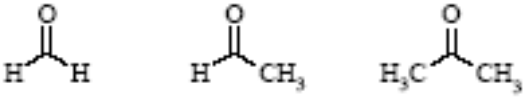
амины



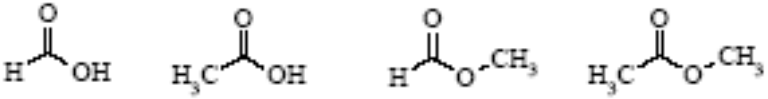
спирты и эфиры



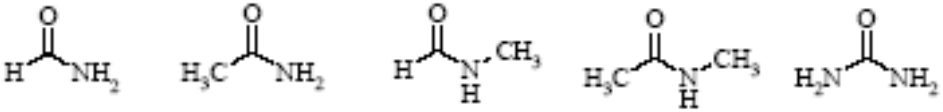
альдегиды и кетоны



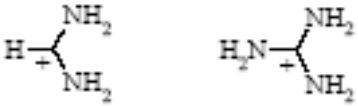
карбоновые кислоты и сложные эфиры



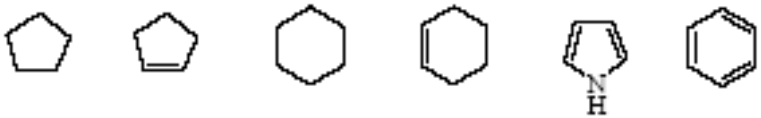
амиды



амидиновые и гуанидиновые группы

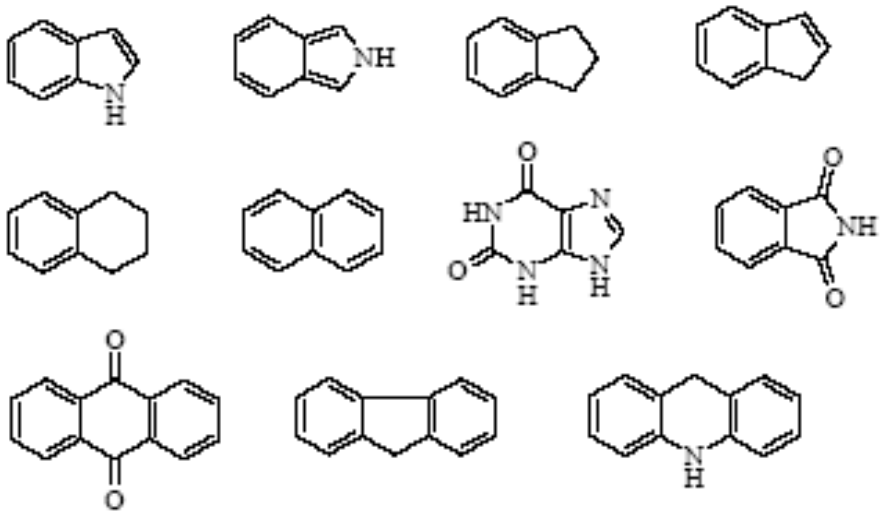


простые циклы

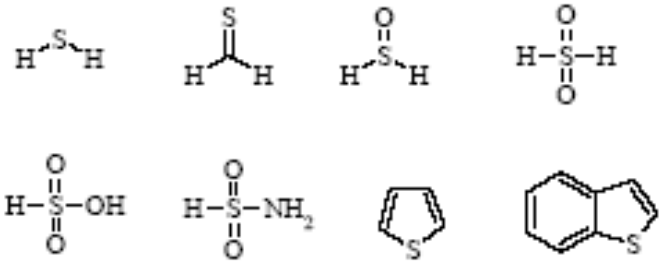


Дизайн *de novo*. Строительные блоки

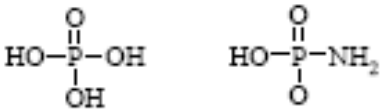
полициклы



S-содержащие фрагменты

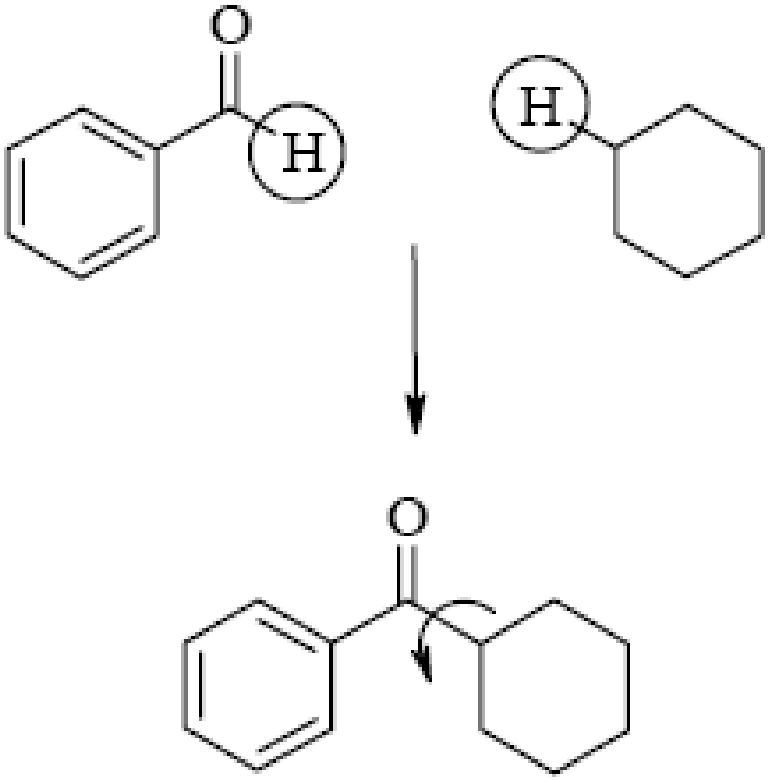


P-содержащие фрагменты

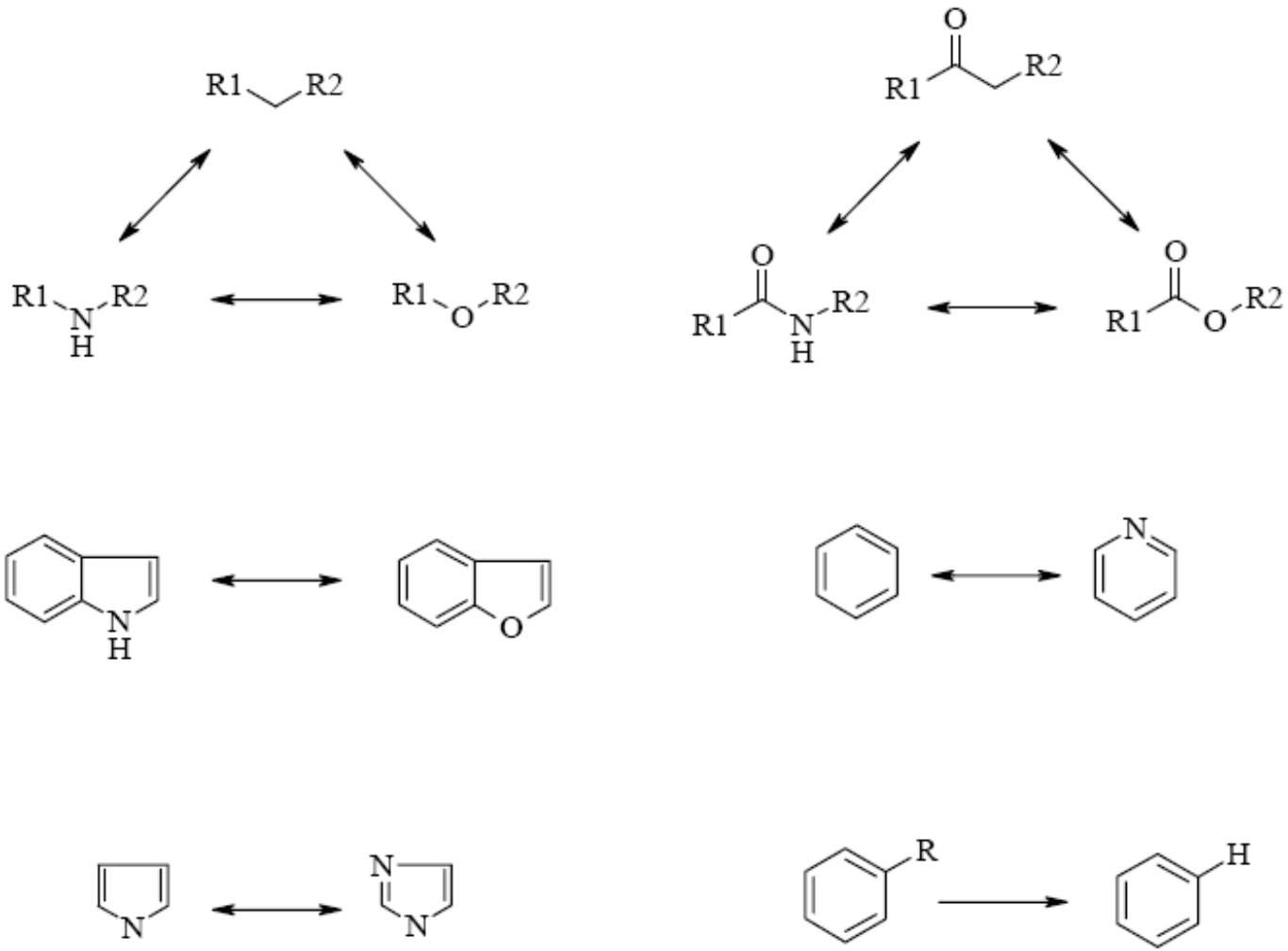


галогены

Операция роста в дизайне *de novo*

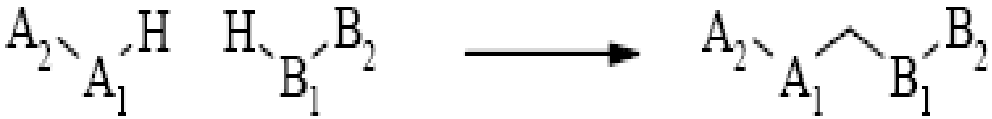


Операция мутации в дизайне *de novo*



Операция связывания в дизайне *de novo*

наведение моста

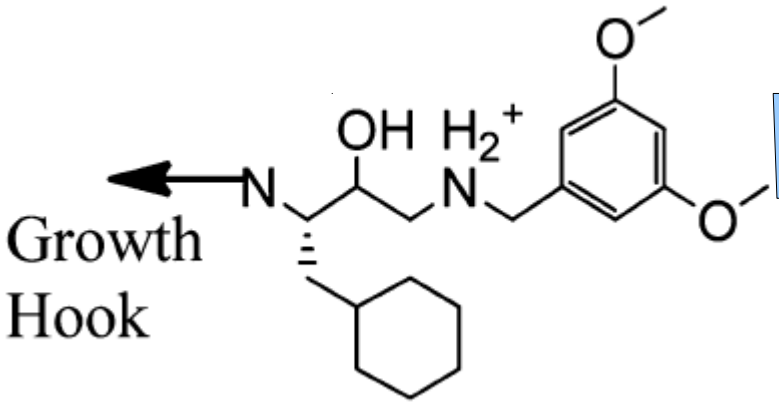


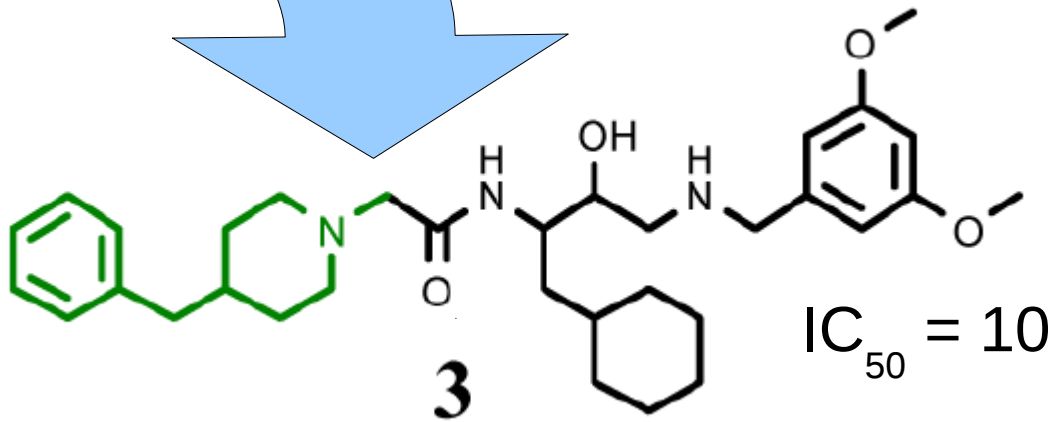
соединение

слияние

Дизайн ингибиторов ренина *de novo*

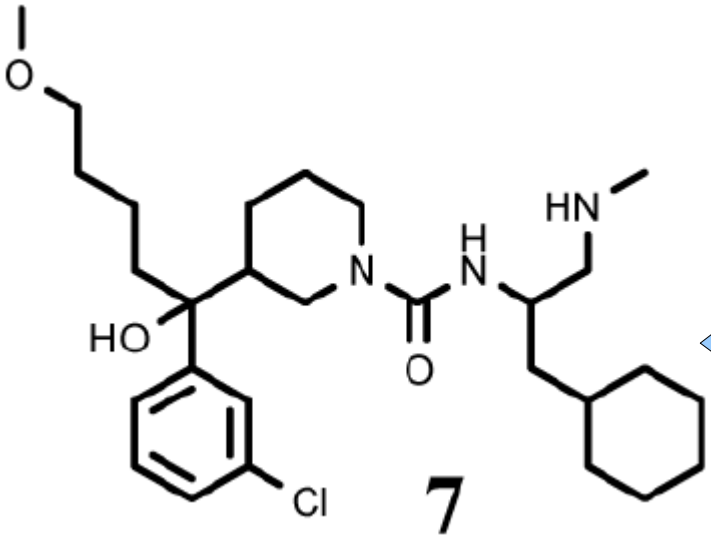
Ингибиторы протеазы **ренина** могут использоваться для лечения сердечно-сосудистых заболеваний





IC_{50} = 10 μM

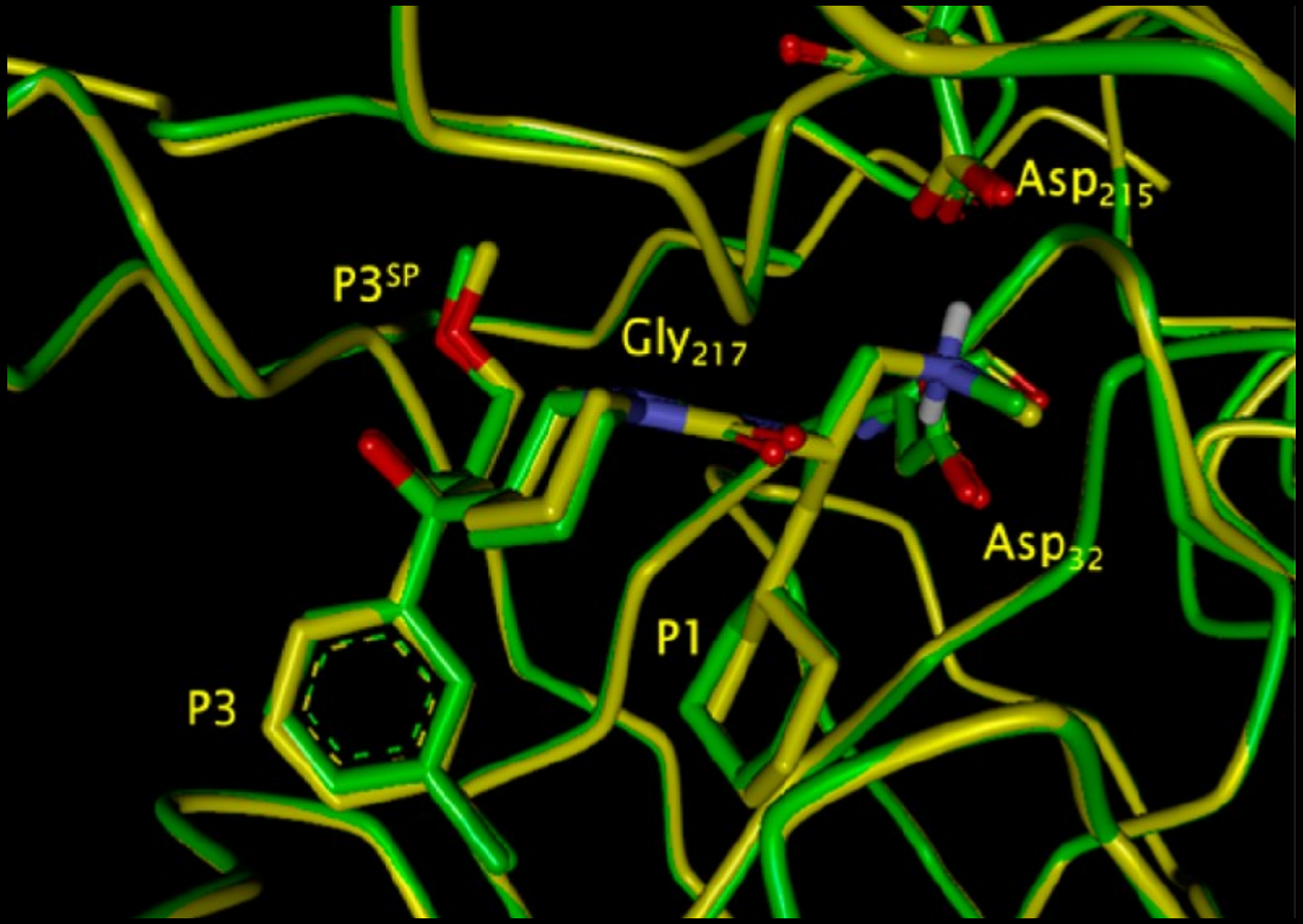
Несколько раундов оптимизации структуры



IC_{50} = 0.5 nM

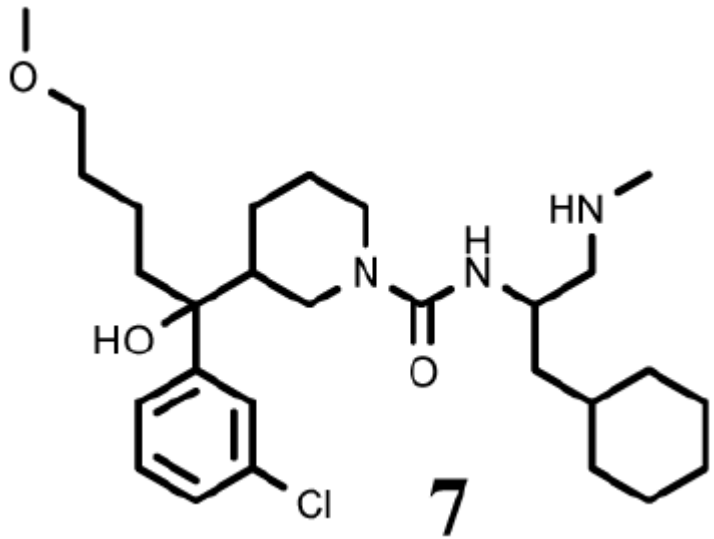
Ishchenko *et al.*, *J. Chem. Inf. Model.*, ***2012,*** *52*, 2089-2097

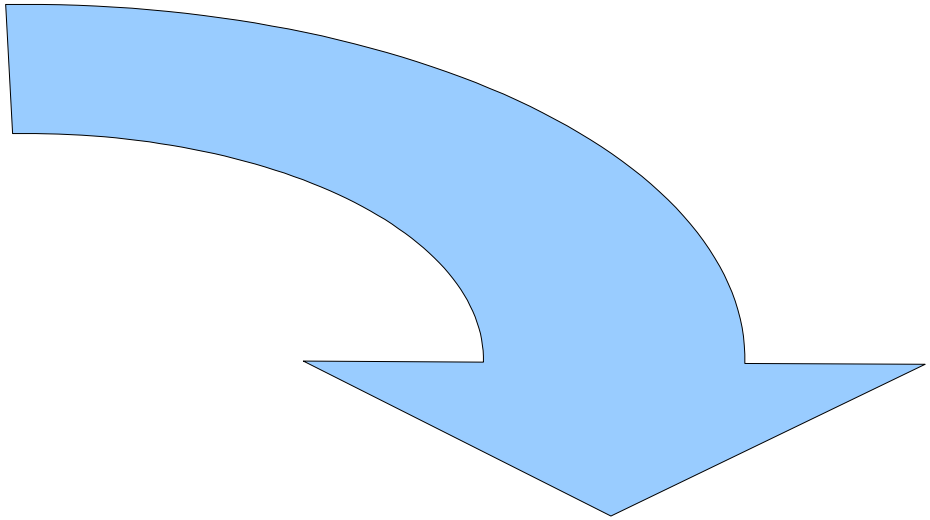
RMSD лиганда = 0.25 Å!

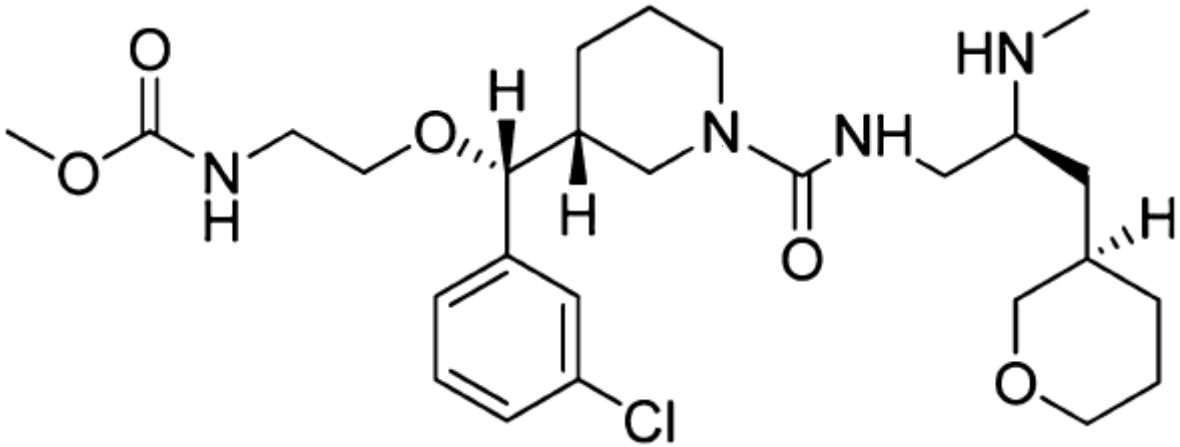


Ishchenko *et al.*, *J. Chem. Inf. Model.*, ***2012,*** *52*, 2089-2097

Однако потребовалась оптимизация биодоступности...







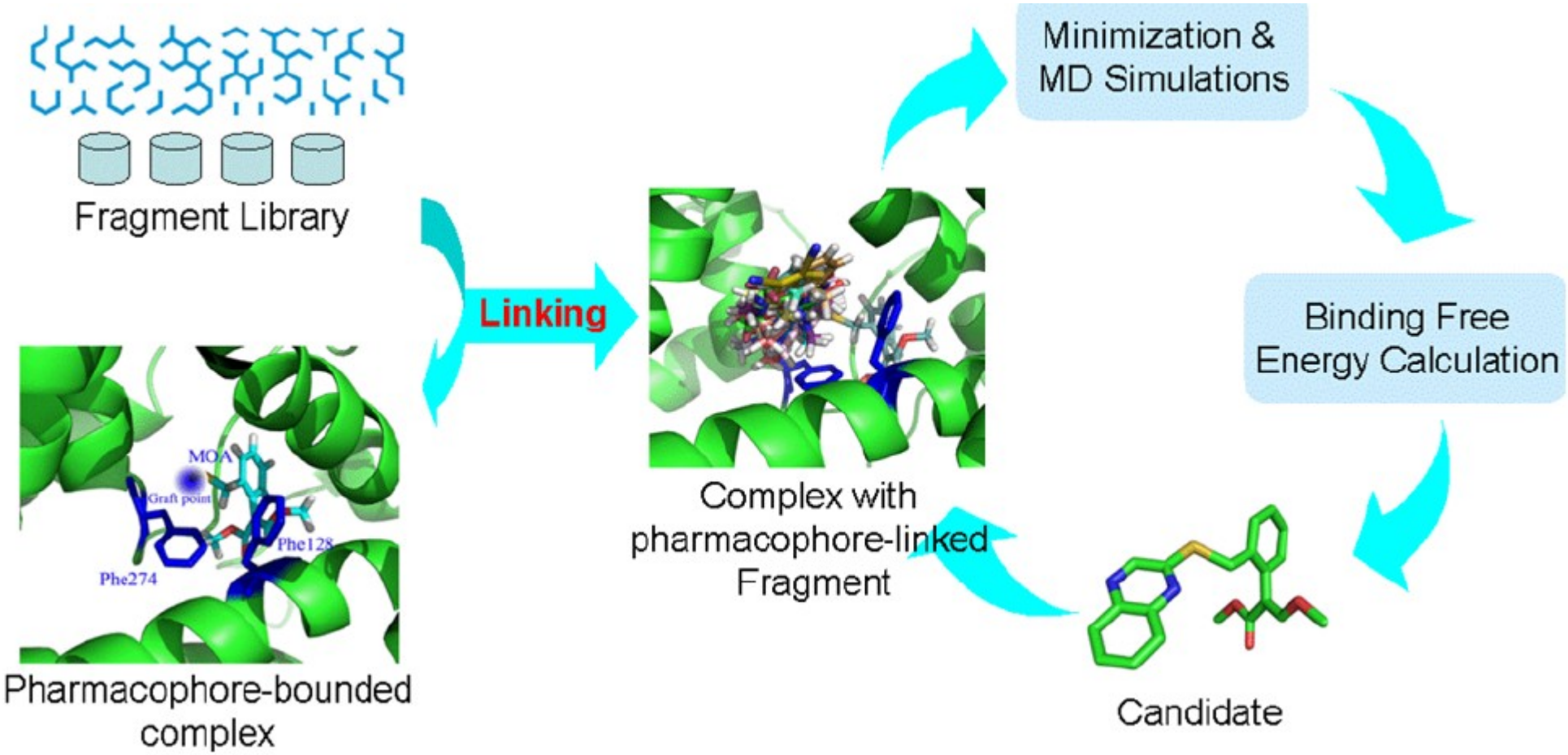
Ishchenko *et al.*, *J. Chem. Inf. Model.*, ***2012,*** *52*, 2089-2097

Фрагментный дизайн

http://practicalfragments.blogspot.com

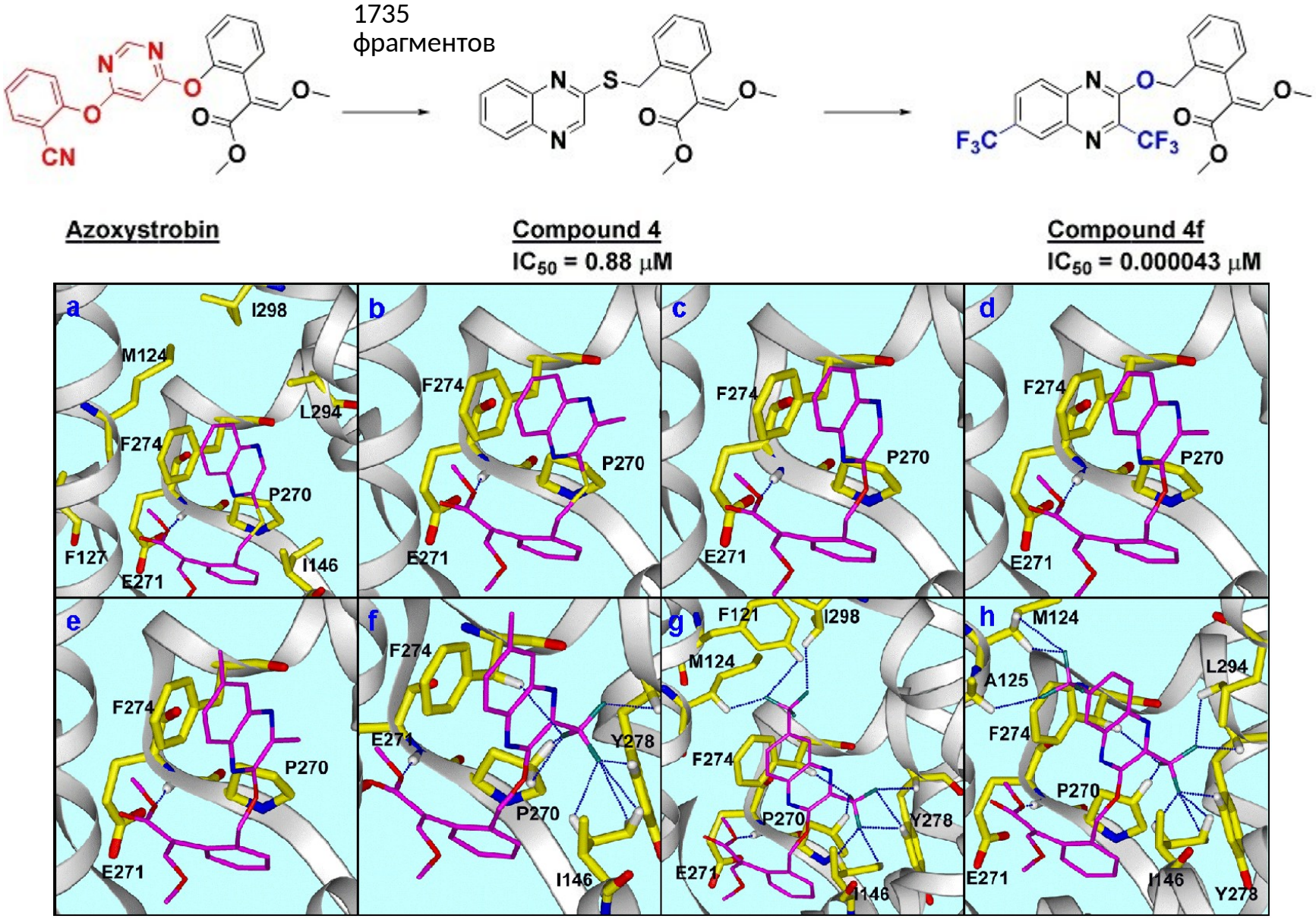
- Близок по сути к дизайну *de novo;*
- Приводит к более доступным для синтеза соединениям;
- Опирается на **библиотеку фрагментов,** построенную определённым образом;
- Можно проводить без использования моделирования

Методика фрагментного дизайна



G.-F. Hao *et al.*, *JACS*, **2012,** *134*, 11168-11176

Фрагментный дизайн ингибитора цитохрома bc_1



Ингибиторы цитохрома bc_1 могут быть использованы как фунгициды

G.-F. Hao *et al.*, *JACS*, **2012,** *134*, 11168-11176

Литература

• **Biomolecular Simulations: Methods and Protocols.** L. Monticelli, E. Salonen, Eds. Methods in Molecular Biology, vol. 924. Springer, 2013. DOI: 10.1007/978-1-62703-017-5 (доступна на факультете).

• Х.-Д. Хёльтье и др. **Молекулярное моделирование: теория и практика.** 2-е изд. М., Бином, 2013.