

Лекция по Компьютерному Моделированию Биологических Объектов:

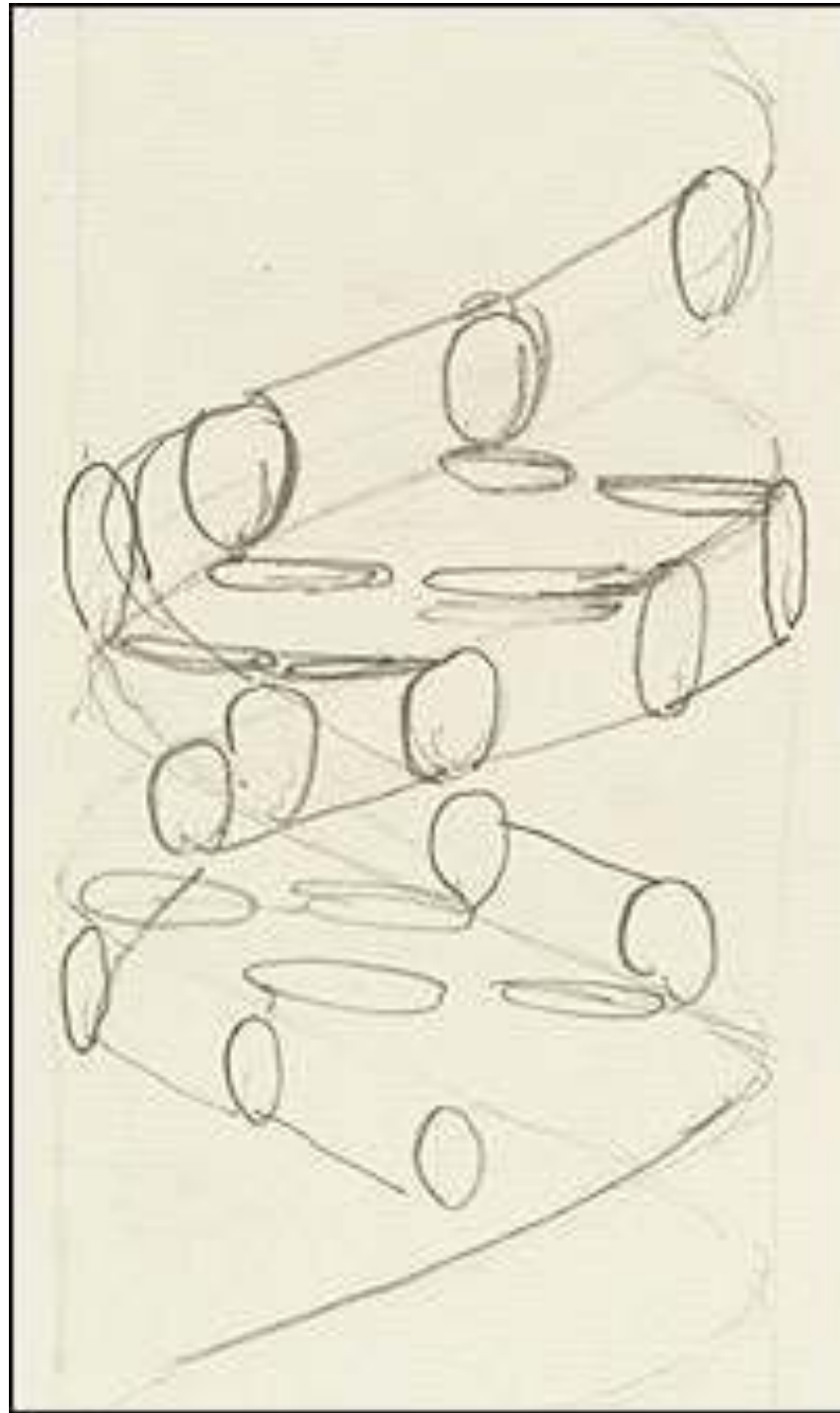
1. Моделирование по Гомологии
2. Метод Монте-Карло
3. Метод Молекулярной Динамики
4. QSAR

Сергей Грудинин, 29.04.2010

Почему Компьютерное Моделирование?

- Компьютерная разработка лекарств
 - Рынок антибиотиков - \$25.0 миллиардов (Nature Biotechnology 2006), “Tufts Report” предполагает, что стоимость разработки и открытия нового лекарства достигла \$800 миллионов для каждого лекарства, доведенного до продаж
 - Огромная потребность в новых лекарствах: За последние 40 лет открыто всего 3 полностью новых антибиотика
 - Время выхода на рынок
 - Понимание процессов
- Биохимические процессы, которые не видны в лабораторных экспериментах
 - Можно посмотреть на редкие события
 - Атомарные детали
 - Фолдинг белков
 - Белок-белок и блок-лиганд взаимодействия
- Дешевле лабораторных экспериментов
 - Платенсимицин: протестировано 250,000 кандидатов
- Нано-дизайн
 - Нано рынок: \$1 триллион к 2015 году (NSF report 2001)

Классические примеры



- молекулярная модель двойной спирали ДНК (Watson and Crick, 1953)

Классические примеры

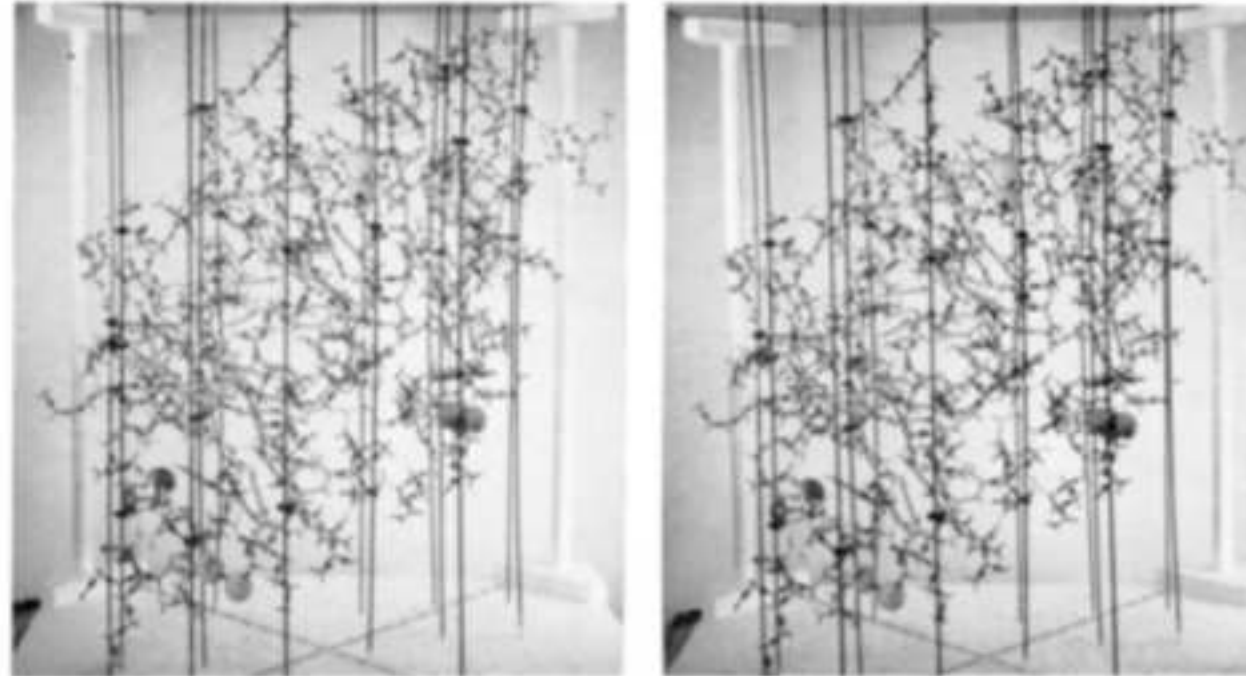
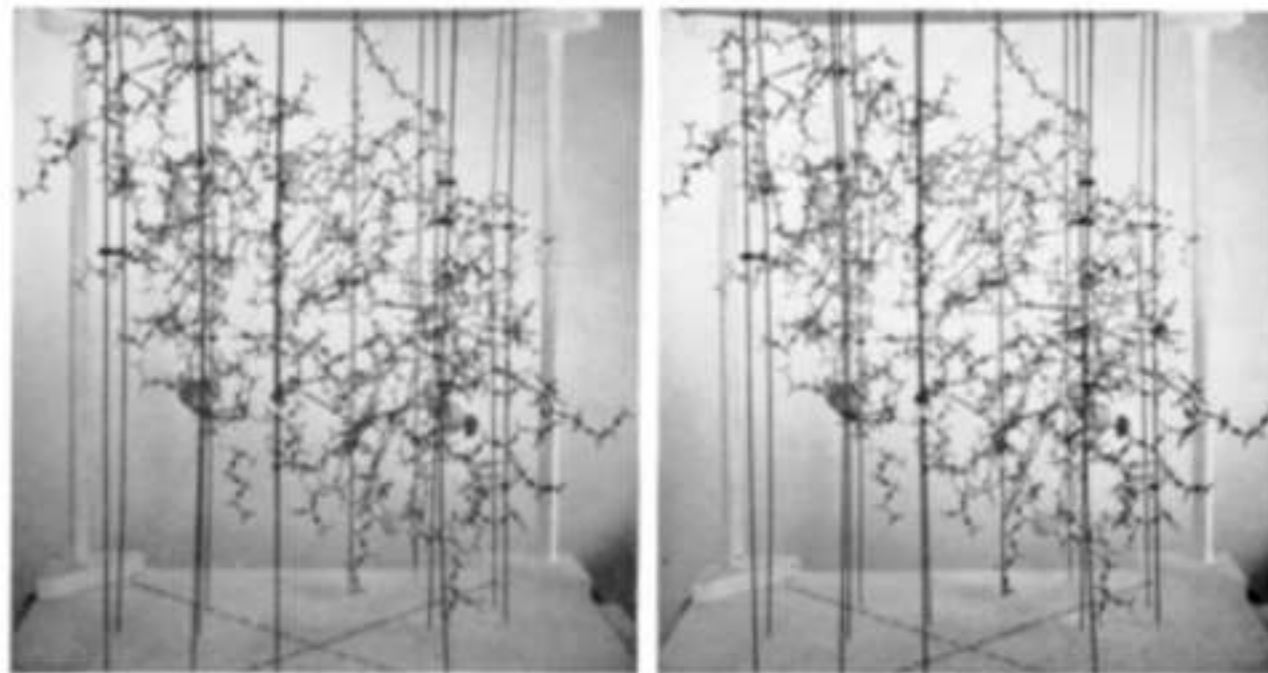


PLATE I



- первая модель гомологичного белка (alpha-lactalbumin, Browne et al., 1969).

Классические примеры

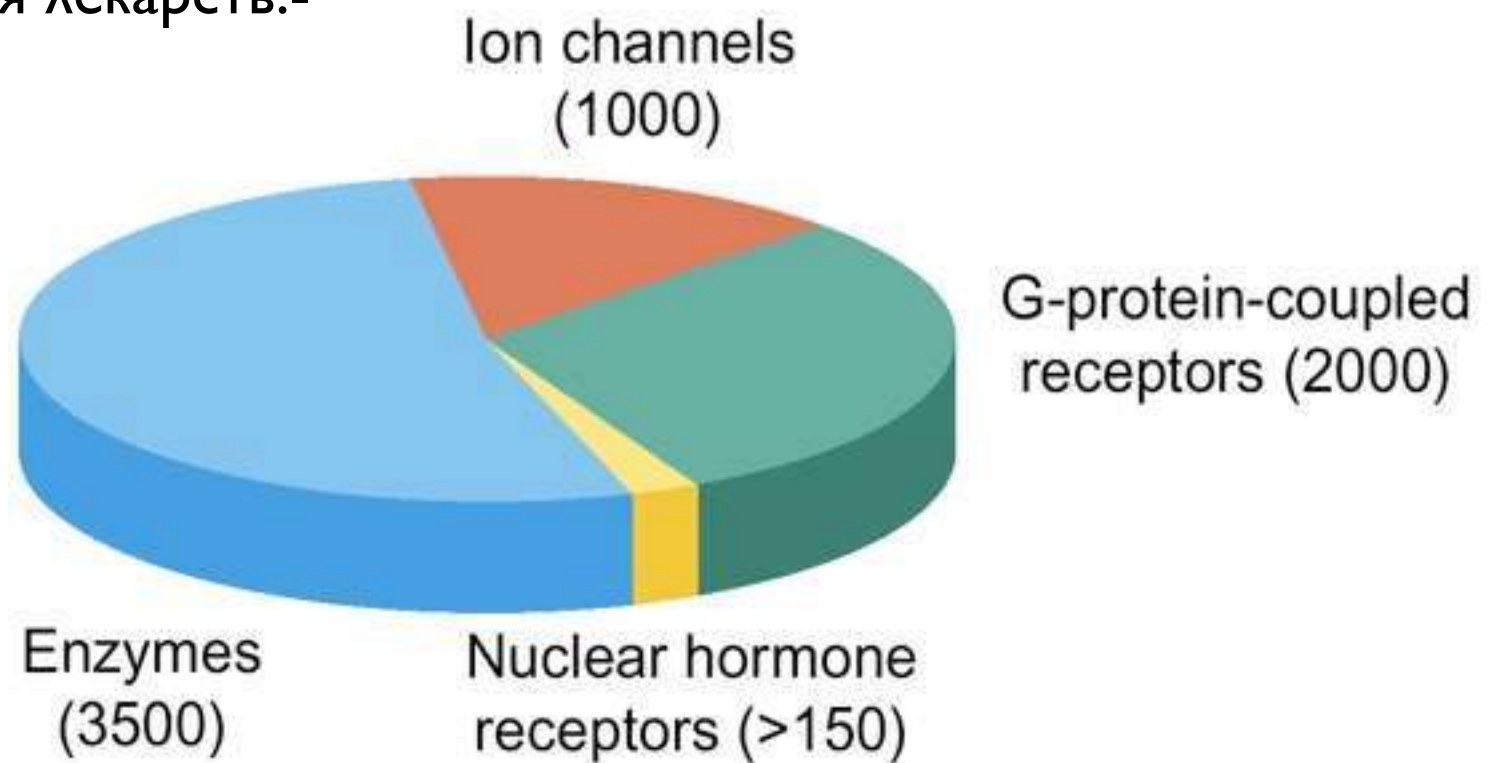


- модели полипептидных альфа-спирали и бета-листа, предложенные Линусом Полингом (Pauling et al., 1951)

Мембранные Белки – Почему?

- Ионные каналы, транспортеры, насосы, энзимы
- Сравнительно мало экспериментальной информации атомарного уровня
- Экспрессия
- Кристаллизация

- Но ключевые мишени для лекарств:-



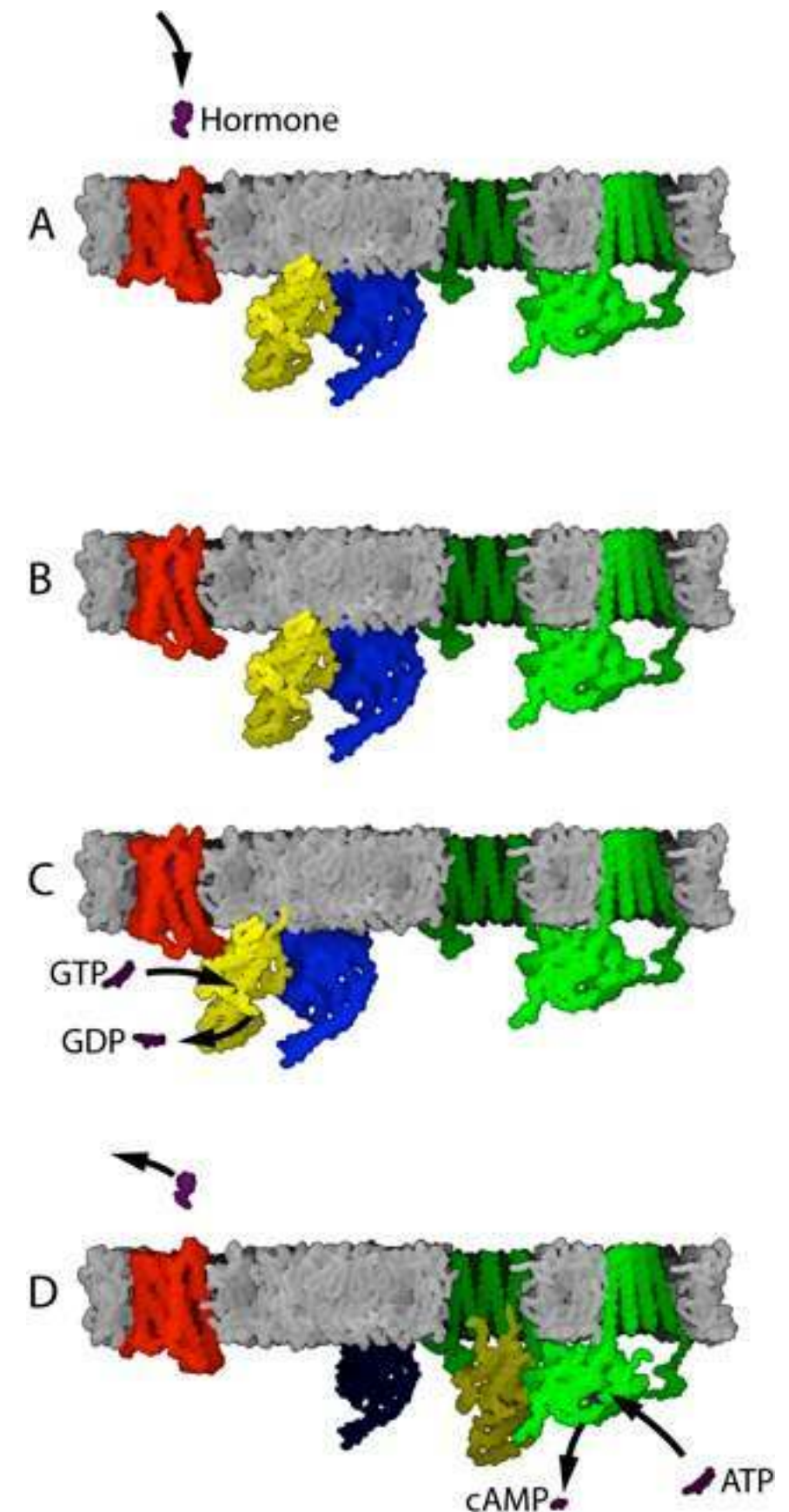
TRENDS in Pharmacological Sciences

From Terstappen & Reggiani, TIPS. 2001

GPCR

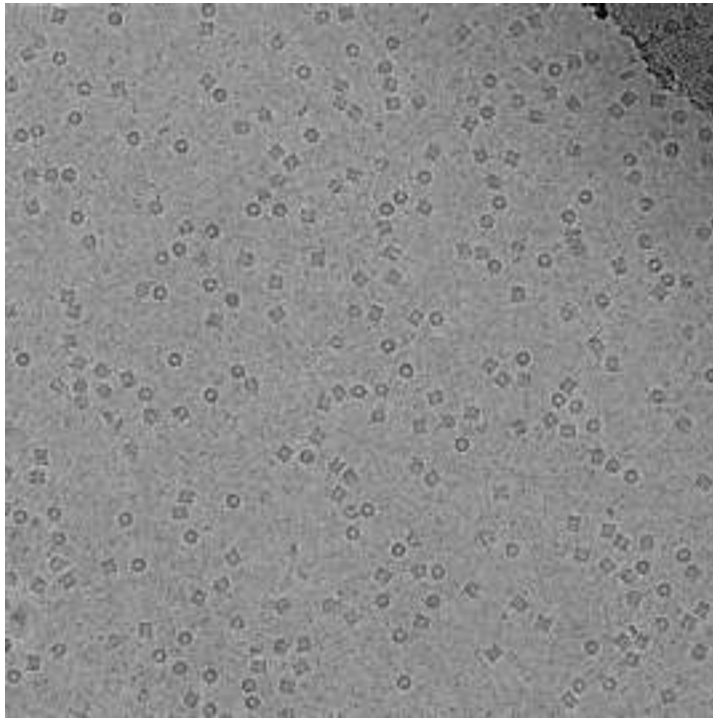
Из Википедии:

If a receptor in an active state encounters a G protein, it may activate it (blue protein in part B). Some evidence suggests that receptors and G proteins are actually pre-coupled. For example, binding of G proteins to receptors affects the receptor's affinity for ligands. Activated G proteins are bound to GTP. Further signal transduction depends on the type of G protein. The enzyme adenylate cyclase (green protein in panel C) is an example of a cellular protein that can be regulated by a G protein, in this case the G protein Gs. Adenylate cyclase activity is activated when it binds to a subunit of the activated G protein (Panel D). Activation of adenylate cyclase ends when the G protein returns to the GDP-bound state (panels E and A). Adenylate cyclases (of which 9 membrane bound and one cytosolic forms are known in humans) may also be activated or inhibited in other ways (e.g. Ca^{2+} /Calmodulin binding) and that this can modify the activity of these enzymes in an additive or synergistic fashion along with the G proteins.

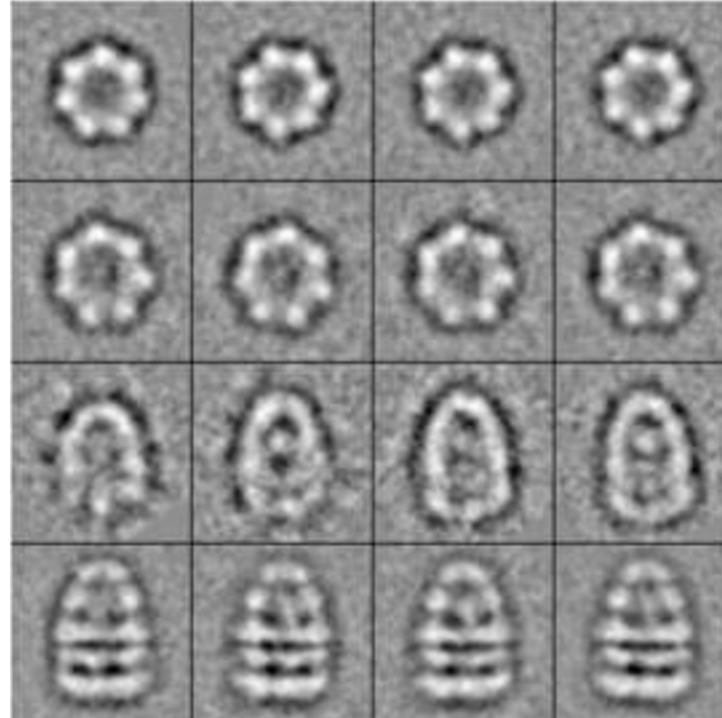


Экспериментальные Методы

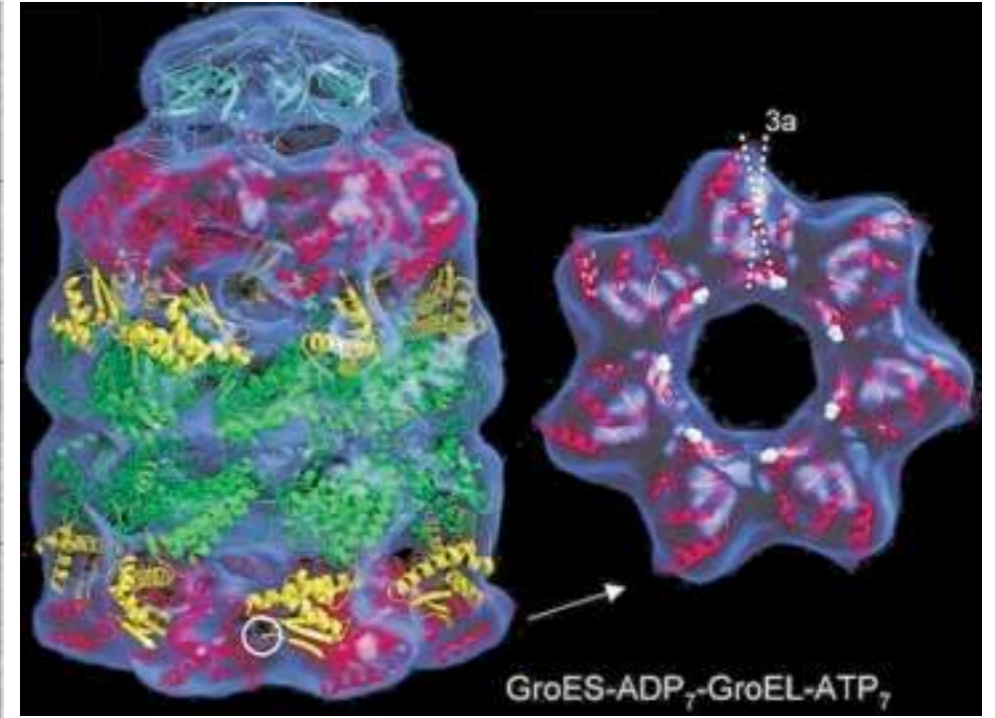
Электронная Микроскопия



EM image of GroEL



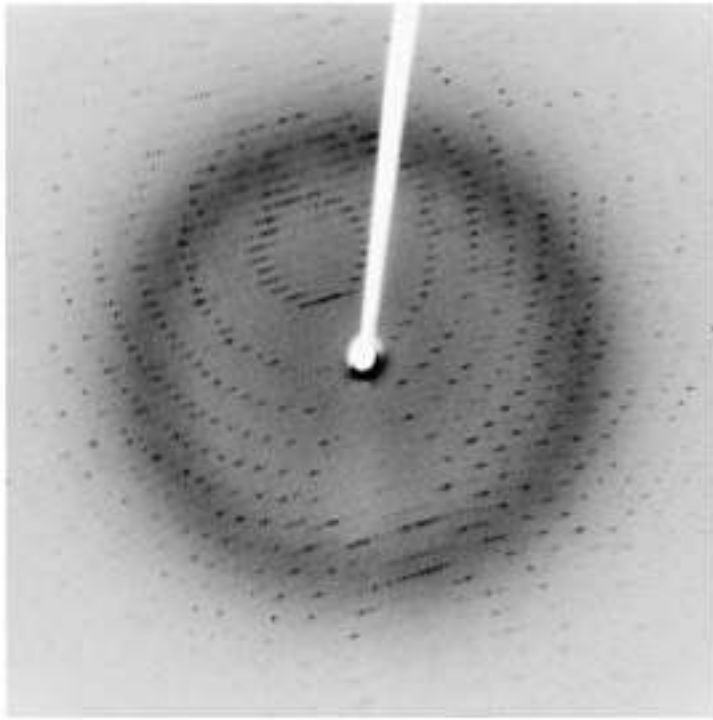
GroEL-GroES complex, Kanno et. al.,
Structure 2009, 17, 287.



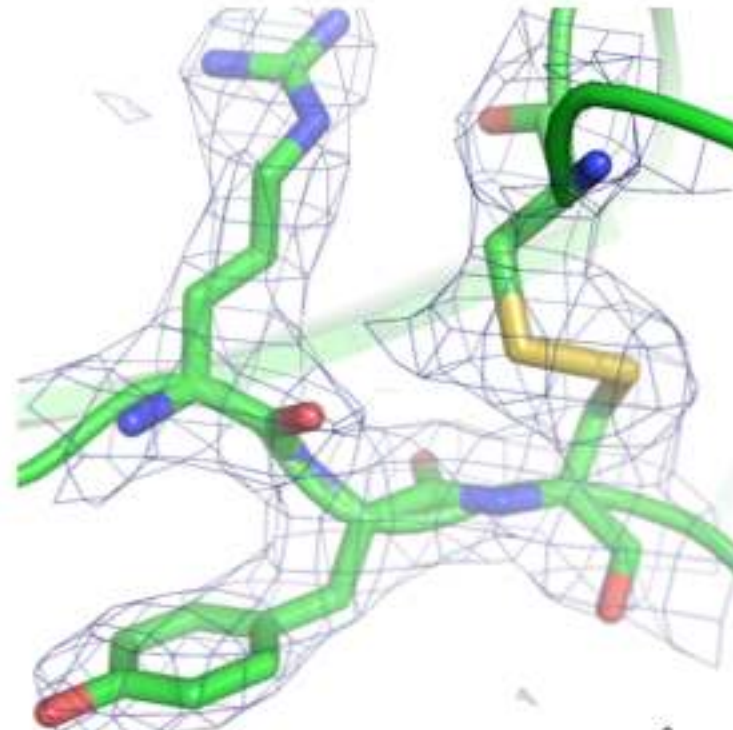
GroEL reconstruction. Ranson et. al., Cell 2001, 107,
869.

- для каждого снимка определить 5 параметров - 3 угла и 2-D положение центра масс
- определить 3-D модель по 2-D проекциям
- определить пространственную структуру белка по его 3-D модели электронной плотности

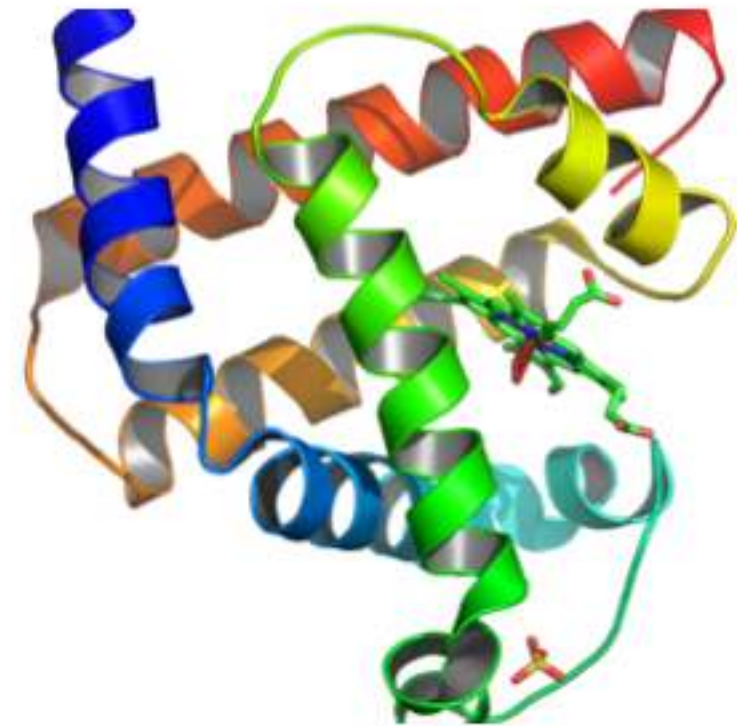
Рентгеновская Кристаллография



(a) An X-ray diffraction pattern



(b) A protein crystal structure at 2.7 Å resolution



(c) Structure of myoglobin

- электронная плотность $\rho(\mathbf{r}) = \int \frac{d\mathbf{q}}{(2\pi)^3} F(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{r}}$
- “проблема фаз” $F(\mathbf{q}) = |F(\mathbf{q})| e^{i\phi(\mathbf{q})}$
- неэмпирические фазы
- фитинг

Проблема Фаз

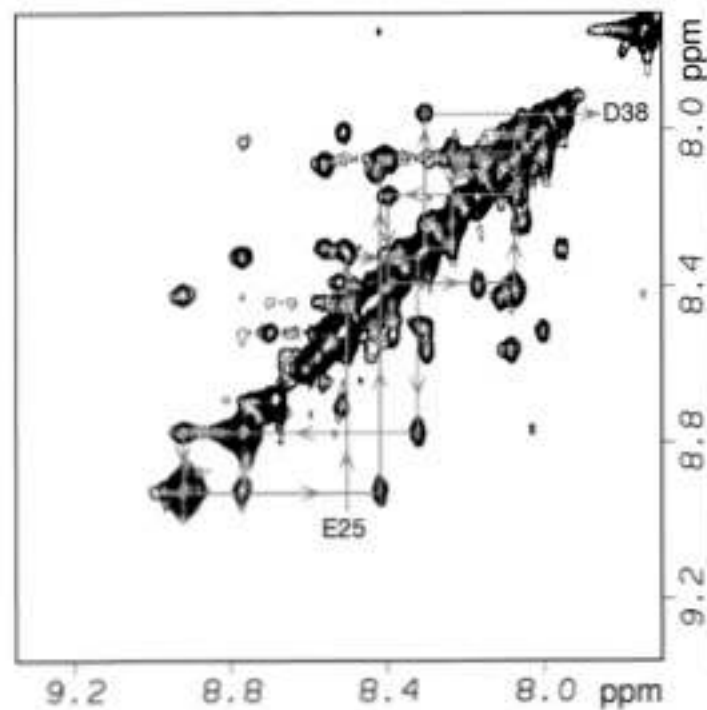
- сверху: фотографии Жерома Карле (Jerome Karle) (слева) и Херба Хауптмана (Herb Hauptman) (справа), которые получили Нобелевскую премию по физике за работы по решению фазовой проблемы для маленьких молекулярных кристаллов
 - посчитаем амплитуды и фазы по фотографии, как если бы она была картой плотностей при помощи Фурье-трансформ
- снизу: обменяли фотографии фазами



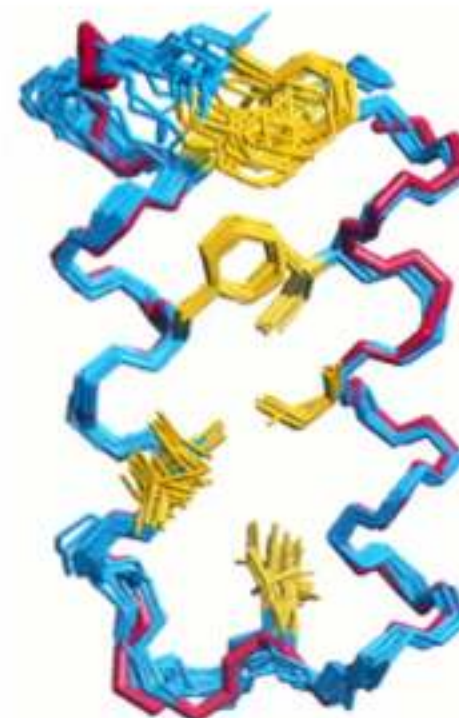
ЯМР



World's First 1 GHz NMR Spectrometer.



2D NOESY spectra, A. Starovasnik et. al. PNAS 1997, 94, 10080

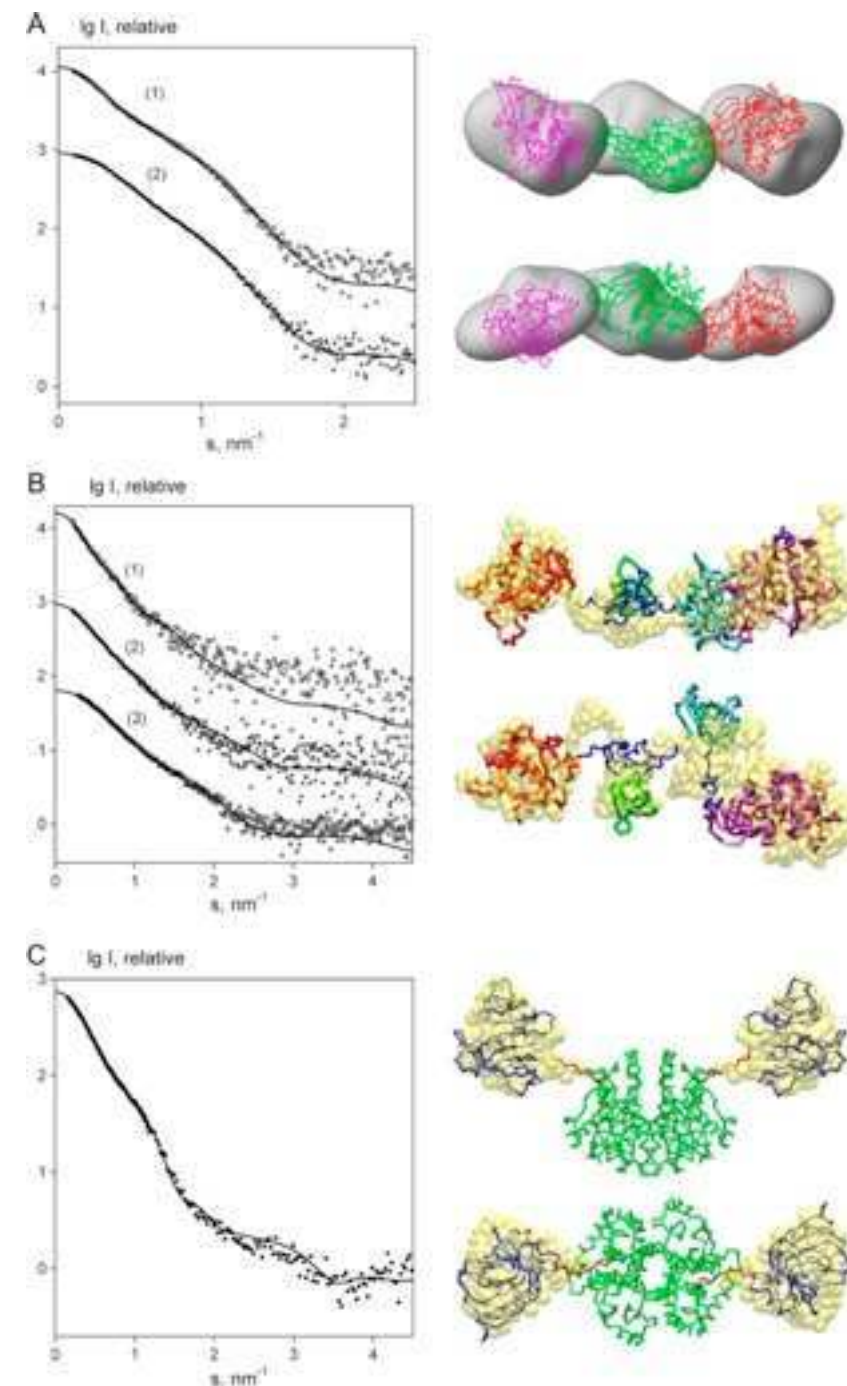


NMR structure ensembles, A. Starovasnik et. al. PNAS 1997, 94, 10080

- $NOE_{ij} \sim \frac{1}{r_{ij}^6}$
- Метод модельного отжига (Simulated Annealing)
- Метрическая геометрия

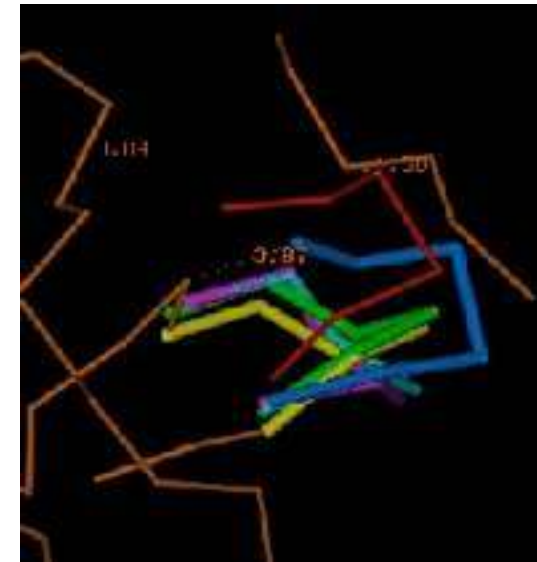
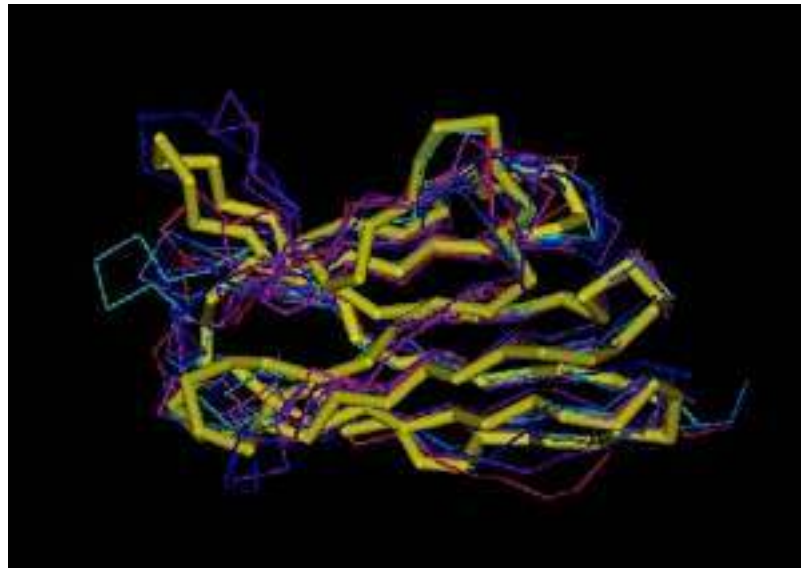
Малоугловое рассеяние (SANS)

- моделирование твердых тел по кривым рассеяния
- определение формы
- недостающие фрагменты
- олигомерные комплексы
- эластичные системы



Моделирование при помощи твердых тел,
M. Petoukhov and D. Svergun, Biophysical J
2005, 89, 1237

Моделирование по Гомологии



ACDEFGHIKLMNPQRST--FGHQWERT-----TYREWYEGHADS
ASDEYAHILRLDPQRSTVAYAYE--KSFAPPGSFKWEYEAHADS
MCDEYAHILRLMNPERS TVAGGHQWERT----GSFKEWYAAHADD

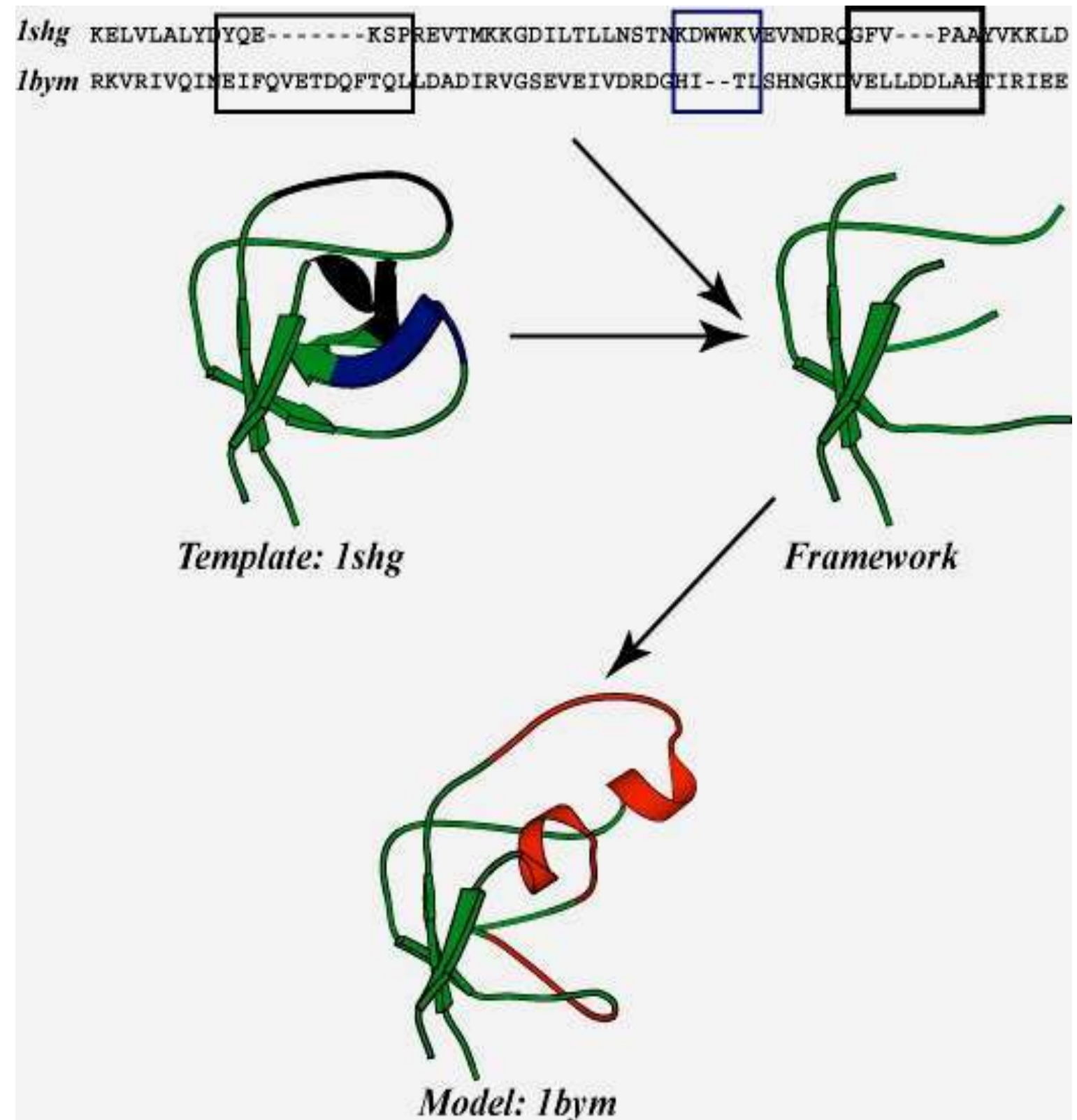
- Основано на наблюдении, что схожие пептидные последовательности формируют схожие структуры
- Известная структура используется как шаблон для моделирования неизвестной (но вероятно похожей) структуры с известной последовательностью
- Впервые применено в 1970' при помощи ранних методов компьютерного распознавания (Tom Blundell)

Моделирование по Гомологии

- Предлагает способ предсказания 3D структуры белков, для которых невозможно получить рентгеновские или ЯМР данные
- Может быть использовано для понимания функции, активности, специфичности, и т.п.
- Интересно фарма-компаниям, которые занимаются созданием лекарств на структурном уровне
- Является центральным звеном в структурной протеомике

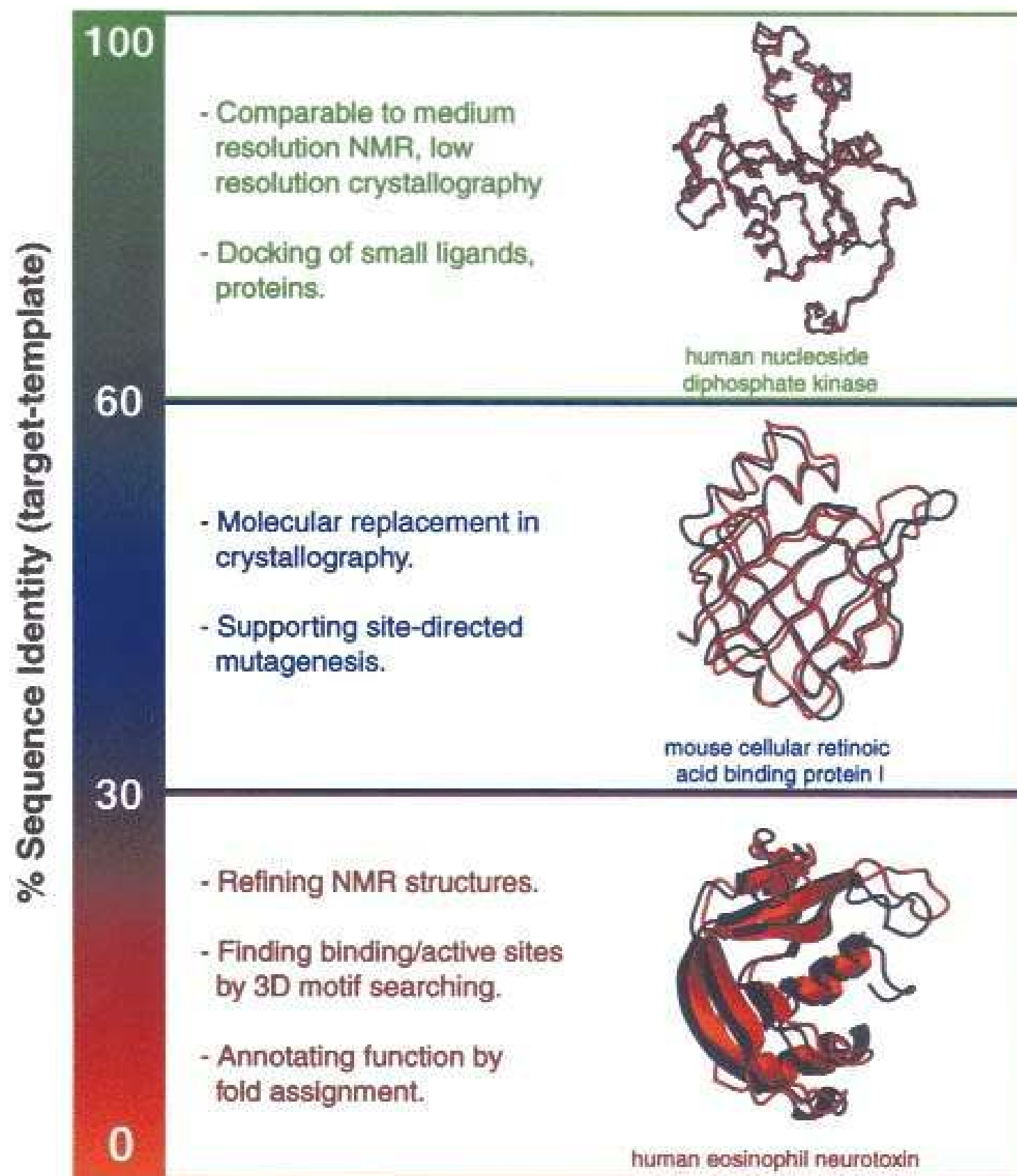
Моделирование по Гомологии

- Находим шаблон
- Выравниваем последовательность мишени с шаблоном
- Собираем модель:
 - добавляем петли
 - добавляем боковые цепи
- Улучшаем качество модели

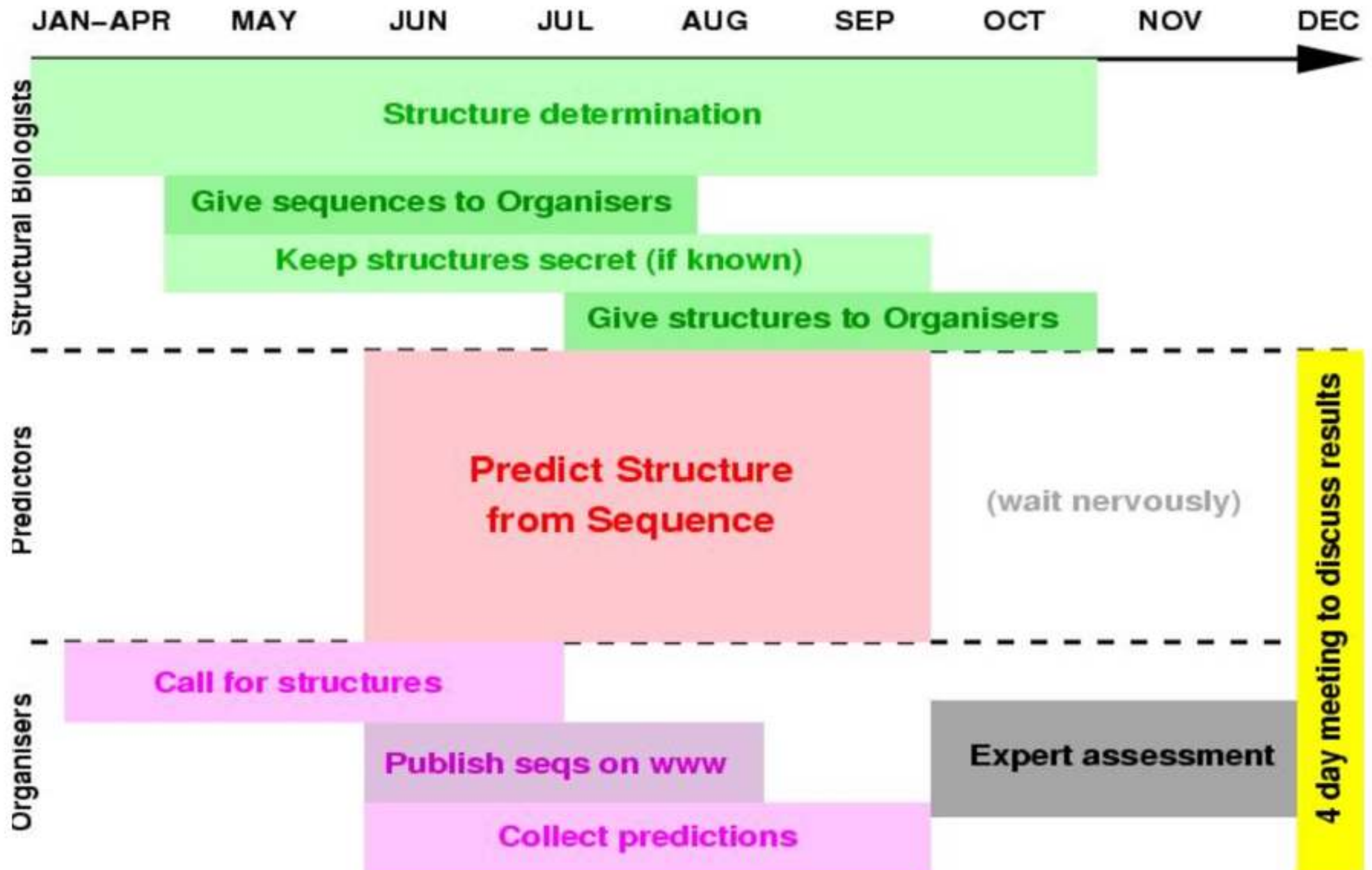


Области применения моделирования по гомологии

Области применения моделирования по гомологии. Из Sánchez et. al. Nature, 2000.

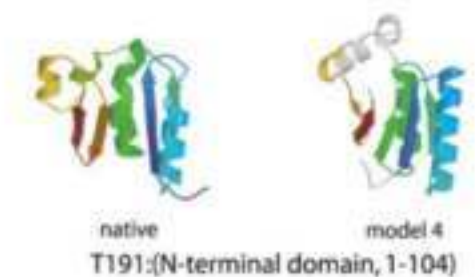
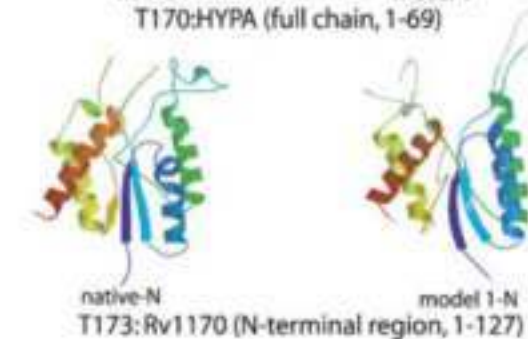
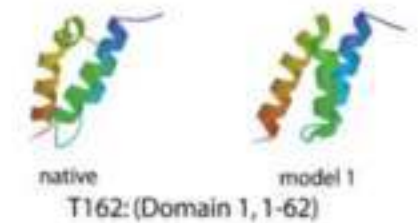
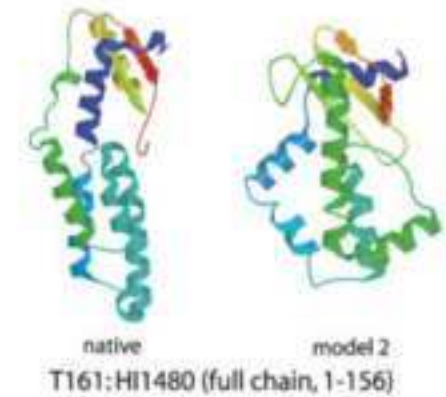
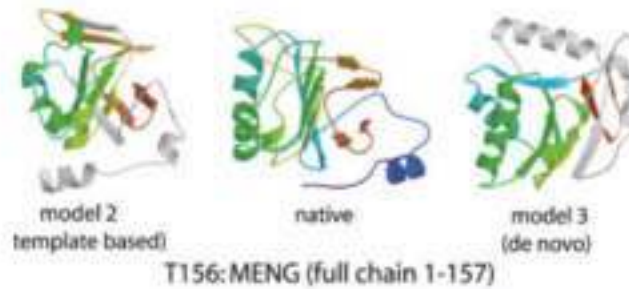
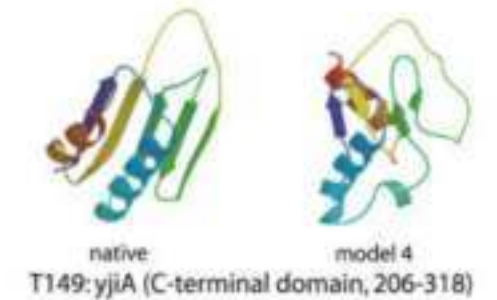
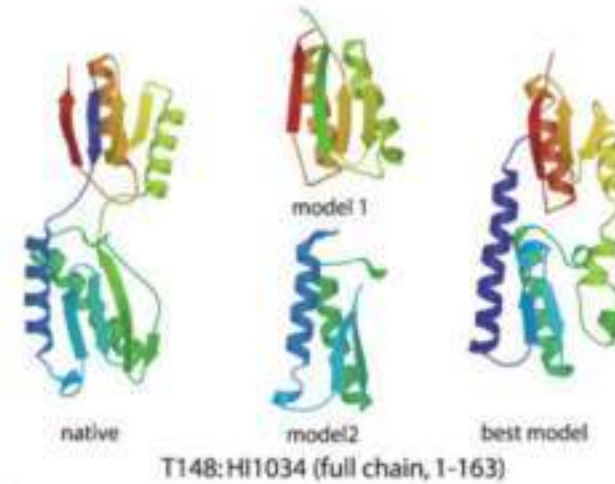
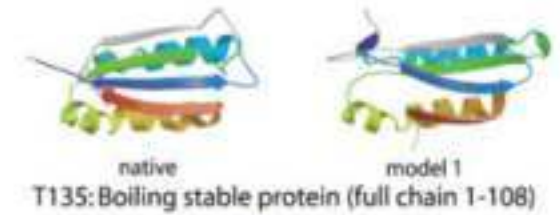
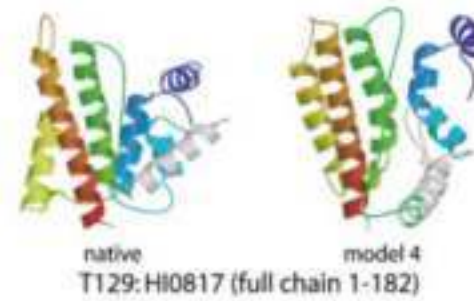


CASP



CASP

- ROSETTA -
предсказания для
CASP5



Top7: Полностью искусственный белок

- Top7 - это 93-аминокислотный α/β белок, чья топология и последовательность не встречаются в базах данных PDB или SCOP
 - Это означает, что фолдинг Top7 находится в той области конформационного пространства, которая еще не исследована решенными биологическими структурами
- кристаллическая структура Top7 имеет RMSD основной цепи 1.7Å по сравнению с предсказанной моделью
 - Для сравнения: <3 - это прекрасный результат для гомологичных моделей маленьких и средних белков с очень высоким процентом совпадения последовательности с шаблоном



компьютерная модель:

спирали = темно-синие, нити = красные

кристаллографическая структура:

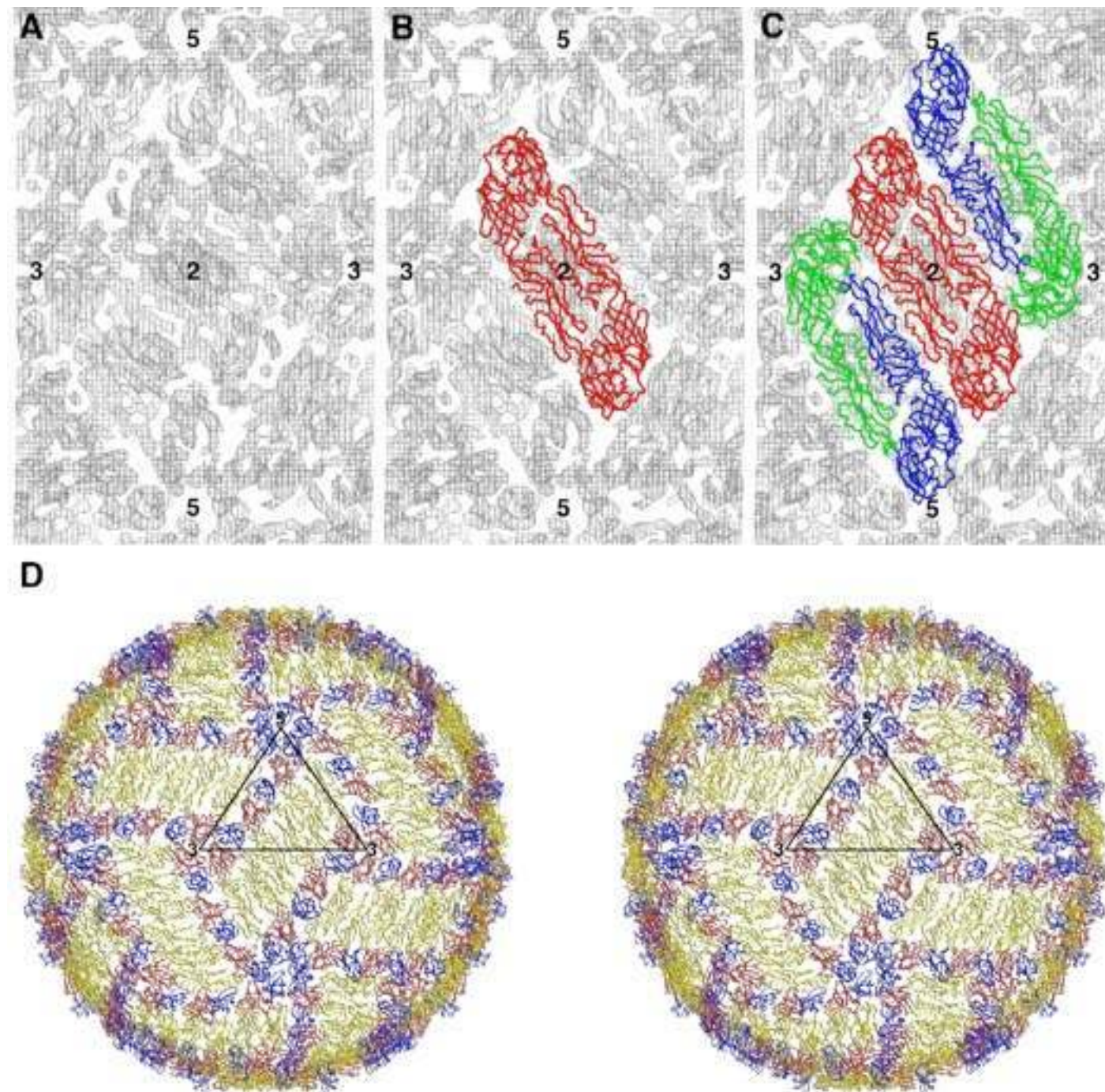
спирали = светло-синие, нити = желтые

Kuhlman B., G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker. "Design of a novel globular protein fold with atomic-level accuracy." *Science* 302, no. 5649 (Nov 21, 2003): 1364-8.

Гибридные методы определения пространственной структуры биологических объектов

- эластичный докинг моделей компонент белка в карту электронных плотностей их комплексов, определяемых при помощи электронной микроскопии (Rossmann et al., 2005 and Topf et al., 2008)
- Определение конфигураций 456 белков комплекса ядерной поры (NPC) из дрожжей (Alber et al., 2008).
- Объединение рентгеновской кристаллографии и электронной микроскопии

Гибридные методы определения пространственной структуры биологических объектов

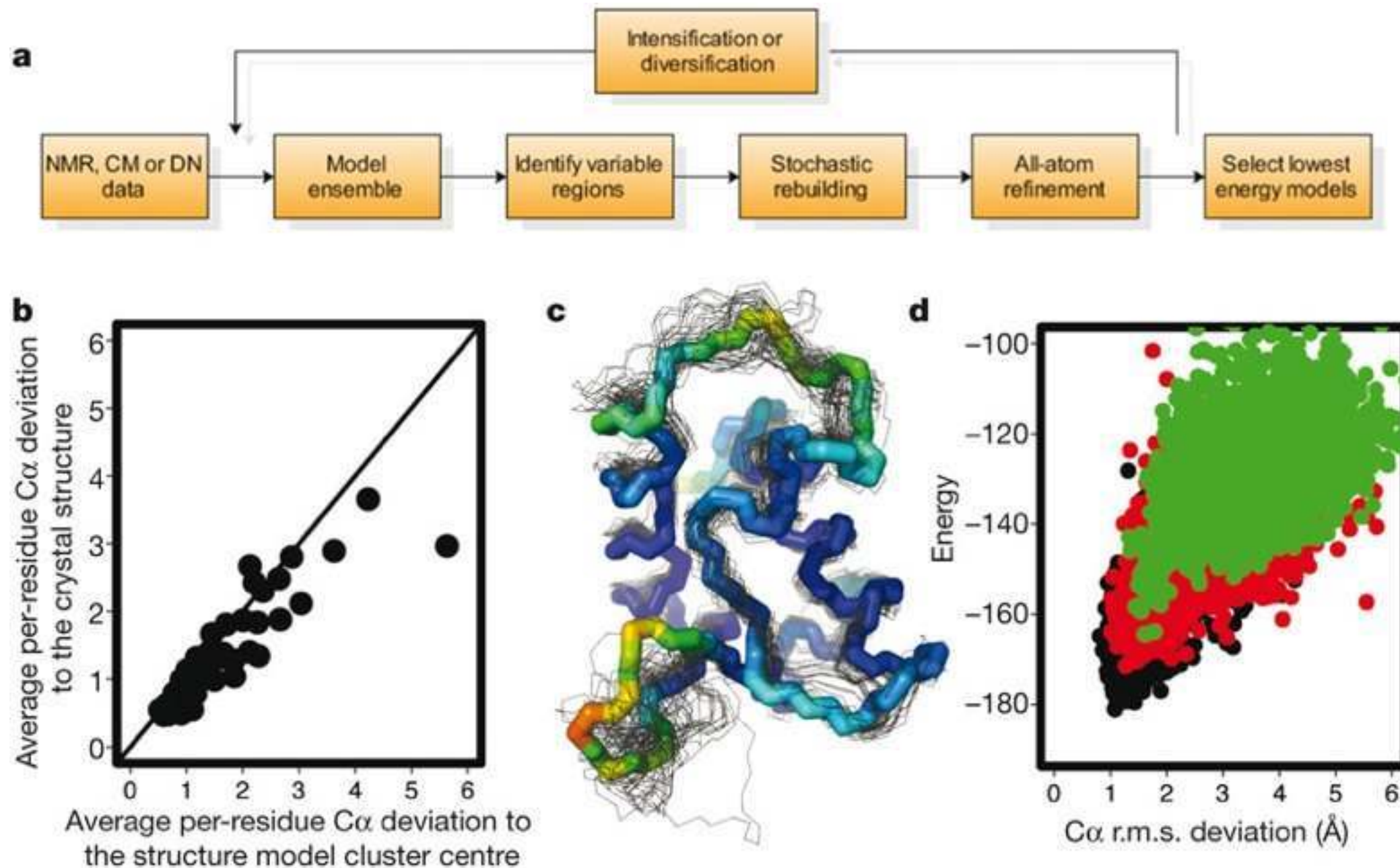


- Наилучшее наложение усредненной по икосаэдру ЕМ-электронной плотности вируса лихорадки денге разрешения 9 Å с основной цепью из кристаллографической структуры

Определение атомарной структуры

- В кристаллографии, “de novo”-предсказания структуры белка могут иногда решить фазовую проблему при помощи моделей молекулярного замещения для белков с достаточно слабой гомологией или даже без гомологии по сравнению с уже решенными структурами (Qian et al., 2007).
- Определение структуры при помощи соблюдения геометрических ограничений из ЯМР; улучшение модели при помощи атомарного потенциала может очень сильно улучшить качество ЯМР ансамбля конечных структур (Bhattacharya et al., 2008)
- Ускоренный, основанный на ЯМР структурный вывод; объединяя недостающие данные химических сдвигов и алгоритмы моделирования, можно получить структуры практически атомарного разрешения (Cavalli et al., 2007)

Определение атомарной структуры, Rosetta



a, Rosetta protocol: Schematic diagram of the rebuilding-and-refinement method applied to structures from NMR, from comparative modelling (CM) and from de novo (DN) modelling approaches.

c, Superposition of the native structure of CASP6 target T0199 with 50 low-energy all-atom refined models.

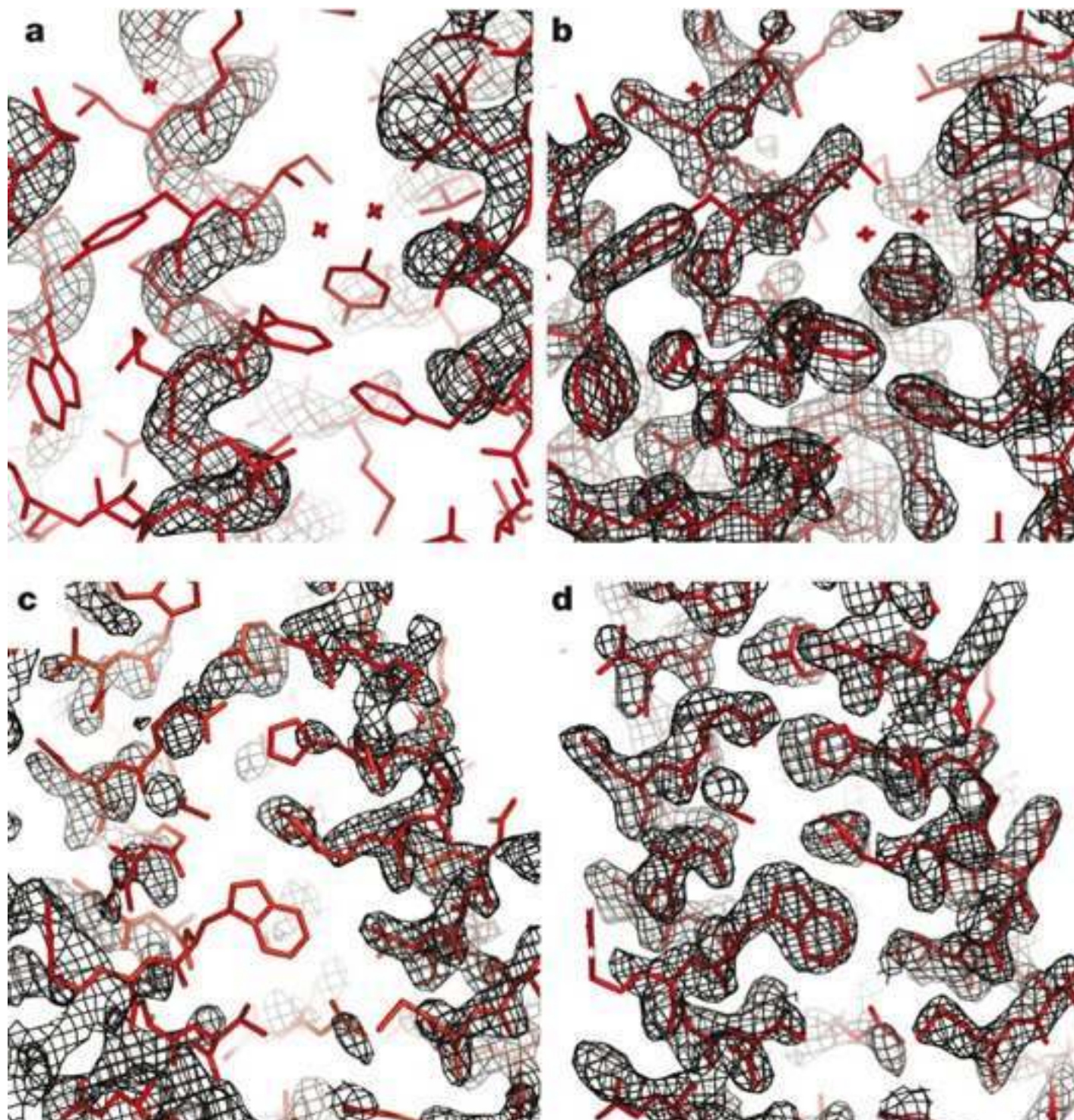
d, Iterative rebuilding and refinement yields low-energy native-like models.

Определение атомарной структуры, Rosetta

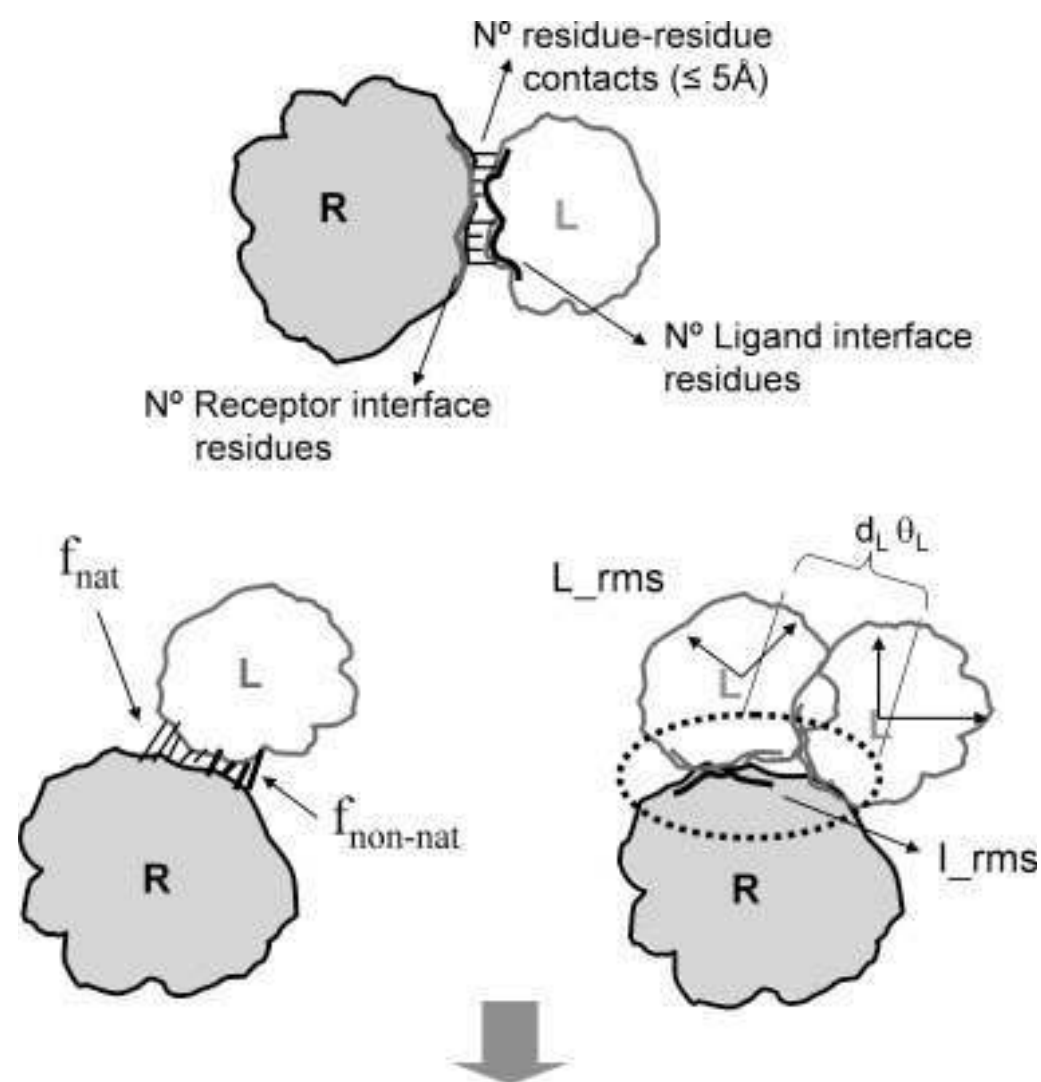
Rosetta Refinement:

(a, b) NMR structure of acyl CoA binding protein 2abd

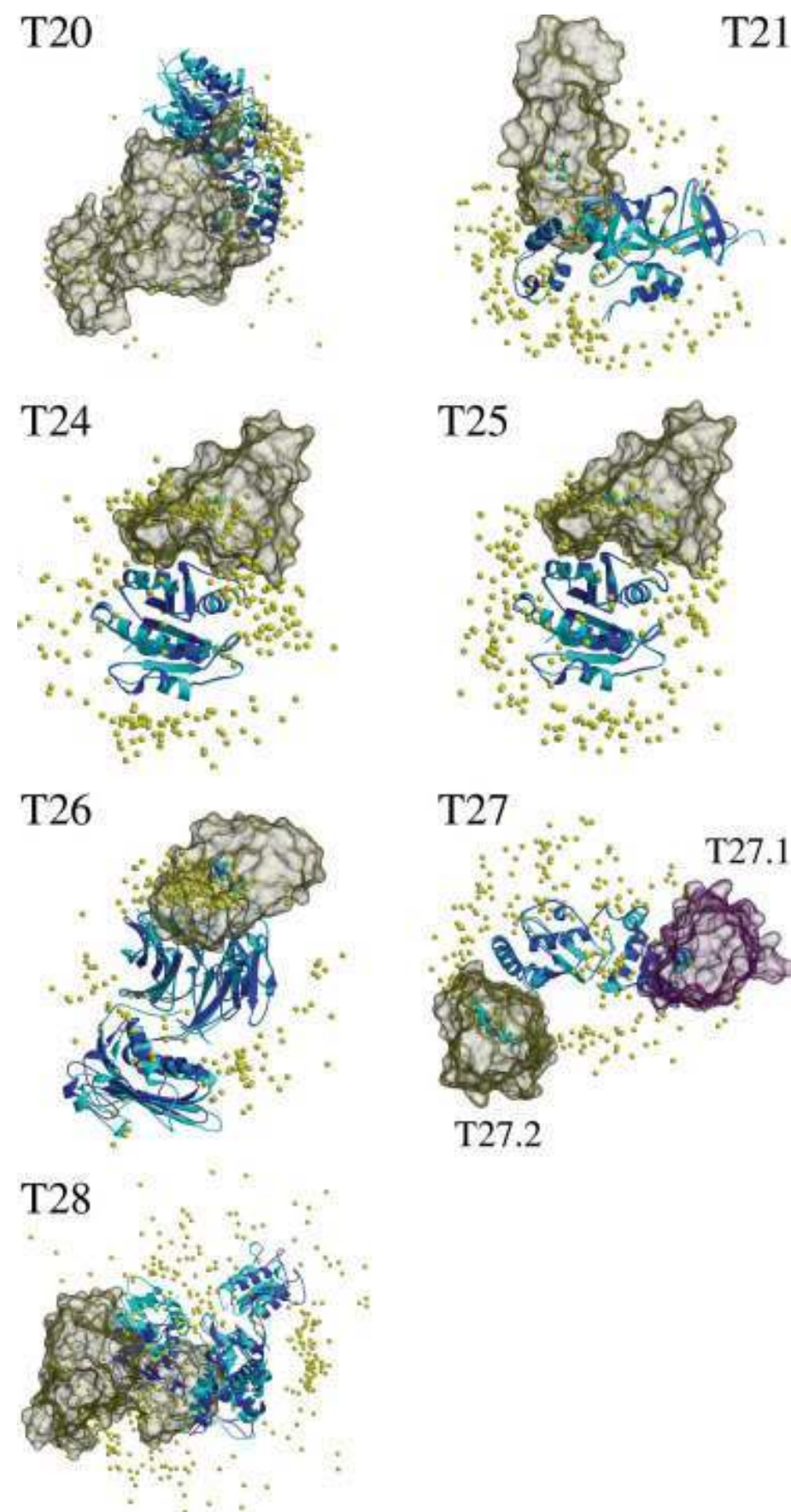
(c, d) CASP7 comparative modelling target T0385 (c and d).



Взаимодействия белок-белок и белок-лиганд



Ranking	Conditions based on Capri computed parameters
High	$f_{nat} \geq 0.5$ AND ($L_{rms} \leq 1.0$ OR $I_{rms} \leq 1.0$)
Medium	($f_{nat} \geq 0.3$ AND $f_{nat} < 0.5$) AND ($L_{rms} \leq 5.0$ OR $I_{rms} \leq 2.0$) OR $f_{nat} \geq 0.5$ AND $L_{rms} > 1.0$ AND $I_{rms} > 1.0$
Acceptable	($f_{nat} \geq 0.1$ AND $f_{nat} < 0.3$) AND ($L_{rms} \leq 10.0$ OR $I_{rms} \leq 4.0$) OR $f_{nat} \geq 0.3$ AND $L_{rms} > 5.0$ AND $I_{rms} > 2.0$
Incorrect	$f_{nat} < 0.1$ OR ($L_{rms} > 10.0$ AND $I_{rms} \leq 4.0$)



- предсказания CAPRI для мишеней T20-T28

Взаимодействия белок-белок и белок-лиганд

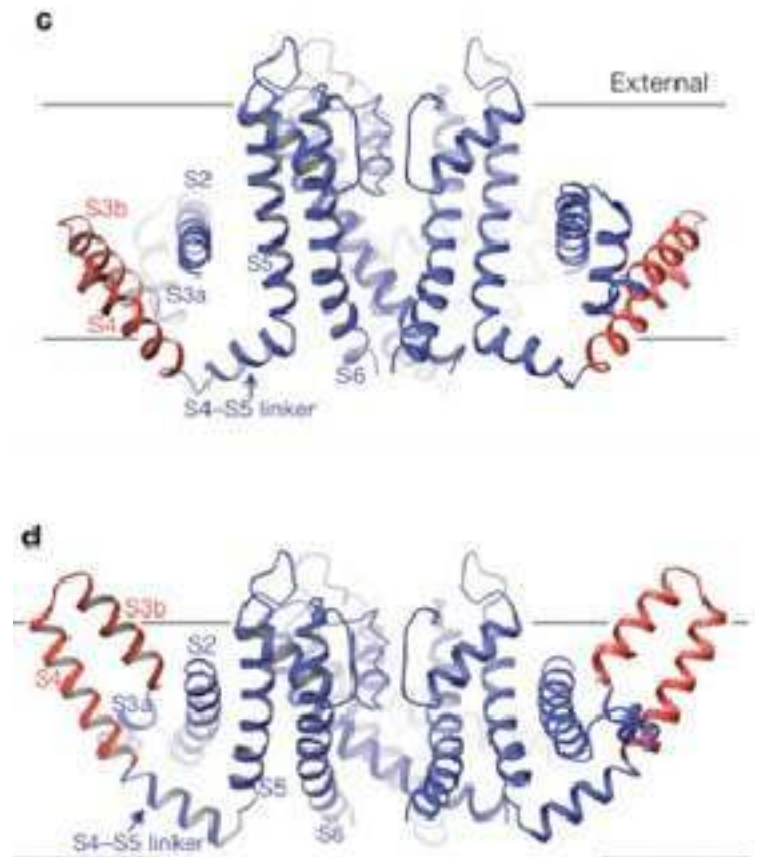
	Target T20	Target T21	Target T24	Target T25	Target T26	Target T27.1	Target T27.2	Target T28
Predictor groups	29	37	37	37	37	38	38	38
Evaluated predictions	266	337	335	336	351	343	348	357
High accuracy(***)	0	0	0	1	0	0	0	0
Medium Accuracy(**)	0	4	0	13	22	0	2	0
Acceptable(*)	3	7	4	20	20	0	55	0
Incorrect	246	307	298	289	291	327	270	322
Predictions with clashes	17	19	33	13	18	16	21	35
Average no. of clashes (SD)	35.30 (39.15)	29.32 (38.17)	22 (29)	19 (24)	25 (29)	21 (26)	20 (26)	41 (51)

- предсказания CAPRI для мишеней T20-T28 (из Lensink et. al. 2007)

Монте-Карло и Молекулярная Динамика

Измеряем среднее

- Калиевый канал
 - регулируется напряжением (открывается — закрывается)
 - Содержит S3b-S4 спирали, чувствительные к напряжению
- Как мы можем его изучить моделированием?
 - Можем помереть *среднюю* силу в открытой структуре
 - Что такое *средняя* сила?



Измеряем среднее

- В эксперименте мы ждем

$$\langle F \rangle = \int_0^\infty F[r^N(t), p^N(t)] dt$$

- Пока система не посетит достаточно состояний фазового пространства (скорости обычно не важны)

Измеряем среднее

- Или вместо ожидания

$$\langle F \rangle = \int_0^\infty F[r^N(t), p^N(t)] dt$$

- Мы можем сами посетить как можно больше состояний

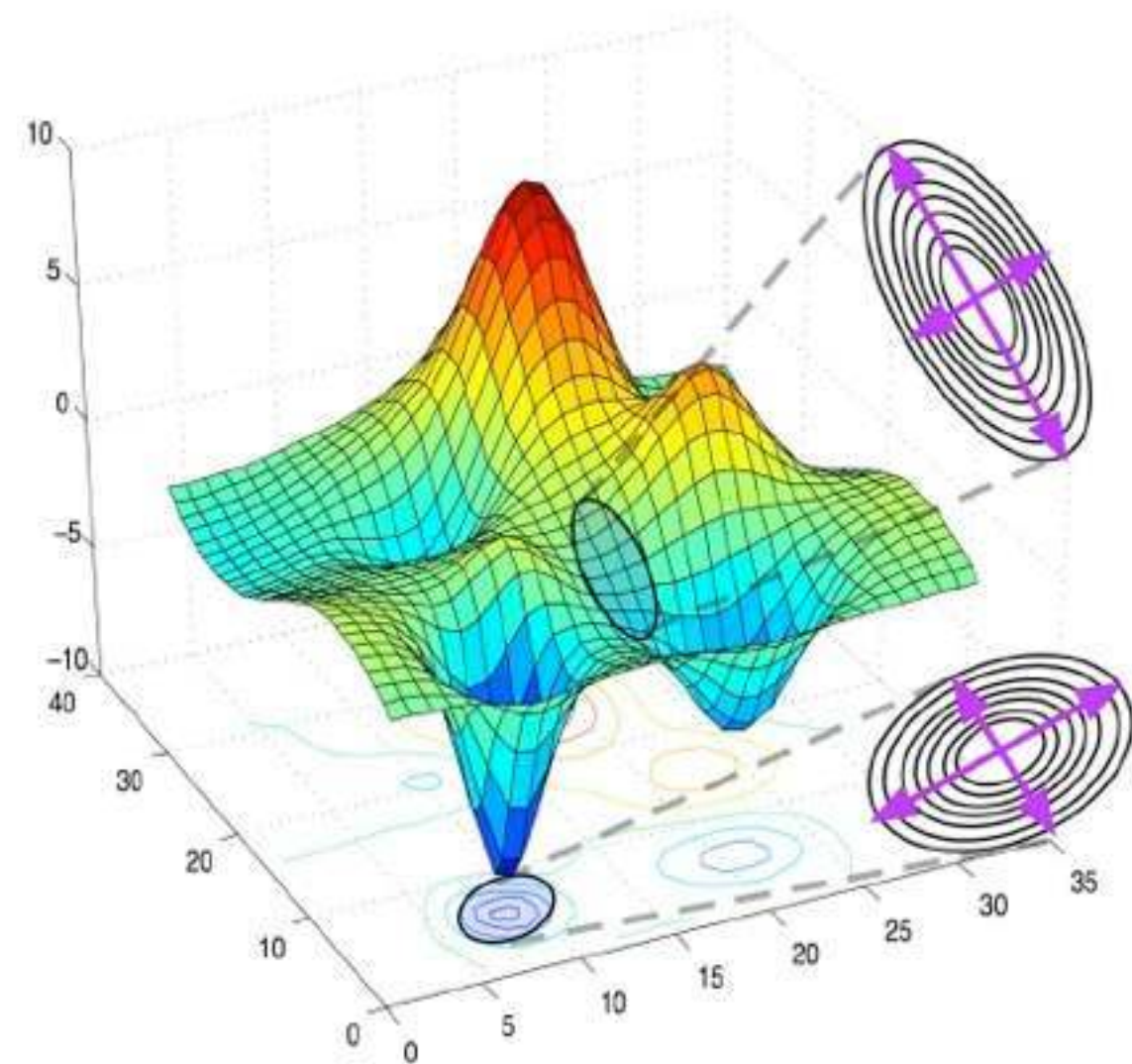
$$\langle F \rangle = \int dr^N dp^N F(r^N, p^N) p(r^N, p^N)$$

$$\langle F \rangle = \int dr^N dp^N F(r^N, p^N) e^{-\beta \mathcal{H}(r^N, p^N)}$$

- Вероятность каждой точки фазового пространства
 $\exp[-\beta H(r^N, p^N)]$

Ландшафт энергии

- Пространство размерности $3N$
- Нативная структура — та, что соответствует минимуму свободной энергии
- В идеале нам нужно сделать полный перебор всех конформаций
- В действительности мы можем это сделать только с наиболее заселенными конформациями

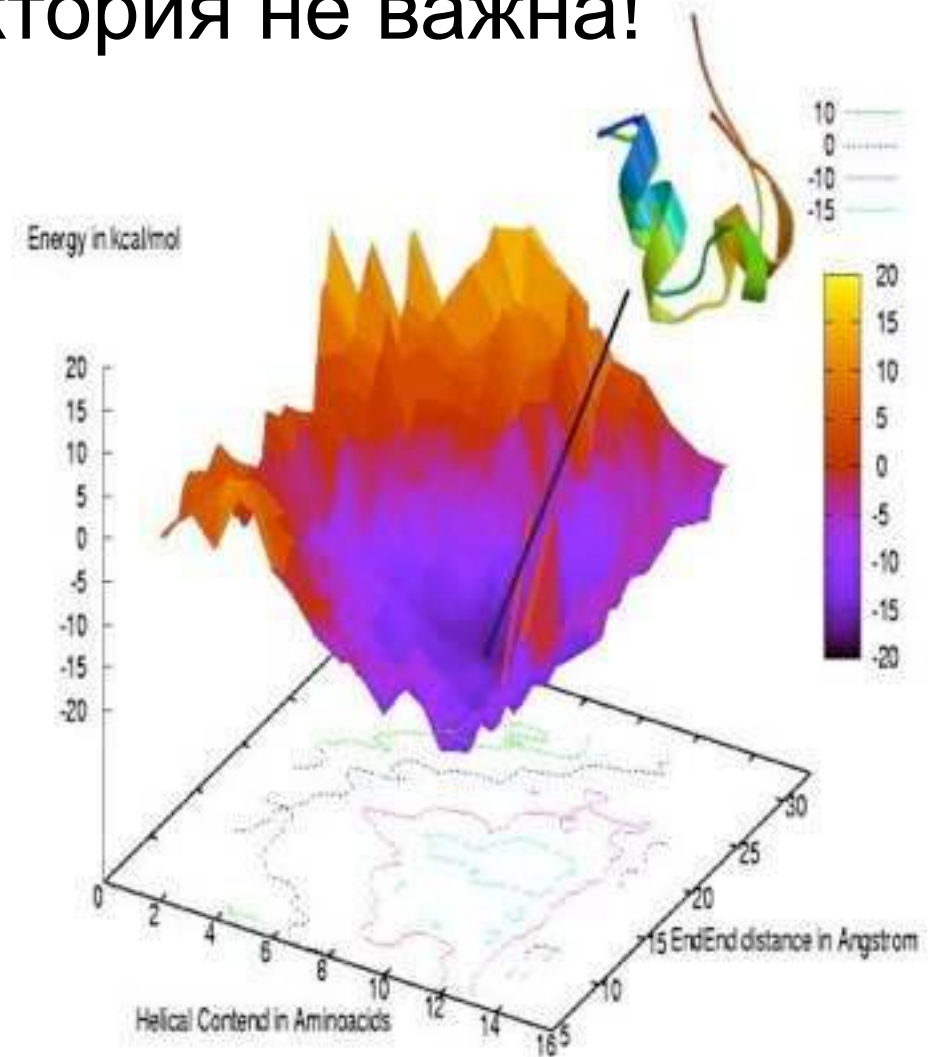


Хаотичные системы

- Иногда даже невозможно получить что-то полезное после моделирования
- Например, проследим за траекторией частицы в воздухе
- Очень маленькое отклонение в начальной конфигурации приведет к большим изменениям
- Неустойчивость Ляпунова

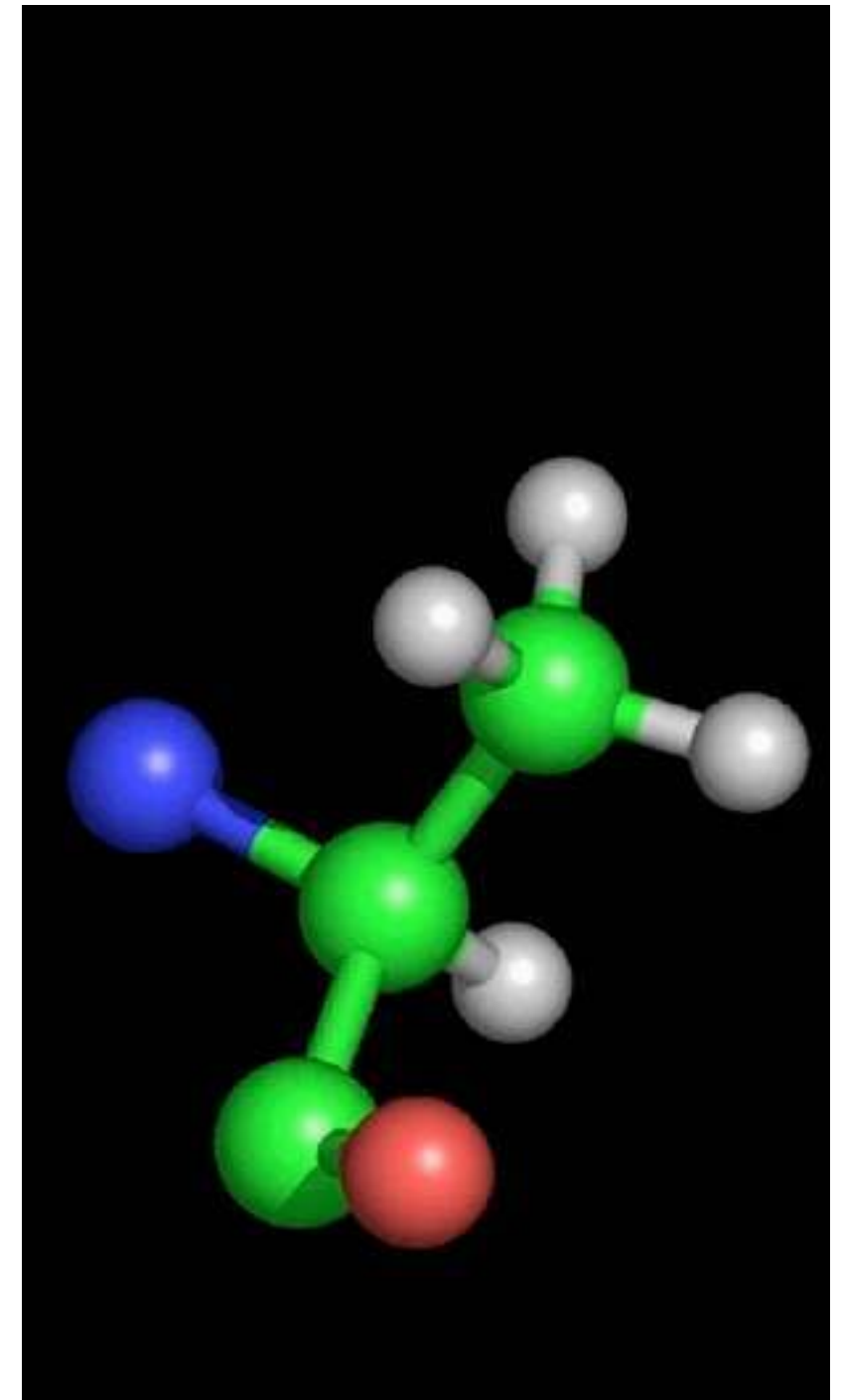
Средние свойства

- Поэтому мы сконцентрируем свои усилия на средних свойствах, усредненных по ансамблю — термодинамика!
- Исследуем типичные конфигурации фазового пространства, конкретная траектория не важна!



Моделирование

- Монте-Карло
 - Нет времени
 - Не требует вычисления сил
 - Требуется очень хорошо задать следующую конфигурацию
- Молекулярная Динамика
 - В уравнениях есть время
 - Стохастические процессы
 - Добавляем шум



Монте-Карло

Монте-Карло

Памятка: все величины, которые нам могут понадобиться - среднее в фазовом пространстве

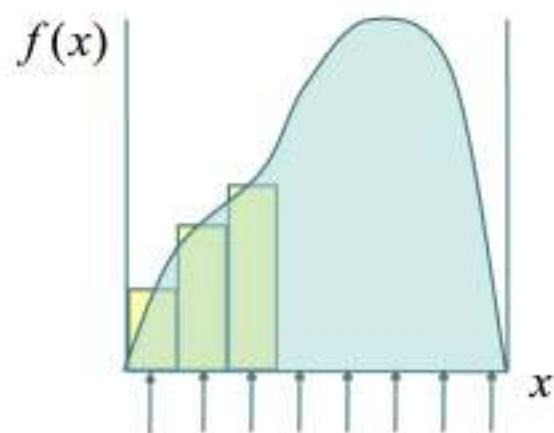
$$\underbrace{\langle F \rangle}_{\text{average}} = \int \underbrace{dr^N dp^N}_{\text{phase space}} \underbrace{F(r^N, p^N)}_{\text{microscopic value}} \underbrace{e^{-\beta \mathcal{H}(r^N, p^N)}}_{\text{probability}} / Z$$

- Интеграл по импульсу может быть посчитан аналитически
- Сложная задача - это вычисление среднего для $F(r^N)$
- Что мы можем? Численное интегрирование?

Монте-Карло

- Численное интегрирование
- Правило треугольника, правило Симпсона

- Стохастический подход
- Те же формулы для квадратур, но другой выбор точек

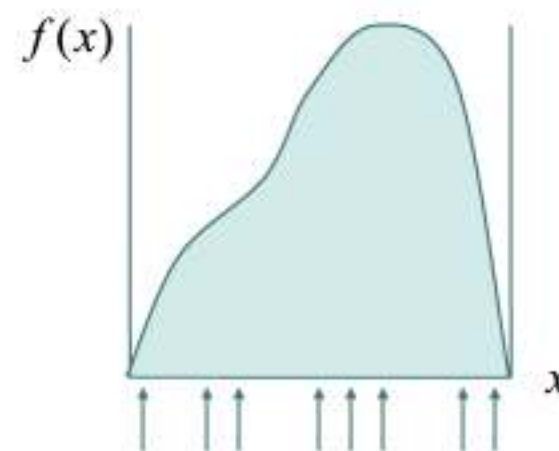


Равномерно распределенные точки

$$I = \int_a^b f(x) dx$$

Численное интегрирование

$$I \approx \Delta x \sum_{i=1}^n f(x_i) = \frac{b-a}{n} \sum_{i=1}^n f(x_i)$$



Монте-Карло

- Численное интегрирование:
 - выбираем m точек по каждой координате
 - у нас размерность D и N степеней свободы
 - общее количество точек будет m^{DN}
 - например, 100 частиц в 3 измерениях с $m=5$ дадут 10^{210} точек!
 - невыполнимо!!!
- Сравнение ошибок:
 - Стандартное интегрирование $E \propto \Delta x^2 / n^{2/d}$
 - Монте-Карло (МС) интегрирование $E \propto 1/n^{1/2}$
 - ошибка МС спадает намного медленнее с увеличением n
 - Для одномерного интеграла, МС не дает никаких преимуществ
 - Этот вывод меняется при увеличении размерности системы
 - МС “выигрывает” при примерно $D = 4$

Вопрос

Попытайтесь оценить $\int_0^a \frac{\sin x}{x} dx$

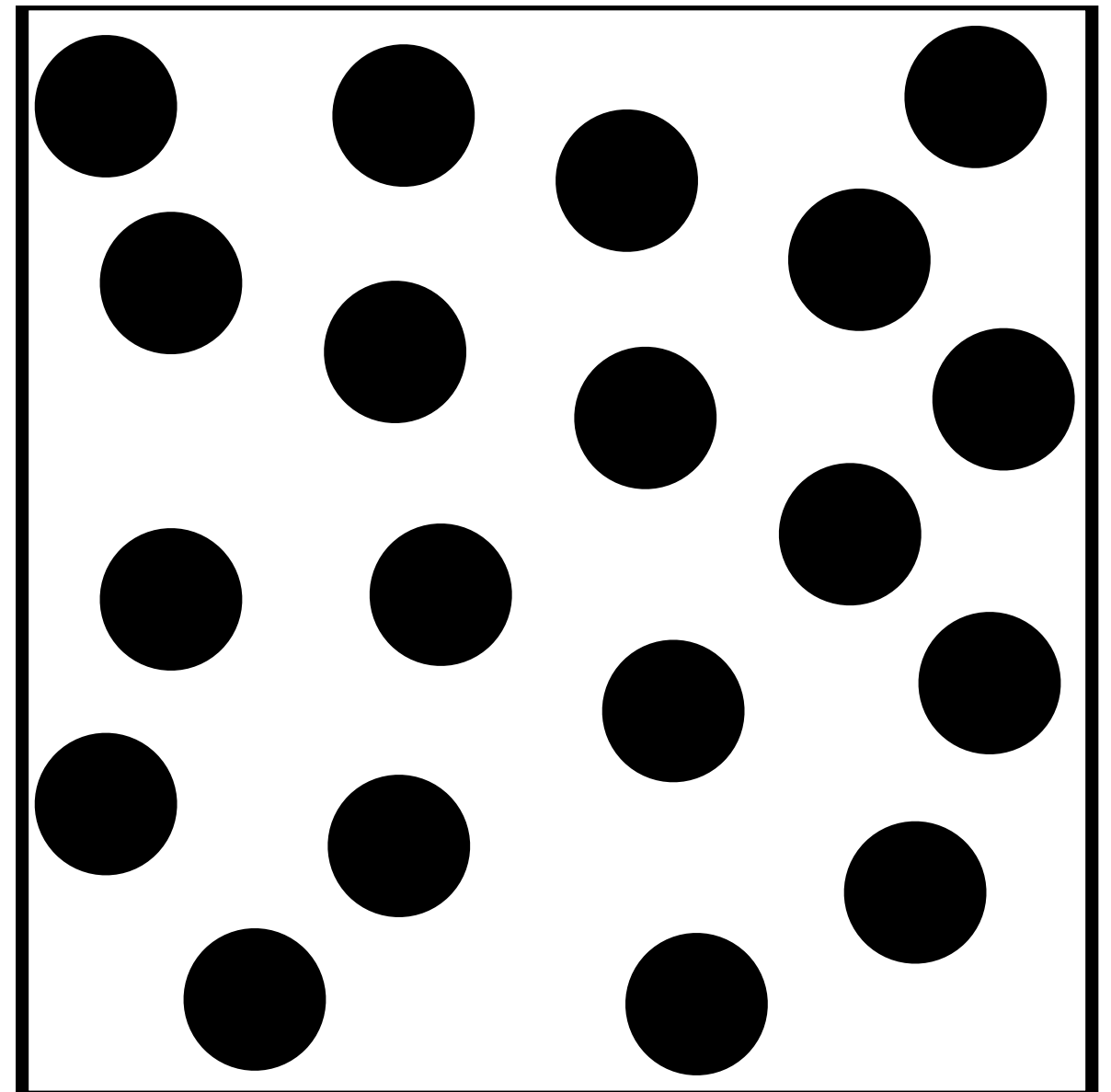
$\int_0^\infty \frac{\sin x}{x} dx$ -берется аналитически

Монте-Карло

$$\underbrace{\langle F \rangle}_{\text{average}} = \int \underbrace{dr^N dp^N}_{\text{phase space}} \underbrace{F(r^N, p^N)}_{\text{microscopic value}} \underbrace{e^{-\beta \mathcal{H}(r^N, p^N)}}_{\text{probability}} / Z$$

Пример жестких сфер:

- 100 жестких сфер
- ненулевой Больцмановский фактор только для 1 из 10^{260} конфигураций



Выборка по значимости

$$I = \int_a^b dx f(x).$$

$$I = (b - a) \langle f(x) \rangle$$

Простейший Монте-Карло $\langle f(x) \rangle$

среднее считается определением значения функции в большом количестве случайно распределенных точек

Как должны быть распределены точки?

$$I = \int_0^1 dx w(x) \frac{f(x)}{w(x)}$$

$w(x)$ - это производная $u(x)$, $u(0) = 0$ and $u(1) = 1$

$$I = \int_0^1 du \frac{f[x(u)]}{w[x(u)]}, \quad I \approx \frac{1}{L} \sum_{i=1}^L \frac{f[x(u_i)]}{w[x(u_i)]}$$

Выборка по значимости

Дисперсия:

$$\sigma_1^2 = \frac{1}{L^2} \sum_{i=1}^L \sum_{j=1}^L \left\langle \left(\frac{f[x(u_i)]}{w[x(u_i)]} - \langle f/w \rangle \right) \left(\frac{f[x(u_j)]}{w[x(u_j)]} - \langle f/w \rangle \right) \right\rangle$$

$$\sigma_1^2 = \frac{1}{L^2} \sum_{i=1}^L \left\langle \left(\frac{f[x(u_i)]}{w[x(u_i)]} - \langle f/w \rangle \right)^2 \right\rangle = \frac{1}{L} \left[\langle (f/w)^2 \rangle - \langle f/w \rangle^2 \right]$$

- Величина дисперсии может быть уменьшена выбором соответствующего $w(x)$
- Лучше всего: $f(x)/w(x)$ постоянно - нулевая дисперсия
- Хуже всего: постоянное $w(x)$
- Для примера жестких сфер относительная ошибка $\frac{1}{L} 10^{260}$

Метод Метрополиса

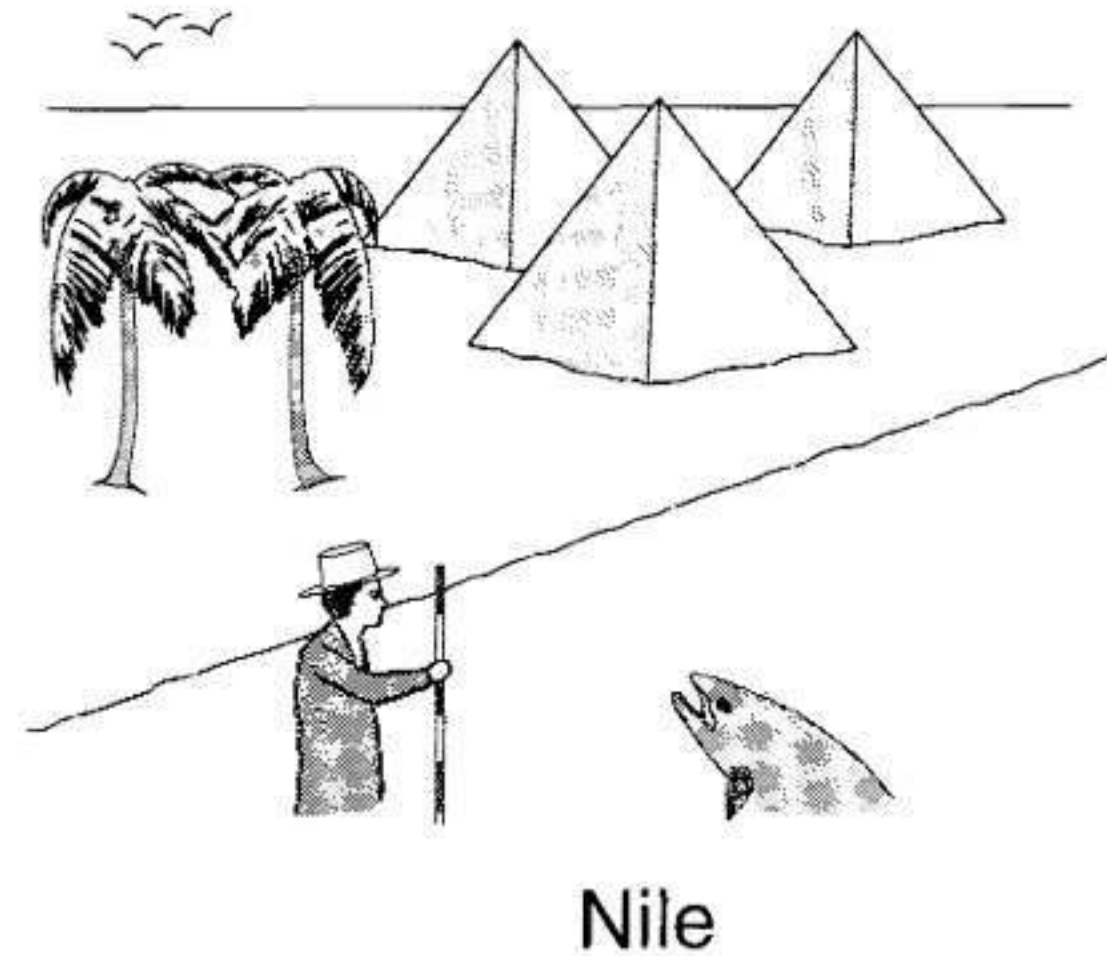
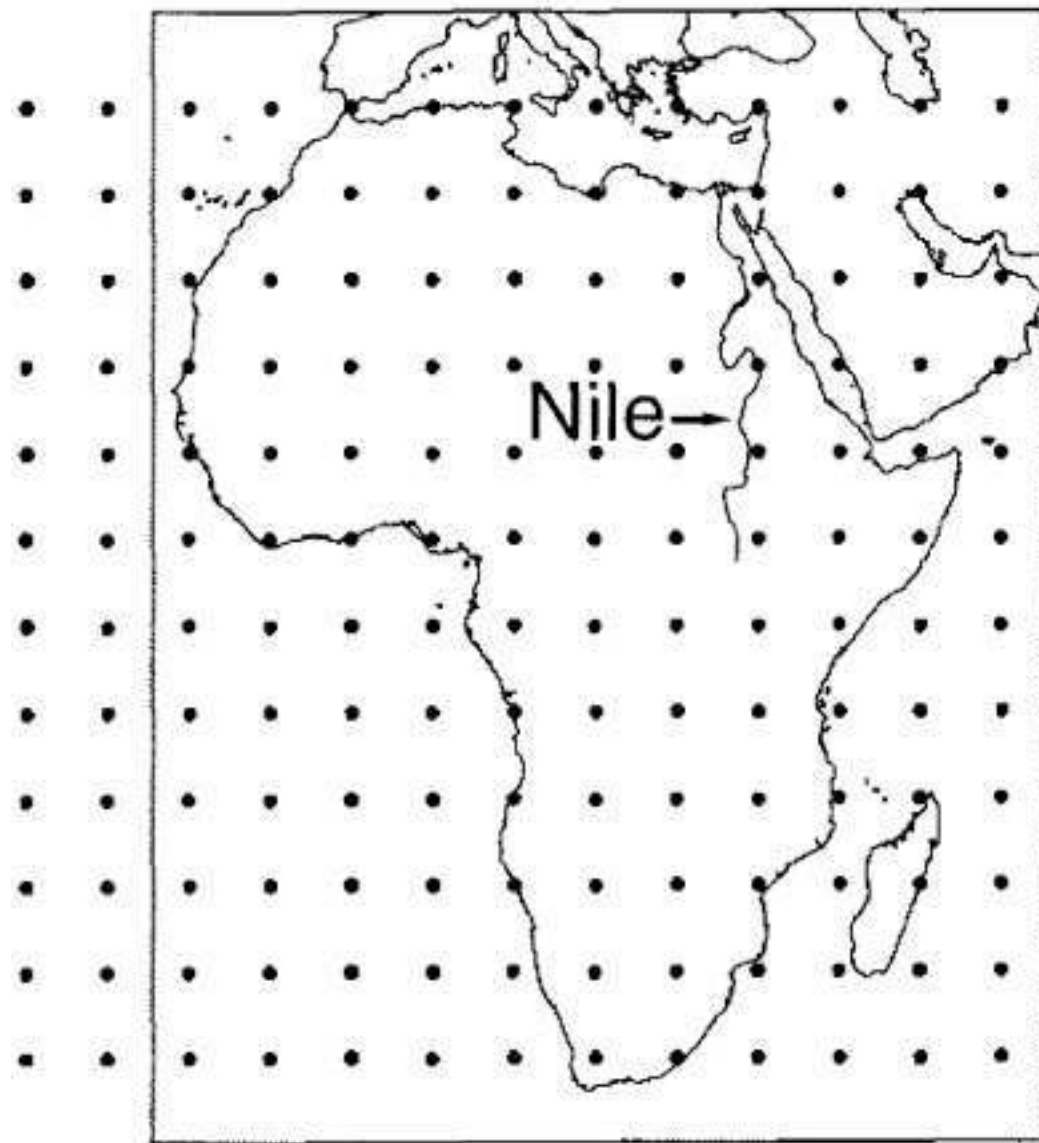
$$\underbrace{\langle F \rangle}_{\text{average}} = \int \underbrace{dr^N dp^N}_{\text{phase space}} \underbrace{F(r^N, p^N)}_{\text{microscopic value}} \underbrace{e^{-\beta \mathcal{H}(r^N, p^N)}}_{\text{probability}} / Z$$

$$\langle F \rangle \approx \frac{1}{L} \sum_{i=1}^L n_i F(r_i^N)$$

$$n_i = LP(r_i^N)/Z = L \exp[-\beta H(r_i^N)]/Z$$

- Метрополис:
 - меряем глубину
 - не можем измерить площадь
- Численное интегрирование:
 - измеряет обе величины, но медленно

Метод Метрополиса

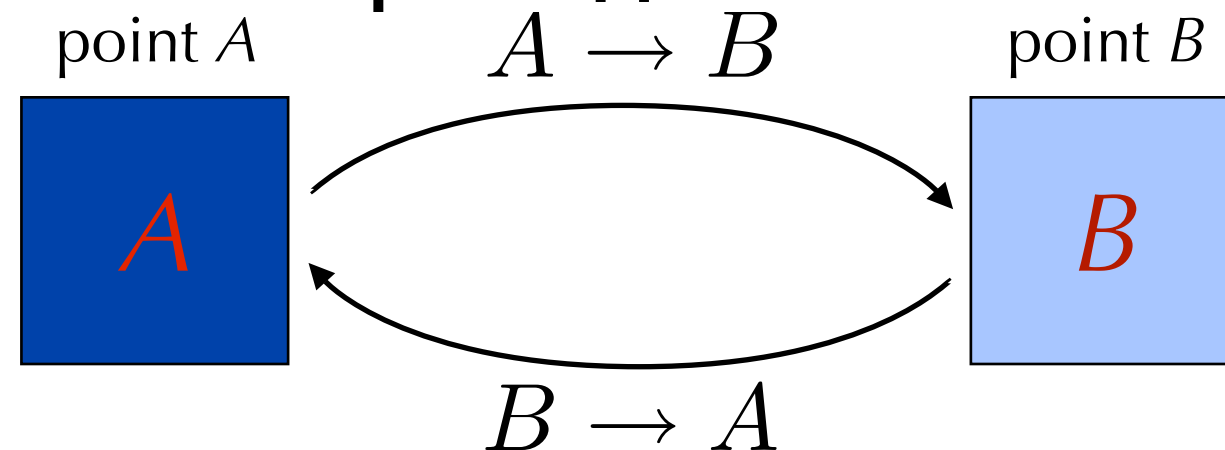


Измеряя глубину Нила, численное интегрирование
(слева) в Метрополис (справа)

Выборка

$$\underbrace{\langle F \rangle}_{\text{average}} = \int \underbrace{dr^N dp^N}_{\text{phase space}} \underbrace{F(r^N, p^N)}_{\text{microscopic value}} \underbrace{e^{-\beta \mathcal{H}(r^N, p^N)}}_{\text{probability}} / Z$$

Посмотрим на переход:

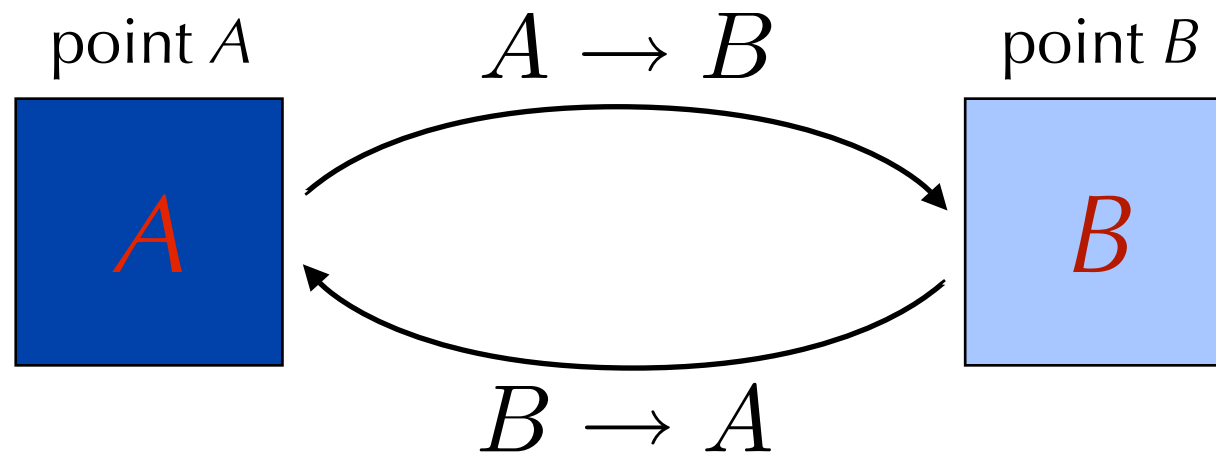


Алгоритм:

- Генерируем новое состояние B с вероятностью $P_{\text{gen}}^{A \rightarrow B}$
- Принимаем или отвергаем с вероятностью $P_{\text{acc}}^{A \rightarrow B}$

Выборка

$$\underbrace{\langle F \rangle}_{\text{average}} = \int \underbrace{dr^N dp^N}_{\text{phase space}} \underbrace{F(r^N, p^N)}_{\text{microscopic value}} \underbrace{e^{-\beta \mathcal{H}(r^N, p^N)}}_{\text{probability}} / Z$$



То, что всегда работает (детальное равновесие):

$$P^{A \rightarrow B} P^A = P^{B \rightarrow A} P^B$$

В нашем случае:

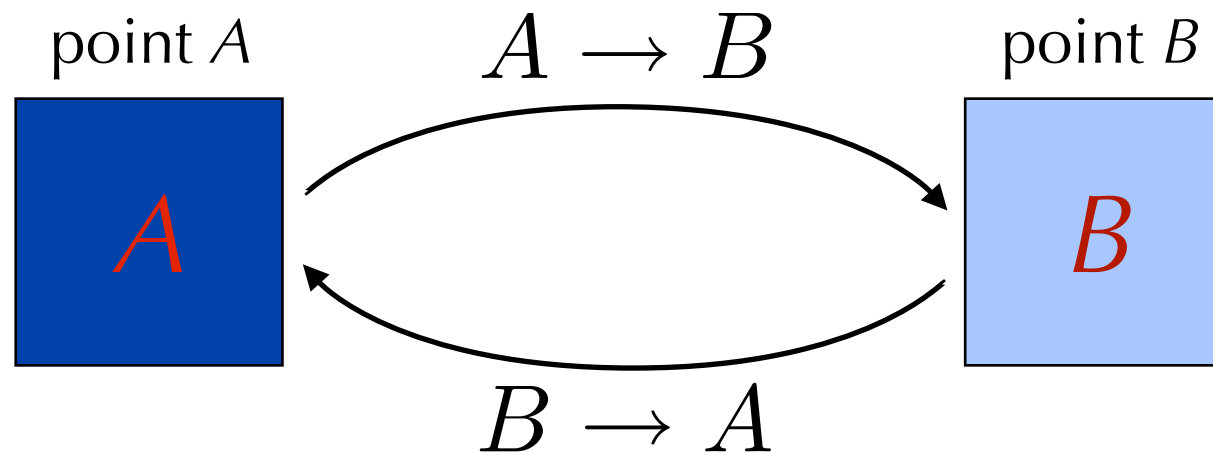
$$P_{\text{gen}}^{A \rightarrow B} P_{\text{acc}}^{A \rightarrow B} P^A = P_{\text{gen}}^{B \rightarrow A} P_{\text{acc}}^{B \rightarrow A} P^B$$

И мы полагаем

$$P_{\text{gen}}^{A \rightarrow B} = P_{\text{gen}}^{B \rightarrow A}$$

Выборка

$$\underbrace{\langle F \rangle}_{\text{average}} = \int \underbrace{dr^N dp^N}_{\text{phase space}} \underbrace{F(r^N, p^N)}_{\text{microscopic value}} \underbrace{e^{-\beta \mathcal{H}(r^N, p^N)}}_{\text{probability}} / Z$$



$$\frac{P_{\text{acc}}^{A \rightarrow B}}{P_{\text{acc}}^{B \rightarrow A}} = \frac{P^B}{P^A} = e^{-\beta \Delta \mathcal{H}}$$

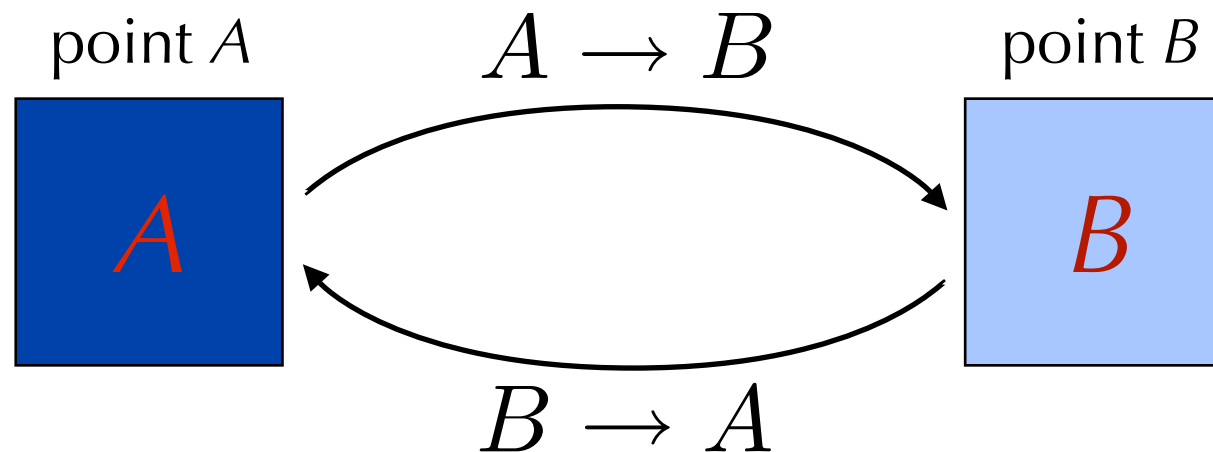
так, например

$$P_{\text{acc}}^{A \rightarrow B} = \min \left(1, e^{-\beta \Delta \mathcal{H}^{A \rightarrow B}} \right)$$

работает

Выборка

$$\underbrace{\langle F \rangle}_{\text{average}} = \int \underbrace{dr^N dp^N}_{\text{phase space}} \underbrace{F(r^N, p^N)}_{\text{microscopic value}} \underbrace{e^{-\beta \mathcal{H}(r^N, p^N)}}_{\text{probability}} / Z$$



Алгоритм:

- Генерируем новое состояние B
 - Делаем это так, чтобы процесс был обратим:

$$P_{\text{gen}}^{A \rightarrow B} = P_{\text{gen}}^{B \rightarrow A}$$

- Принимаем или отвергаем с вероятностью

$$P_{\text{acc}}^{A \rightarrow B} = \min \left(1, e^{-\beta \Delta \mathcal{H}^{A \rightarrow B}} \right)$$

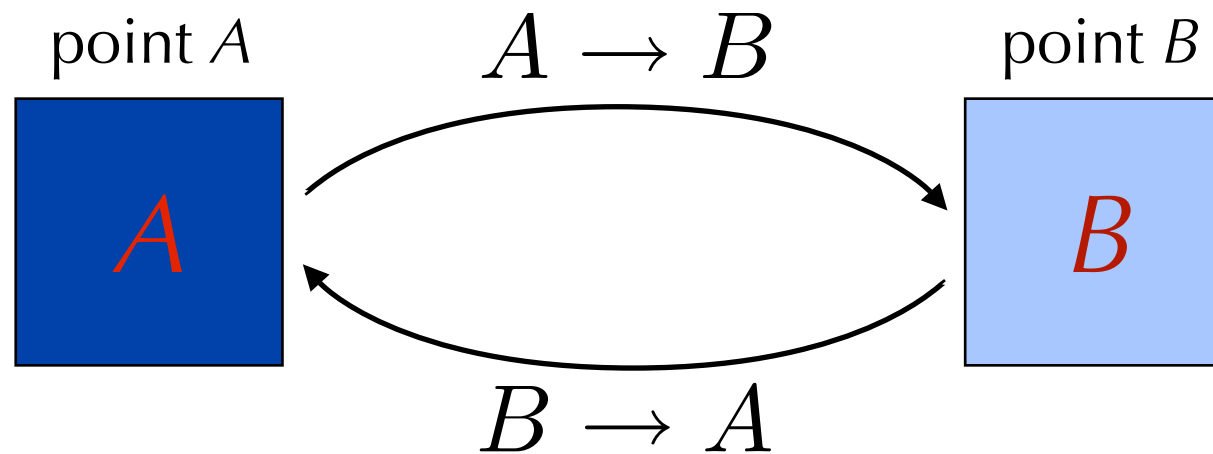
Вопрос

Разные температуры:

$$P_{acc}^{A(T_1), B(T_2) \rightarrow B(T_1), A(T_2)} = ?$$

Ответ: Parallel Tempering

$$\underbrace{\langle F \rangle}_{\text{average}} = \int \underbrace{dr^N dp^N}_{\text{phase space}} \underbrace{F(r^N, p^N)}_{\text{microscopic value}} \underbrace{e^{-\beta \mathcal{H}(r^N, p^N)}}_{\text{probability}} / Z$$



$$P_{gen}^{A(T_1), B(T_2) \rightarrow B(T_1), A(T_2)} \times P_{acc}^{A(T_1), B(T_2) \rightarrow B(T_1), A(T_2)} \times P^A(T_1) P^B(T_2) \\ = P_{gen}^{A(T_2), B(T_1) \rightarrow B(T_2), A(T_1)} \times P_{acc}^{A(T_2), B(T_1) \rightarrow B(T_2), A(T_1)} \times P^A(T_2) P^B(T_1)$$

$$\frac{P_{acc}^{A(T_1), B(T_2) \rightarrow B(T_1), A(T_2)}}{P_{acc}^{A(T_2), B(T_1) \rightarrow B(T_2), A(T_1)}} = \frac{P^A(T_2) P^B(T_1)}{P^A(T_1) P^B(T_2)} = \frac{\exp^{-\beta_2 H^A - \beta_1 H^B}}{\exp^{-\beta_1 H^A - \beta_2 H^B}}$$

$$P_{acc}^{A(T_1), B(T_2) \rightarrow B(T_1), A(T_2)} = \min \left(1, \exp^{-(\beta_1 - \beta_2) \Delta H^{A \rightarrow B}} \right)$$

МС Алгоритм

1. Случайно выбираем частицу (A) и вычисляем ее энергию $H^A(r)$

2. Смещаем ее на случайное расстояние и вычисляем новую энергию (B)
 $H^B(r + \delta r)$

3. Принимаем новое положение с вероятностью
 $P_{\text{acc}}^{A \rightarrow B} = \min \left(1, e^{-\beta \Delta \mathcal{H}^{A \rightarrow B}} \right)$

МС

- Очень просто!
- Нужны только энергии, без сил!
- Возможны нефизические смещения атомов!

Детальное равновесие

детальное равновесие:

$$P^{A \rightarrow B} P^A = P^{B \rightarrow A} P^B$$

- возможны более слабые условия равновесия:

$$P^{A \rightarrow B} P^A = \sum_i P^{B_i \rightarrow A} P^{B_i}$$

- например, если $P_{gen}^{A \rightarrow B} \neq P_{gen}^{B \rightarrow A}$
- смещаем атомы один за одним - работает!!!
- но выполнение детального равновесия намного проще проверить

МС: Пробные смещения

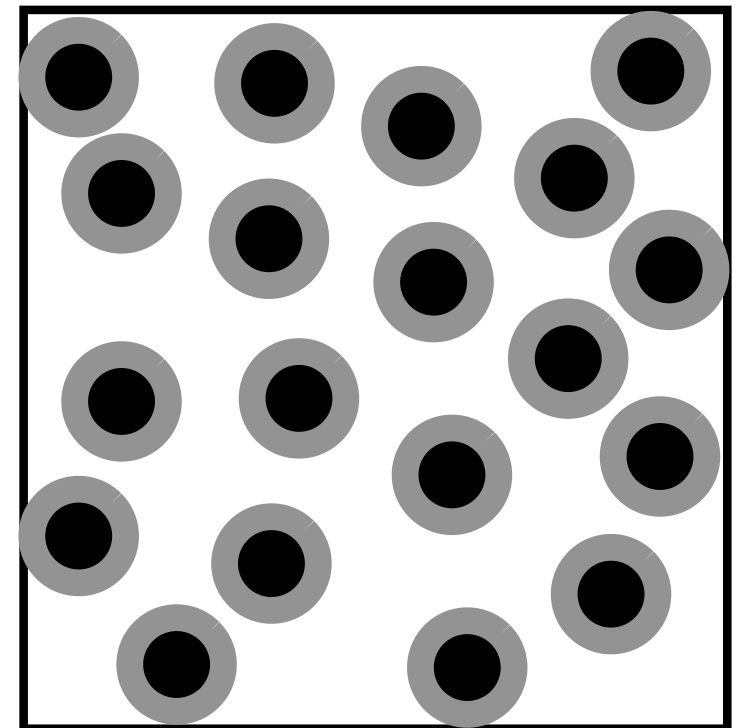
$$x'_i \rightarrow x_i + \Delta(Ranf - 0.5)$$

$$y'_i \rightarrow y_i + \Delta(Ranf - 0.5)$$

$$z'_i \rightarrow z_i + \Delta(Ranf - 0.5)$$

- обратное пробное смещение равновероятно!
- также возможны смещения в сфере (а не кубе)
- Каков размер смещения?
- Следует ли нам двигать частицы все вместе или одну за одной?
 - если двигаем поодиночке - должны выбираться случайно!
 - для конденсированных фазовых состояний
рекомендуемо одно смещение за шаг!
- Система из N сферических частиц с энергией $U(\mathbf{r}^N)$
 - мы ожидаем, что смещение будет отвергнуто при энергии больше $k_B T$
 - мы хотим сделать смещения как можно больше
 - если двигаем поодиночке:

$$\begin{aligned}\langle \Delta \mathcal{U} \rangle &= \left\langle \frac{\partial \mathcal{U}}{\partial r_i^\alpha} \right\rangle \overline{\Delta r_i^\alpha} + \frac{1}{2} \left\langle \frac{\partial^2 \mathcal{U}}{\partial r_i^\alpha \partial r_i^\beta} \right\rangle \overline{\Delta r_i^\alpha \Delta r_i^\beta} + \dots \\ &= 0 + f(\mathcal{U}) \overline{\Delta r_i^2} + \mathcal{O}(\Delta^4),\end{aligned}$$



МС: Пробные смещения

- Система из N сферических частиц с энергией $U(\mathbf{r}^N)$

- двигаем частицы поодиночке:

$$\overline{\Delta r_i^2} \approx k_B T / f(\mathcal{U})$$

$$N \overline{\Delta r^2} \sim N k_B T / f(\mathcal{U})$$

- двигаем N частиц одну за другой, у каждой по n соседей

$$CPU \approx nN$$

- MSD на единицу CPU: $k_B T / (n f(\mathcal{U}))$

- Все частицы разом:

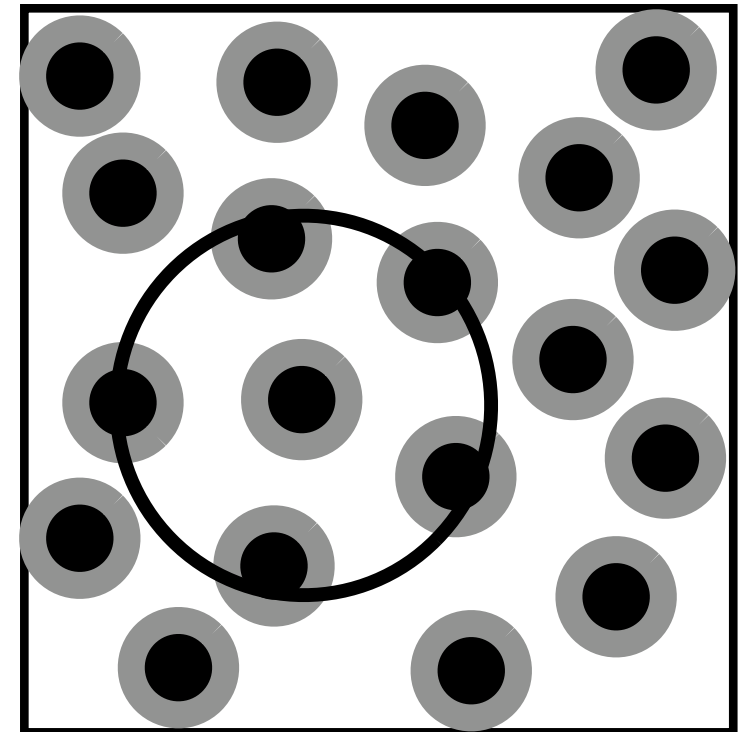
- делаем N независимых пробных смещений:

$$\langle \Delta U \rangle = \left\langle \sum_i^N \Delta U_i \right\rangle = 0 + N \times f(U) \Delta r_i^2 + O(\Delta^4)$$

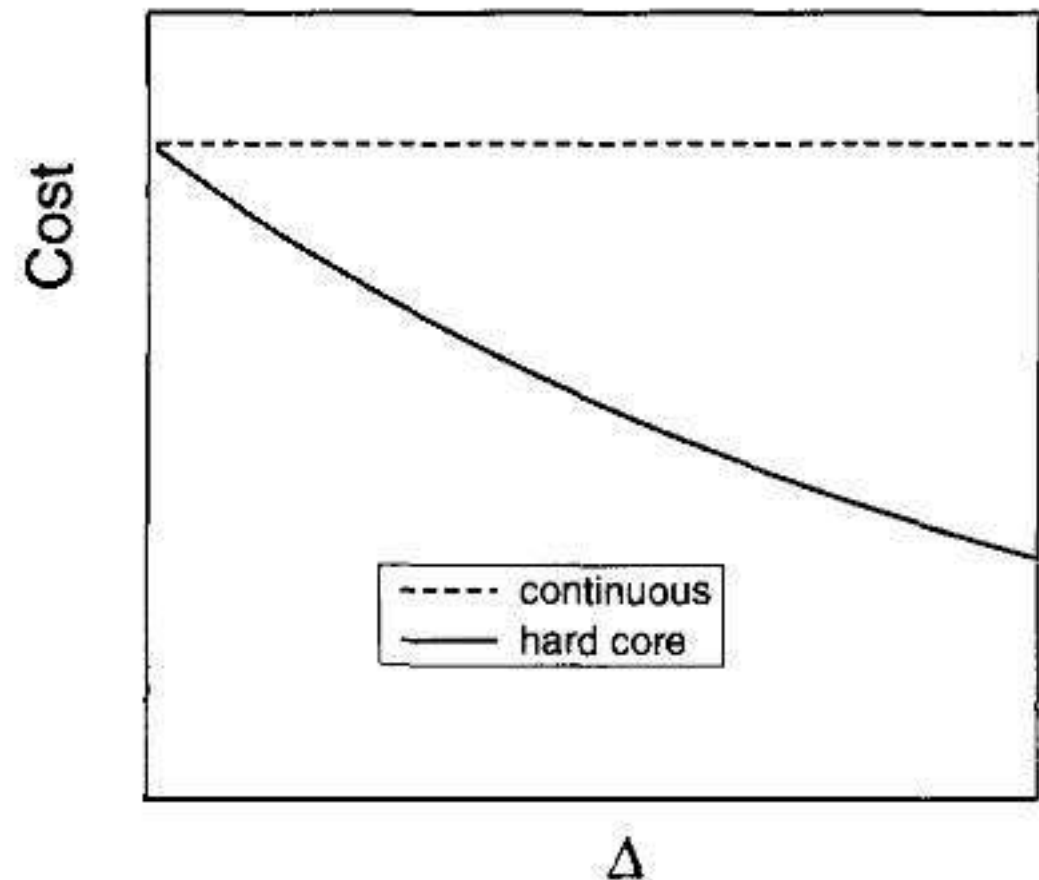
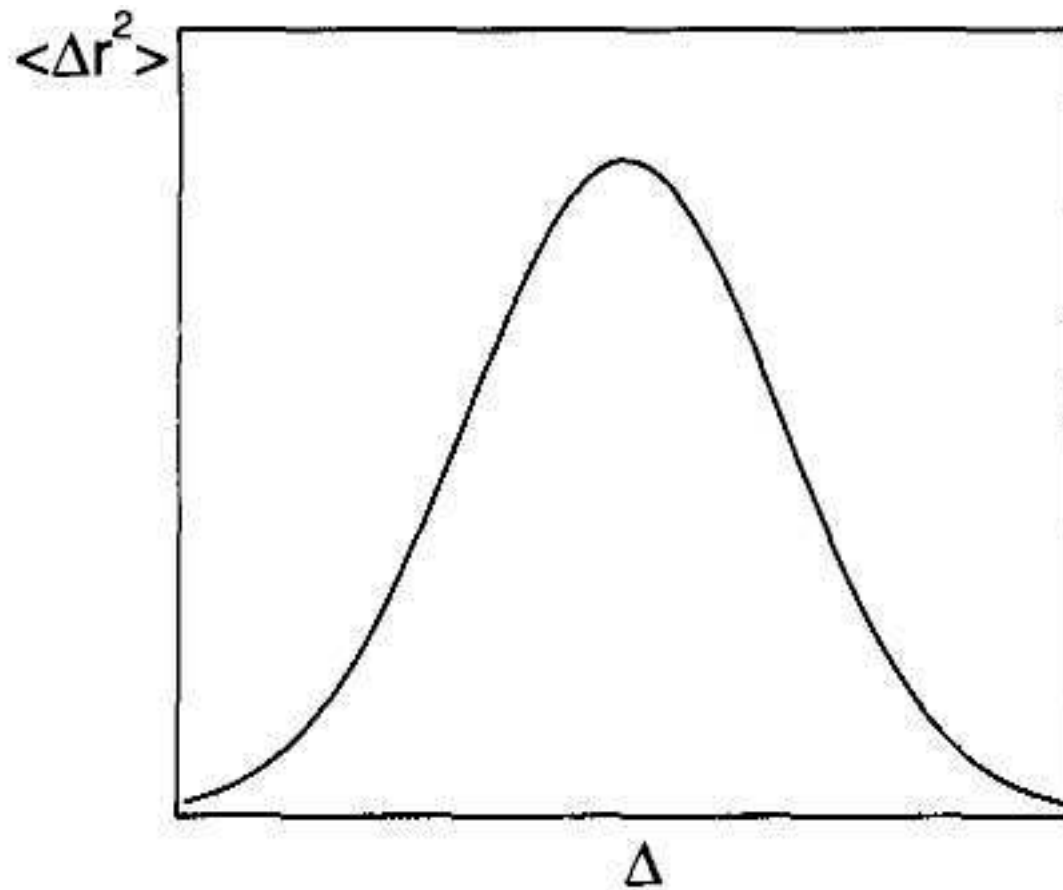
- MSD на единицу CPU : $k_B T / (n f(\mathcal{U})) / N$

- большинство МС алгоритмов используют одиночные смещения

- но также возможны эффективные коллективные смещения!



МС: Пробные смещения



- Какой размер смещения выбрать???
- непрерывные потенциалы: затраты CPU не зависят от размера шага
- жесткое ядро: при отклонении пробного шага затраты меньше
- жесткое ядро: оптимальное значение коэффициента принятия пробного шага будет меньше

МС: Связанные схемы

детальное равновесие:

$$P_{\text{gen}}^{A \rightarrow B} P_{\text{acc}}^{A \rightarrow B} P^A = P_{\text{gen}}^{B \rightarrow A} P_{\text{acc}}^{B \rightarrow A} P^B$$

- Матрица перехода больше не симметрична
- Для удовлетворения условиям детального равновесия нужно изменить правила приема пробного шага

- Общая идея:

$$P_{\text{gen}}^{A \rightarrow B} = f(H^B)$$

- и обратно:

$$P_{\text{gen}}^{B \rightarrow A} = f(H^A)$$

- Накладывая детальное равновесие получаем:

$$\frac{P_{\text{acc}}^{A \rightarrow B}}{P_{\text{acc}}^{B \rightarrow A}} = \frac{f(H^A)}{f(H^B)} \exp[-\beta \Delta H^{A \rightarrow B}]$$

- Работает: $P_{\text{acc}}^{A \rightarrow B} = \min \left(1, \frac{f(H^A)}{f(H^B)} \exp[-\beta \Delta H^{A \rightarrow B}] \right)$

Примеры: связывание по ориентации

- Случайно выбранные частицы переходят из “o” в “n” и меняют ориентацию:
 - Сдвигаем центр масс и вычисляем $u^{\text{pos}}(n)$
 - Генерируем k пробных ориентаций $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k\}$ и для каждой вычисляем $u^{\text{or}}(\mathbf{b}_i)$
 - Вычисляем фактор Розенблюта:
$$W(n) = \sum_{j=1}^k \exp[-\beta u^{\text{or}}(\mathbf{b}_j)]$$

из k ориентаций выбираем одну, n-ю с вероятностью

$$p(\mathbf{b}_n) = \frac{\exp[-\beta u^{\text{or}}(\mathbf{b}_n)]}{\sum_{j=1}^k \exp[-\beta u^{\text{or}}(\mathbf{b}_j)]}$$

- Для старой конфигурации позиционная часть энергии $u^{\text{pos}}(o)$
старая ориентация \mathbf{b}_o и мы генерируем k-1 пробные ориентации $\mathbf{b}_2, \dots, \mathbf{b}_k$
затем вычисляем

$$W(o) = \exp[-\beta u^{\text{or}}(\mathbf{b}_o)] + \sum_{j=2}^k \exp[-\beta u^{\text{or}}(\mathbf{b}_j)]$$

- Шаг принимается если:
$$\text{acc}(o \rightarrow n) = \min \left(1, \frac{W(n)}{W(o)} \exp\{-\beta [u^{\text{pos}}(n) - u^{\text{pos}}(o)]\} \right)$$

Примеры: решеточная модель

- Молекулы в решетке имеют k ориентаций

решеточная модель с k=4 ориентациями

- Потенциал $\mathcal{U} = u^{or} + u^{pos}$

- Условие детального равновесия:

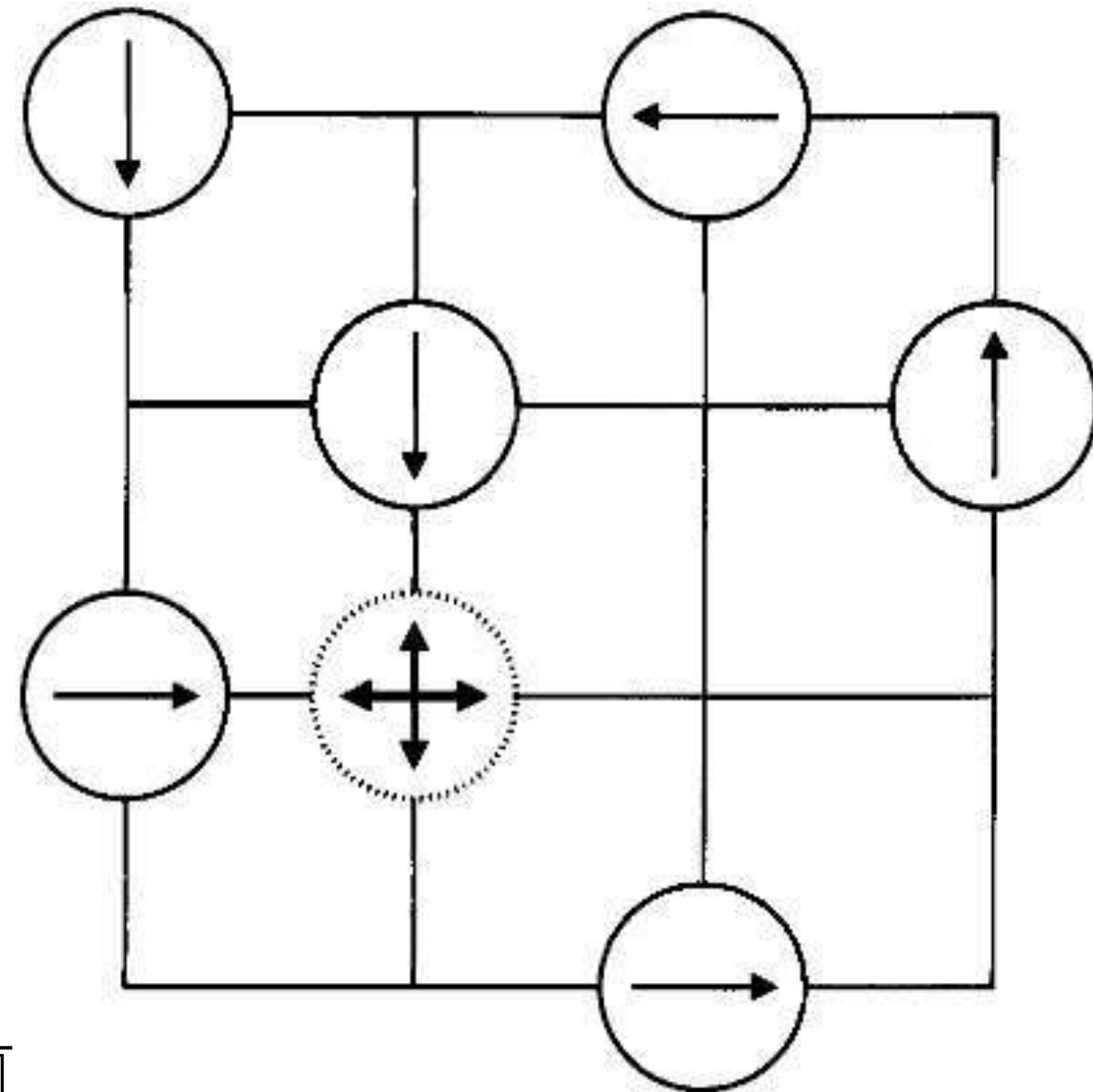
$$P_{gen}^{A \rightarrow B} P_{acc}^{A \rightarrow B} P^A = P_{gen}^{B \rightarrow A} P_{acc}^{B \rightarrow A} P^B$$

- Вероятность выбрать B:

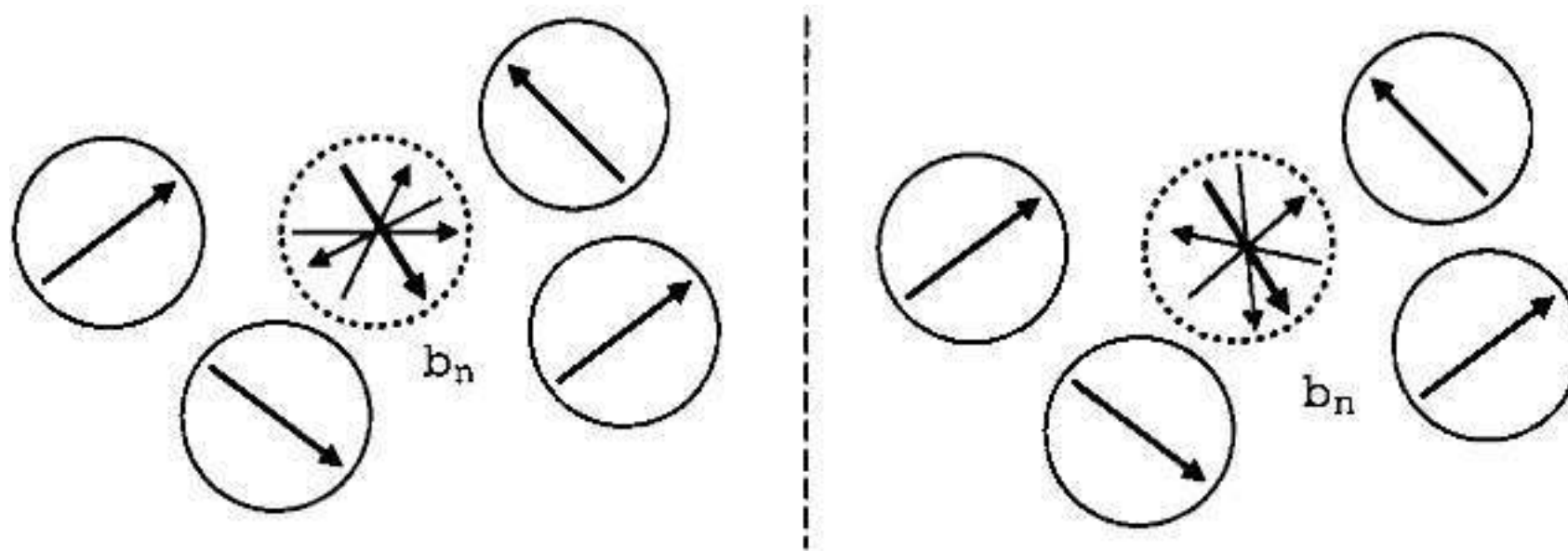
$$P_{gen}^{A \rightarrow B} = \frac{\exp[-\beta u^{or}(B)]}{W(B)}$$

- Правила выбора:

$$\begin{aligned} & \frac{P_{acc}^{A \rightarrow B}}{P_{acc}^{B \rightarrow A}} = \exp[-\beta \Delta U^{A \rightarrow B}] \\ & \times \frac{\exp[-\beta u^{or}(A)]}{W(A)} \times \frac{W(B)}{\exp[-\beta u^{or}(B)]} \\ & = \frac{W(B)}{W(A)} \exp\{-\beta[u^{pos}(B) - u^{pos}(A)]\} \end{aligned}$$



Примеры: непрерывная модель



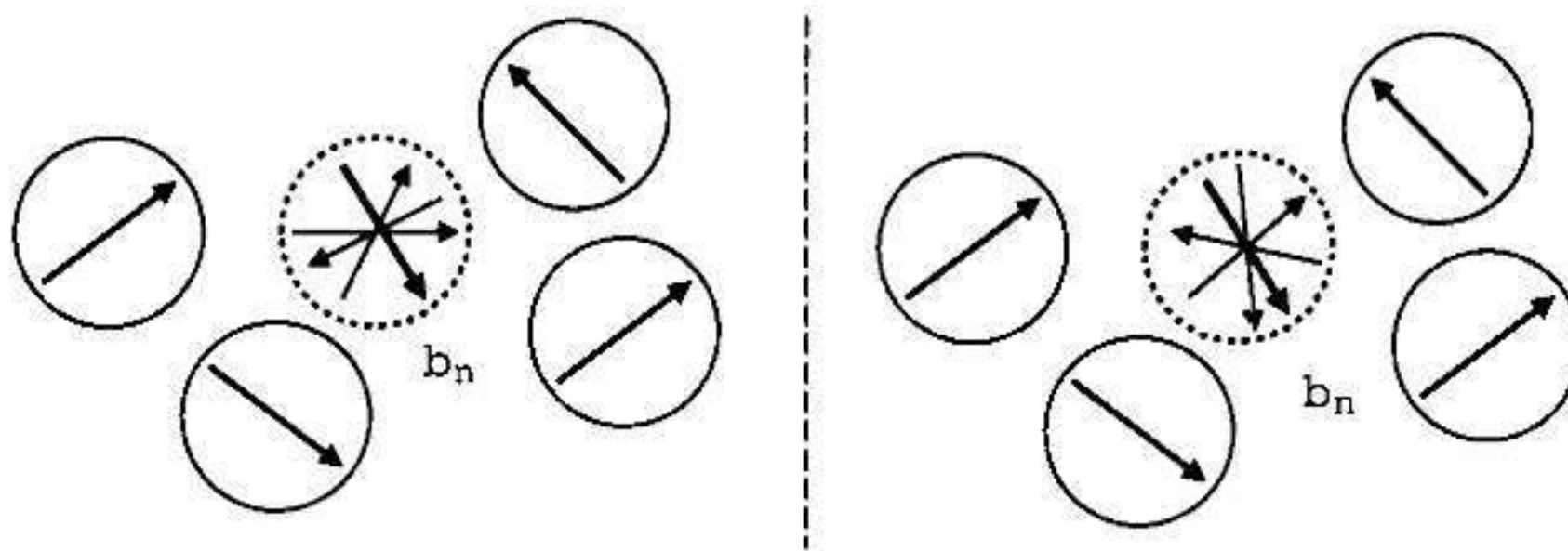
- k пробных ориентаций $\{\mathbf{b}\}_k = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k\}$
- набор всех $\{\mathbf{b}\}_k$ в которых есть конформация n будет $\mathcal{B}_n = \{\{\mathbf{b}\}_k | \mathbf{b}_n \in \{\mathbf{b}\}_k\}$
- каждый элемент \mathcal{B}_n может быть записан как $(\mathbf{b}_n, \mathbf{b}^*)$ где второй элемент - это набор $k-1$ дополнительных ориентаций
- поток конфигураций (принимая во внимание все наборы в \mathcal{B}_n):

$$K(o \rightarrow n) = \mathcal{N}(o) \sum_{i \in \mathcal{B}_n} \alpha(o \rightarrow n, i) \times \text{acc}(o \rightarrow n, i)$$

- обратный шаг: $\mathcal{B}_o = \{\{\mathbf{b}\}_k | \mathbf{b}_o \in \{\mathbf{b}\}_k\}$
каждый элемент - это $(\mathbf{b}_o, \mathbf{b}'^*)$
- обратный поток:

$$K(n \rightarrow o) = \mathcal{N}(n) \sum_{j \in \mathcal{B}_o} \alpha(n \rightarrow o, j) \times \text{acc}(n \rightarrow o, j)$$

Примеры: непрерывная модель



- Условие, более сильное чем детальное равновесие:

$$K(o \rightarrow n, \mathbf{b}^*, \mathbf{b}'^*) = K(n \rightarrow o, \mathbf{b}'^*, \mathbf{b}^*)$$

$$\begin{aligned} \mathcal{N}(o) \alpha(o \rightarrow n, \mathbf{b}^*, \mathbf{b}'^*) \text{acc}(o \rightarrow n, \mathbf{b}^*, \mathbf{b}'^*) \\ = \mathcal{N}(n) \alpha(n \rightarrow o, \mathbf{b}'^*, \mathbf{b}^*) \text{acc}(n \rightarrow o, \mathbf{b}'^*, \mathbf{b}^*) \end{aligned}$$

$$\begin{aligned} \frac{\text{acc}(o \rightarrow n, \mathbf{b}^*, \mathbf{b}'^*)}{\text{acc}(n \rightarrow o, \mathbf{b}'^*, \mathbf{b}^*)} &= \frac{\exp[-\beta \mathcal{U}(n)]}{\exp[-\beta \mathcal{U}(o)]} \frac{\exp[-\beta u^{\text{or}}(o)]}{W(\mathbf{b}_o, \mathbf{b}'^*)} \frac{W(\mathbf{b}_n, \mathbf{b}^*)}{\exp[-\beta u^{\text{or}}(n)]} \\ &= \frac{W(\mathbf{b}_n, \mathbf{b}^*)}{W(\mathbf{b}_o, \mathbf{b}'^*)} \exp\{-\beta [u^{\text{pos}}(n) - u^{\text{pos}}(o)]\}. \end{aligned}$$

Гибридный Монте-Карло

- Создаем новую конфигурацию при помощи MD (используя большие шаги и симплектический интегратор)
 - при каждом шаге скорости частиц выбираются случайным образом
 - каждый шаг может быть набором из нескольких MD шагов
- Шаг по времени не может быть слишком большим - иначе конфигурация будет отклонена
- MD шаги могут быть выполнены используя упрощенный потенциал (который намного быстрее посчитать). После, мы принимаем или отвергаем новую конфигурацию используя настоящий потенциал.

Молекулярная Динамика

Молекулярная Динамика

- Вычисляем траектории отдельных частиц
- Конечно, они будут с ошибками
- Средняя наблюдаемая величина не должна быть чувствительна к ошибкам интегрирования
- Как мы интегрируем?

Уравнения Ньютона

- Начинаем с набора координат
- Вычисляем потенциал и силы на каждый атом
- Из силы на каждый атом вычисляем ускорение

$$a_i = F_i / m_i$$

Уравнения Ньютона

- Из ускорений можно записать изменение скорости:

$$v_i(t + \Delta t) = v_i(t) + a_i \Delta t$$

- Из скоростей – изменение координат:

$$x_i(t + \Delta t) = x_i(t) + v_i \Delta t$$

Уравнения Ньютона

- Решаем уравнения движения Ньютона

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i, \quad i = 1..N$$

$$F_i = - \left. \frac{\partial V(r)}{\partial r} \right|_{r=r_i}$$

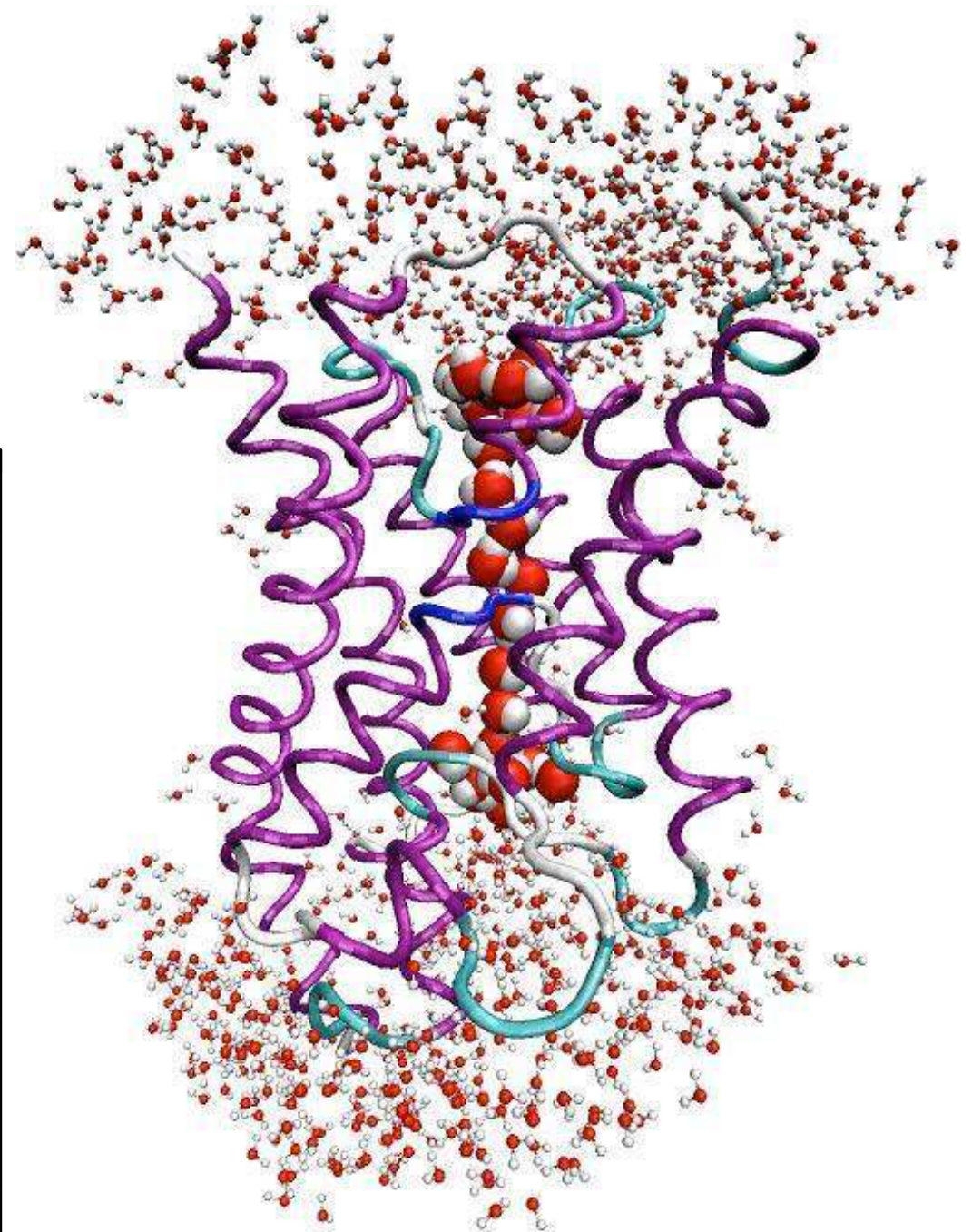
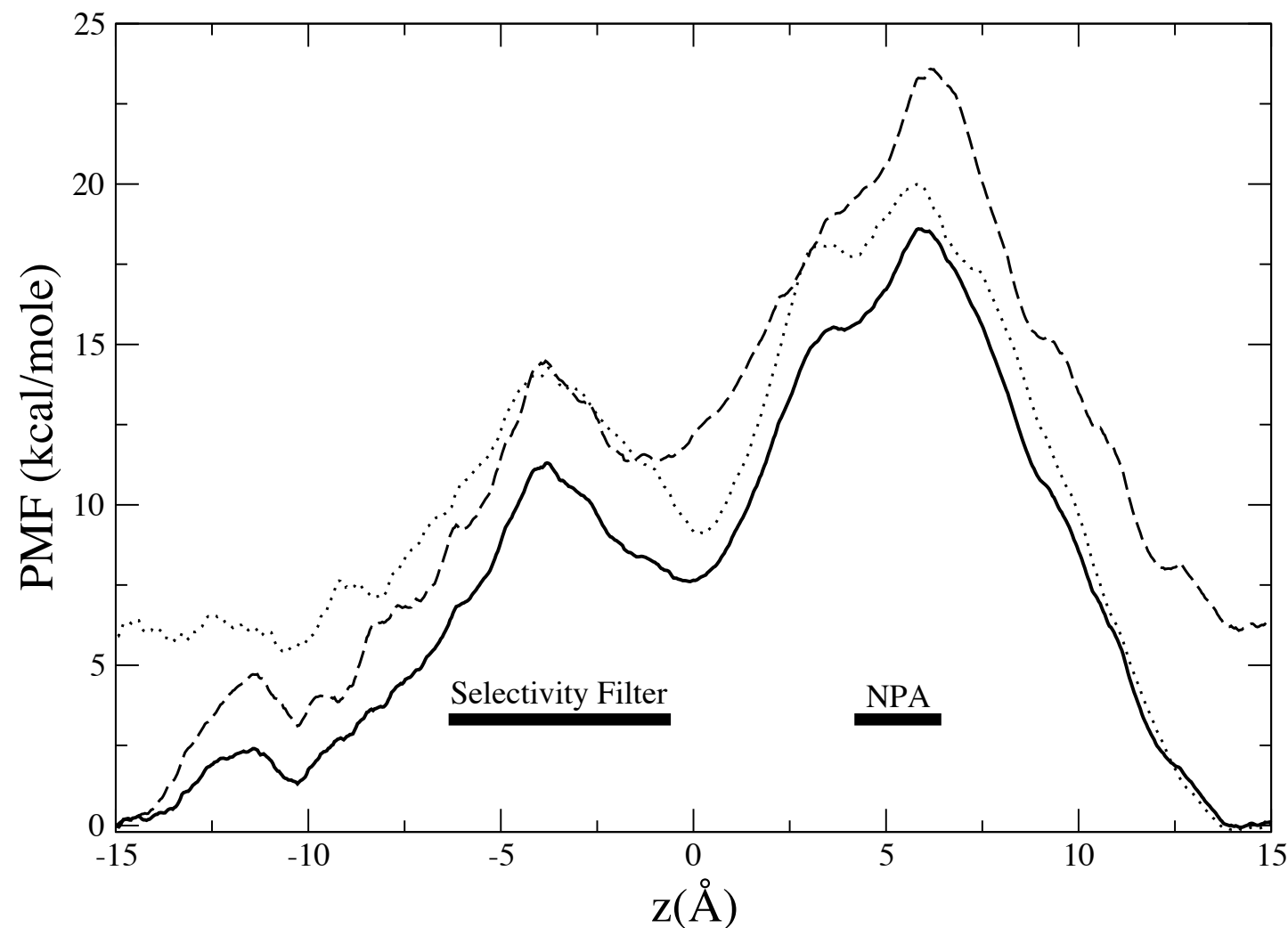
- Шаг должен быть маленьким (фемтосекунды)
- Силы зависят от всех координат в системе (трудоемко для вычислений)
- Получаем траекторию как функцию времени

Что такое траектория?

- НЕ точное предсказание поведения одной частицы
- Молекулярная Динамика – процесс хаотичный, разница в конфигурации растёт экспоненциально
- Но они ведь не случайны?
- «Shadow trajectory»: короткая траектория, идущая вдоль настоящей короткое время

МД не должна быть использована

- Аквапорин
- PMF для протона



Интеграторы

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{\vec{a}(t)\Delta t^2}{2} + \frac{\vec{b}(t)\Delta t^3}{6} + \mathcal{O}(\Delta t^4)$$

$$\vec{x}(t - \Delta t) = \vec{x}(t) - \vec{v}(t)\Delta t + \frac{\vec{a}(t)\Delta t^2}{2} - \frac{\vec{b}(t)\Delta t^3}{6} + \mathcal{O}(\Delta t^4)$$

Верле:

$$\vec{x}(t + \Delta t) = 2\vec{x}(t) - \vec{x}(t - \Delta t) + \vec{a}(t)\Delta t^2 + \mathcal{O}(\Delta t^4)$$

- Ошибка $\mathcal{O}(\Delta t^4)$
- Симплектичен
- Нет явных скоростей

Интеграторы

Скоростной Верле:

$$\vec{x}(t + \Delta t) = 2\vec{x}(t) - \vec{x}(t - \Delta t) + \vec{a}(t)\Delta t^2 + \mathcal{O}(\Delta t^4)$$

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \frac{\vec{a}(t) + \vec{a}(t + \Delta t)}{2}\Delta t + \mathcal{O}(\Delta t^2)$$

Алгоритм:

1. $\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + (1/2)\mathbf{a}(t)\Delta t^2$
2. $\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t) + (1/2)\mathbf{a}(t)\Delta t$
3. $\mathbf{a}(t + \Delta t) = -(1/m)\nabla V(\mathbf{r}(t + \Delta t))$
4. $\mathbf{v}(t + \Delta t) = \mathbf{v}(t + \Delta t/2) + (1/2)\mathbf{a}(t + \Delta t)\Delta t$

- Ошибка в координатах $\mathcal{O}(\Delta t^4)$
- Ошибка в скоростях $\mathcal{O}(\Delta t^2)$
- Симплектичен

Интеграторы

Скорректированный Скоростной Верле:

$$r(t + 2\Delta t) = r(t) + 2v(t)\Delta t + \dot{v}(t)(2\Delta t)^2/2! + \ddot{v}(2\Delta t)^3/3! + \dots$$

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \dot{v}(t)\Delta t^2/2! + \ddot{v}\Delta t^3/3! + \dots$$

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \dot{v}(t)\Delta t^2/2! - \ddot{v}\Delta t^3/3! + \dots$$

$$r(t - 2\Delta t) = r(t) - 2v(t)\Delta t + \dot{v}(t)(2\Delta t)^2/2! - \ddot{v}(2\Delta t)^3/3! + \dots$$

$$12v(t)\Delta t = 8[r(t + \Delta t) - r(t - \Delta t)] - [r(t + 2\Delta t) - r(t - 2\Delta t)] + \mathcal{O}(\Delta t^4)$$

$$v(t) = \frac{v(t + \Delta t/2) + v(t - \Delta t/2)}{2} + \frac{\Delta t}{12}[\dot{v}(t - \Delta t) - \dot{v}(t + \Delta t)] + \mathcal{O}(\Delta t^4)$$

- Мы можем посчитать скорости только после следующего шага

Почему это работает?

$$A(\mathbf{r}^N, \mathbf{p}^N)$$

дифференцируем

$$\dot{A} = \sum_i \dot{\mathbf{r}}_i \cdot \frac{\partial A}{\partial \mathbf{r}_i} + \dot{\mathbf{p}}_i \cdot \frac{\partial A}{\partial \mathbf{p}_i} \equiv i\mathbf{L}A$$

После формального
интегрирования

$$A(t) = e^{i\mathbf{L}t} A(0)$$

$$e^{i\mathbf{L}t} = \left(e^{i\mathbf{L}\delta t} \right)_{\text{approx}}^{n_{\text{step}}} + \mathcal{O}(n_{\text{step}}\delta t^3) \quad \delta t = t/n_{\text{step}}$$

Теорема Троттера

$$e^{(A+B)} = \left(e^{A/2P} e^{B/P} e^{A/2P} \right)^P e^{\mathcal{O}(1/P^2)}$$

$$i\mathbf{L} = i\mathbf{L}_p + i\mathbf{L}_r$$

Почему это работает?

$$iL_p = \sum_i \dot{\mathbf{p}}_i \cdot \frac{\partial}{\partial \mathbf{p}_i} = \sum_i \mathbf{f}_i \cdot \frac{\partial}{\partial \mathbf{p}_i}$$

$$iL_r = \sum_i \dot{\mathbf{r}}_i \cdot \frac{\partial}{\partial \mathbf{r}_i} = \sum_i m_i^{-1} \mathbf{p}_i \cdot \frac{\partial}{\partial \mathbf{r}_i}$$

Раскладываем оператор

$$\exp\left(a \frac{\partial}{\partial x}\right) f(x) \equiv 1 + a \frac{\partial f}{\partial x} + \frac{a^2}{2!} \frac{\partial^2 f}{\partial x^2} + \dots = f(x + a)$$

$$e^{iL_r \delta t} A(\mathbf{r}, \mathbf{p}) = A(\mathbf{r} + m^{-1} \mathbf{p} \delta t, \mathbf{p})$$

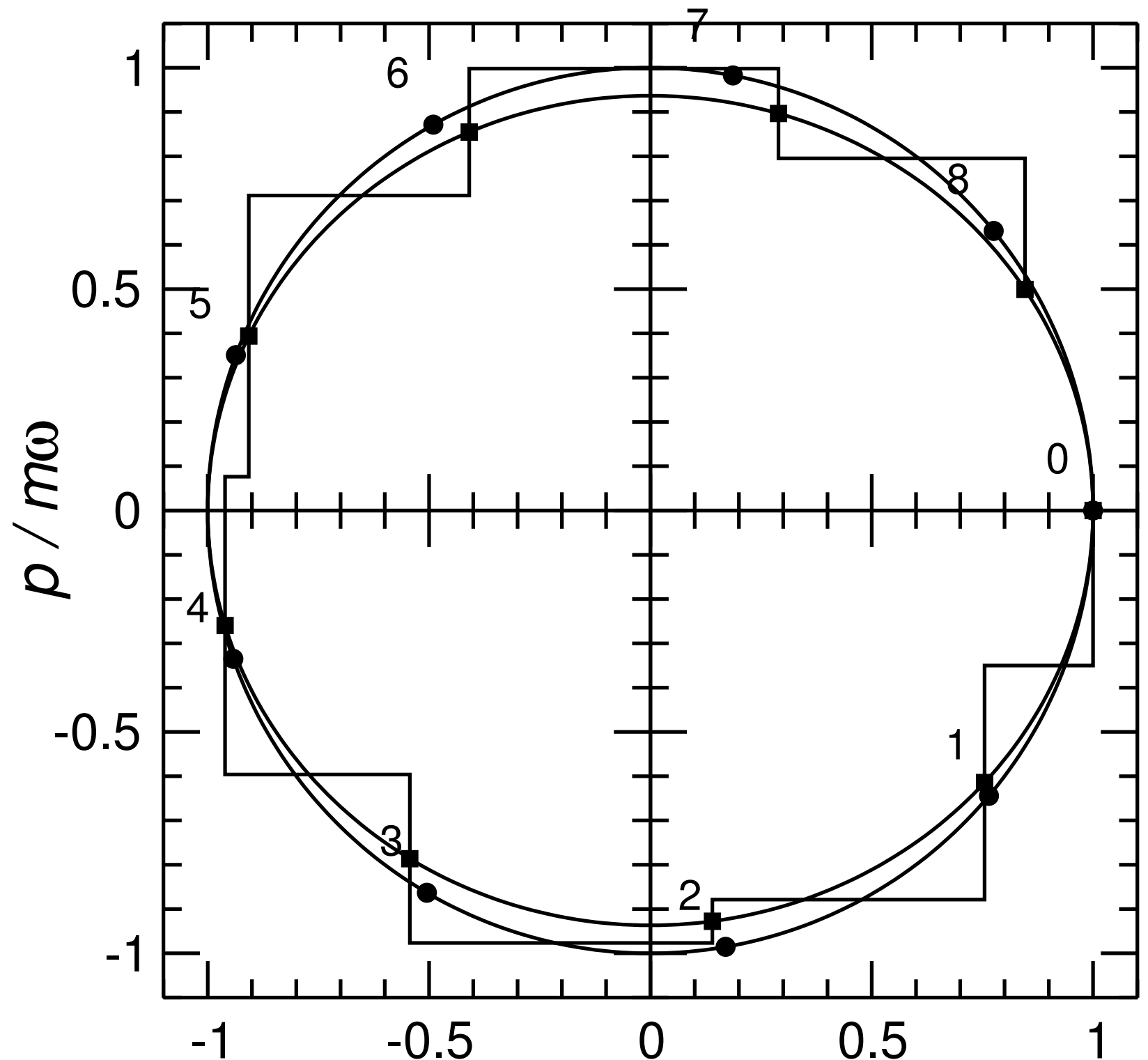
$$e^{iL_p \delta t} A(\mathbf{r}, \mathbf{p}) = A(\mathbf{r}, \mathbf{p} + \mathbf{f} \delta t)$$

$$e^{iL \delta t} = e^{(iL_p + iL_r) \delta t} \approx e^{iL_p \delta t / 2} e^{iL_r \delta t} e^{iL_p \delta t / 2}$$

- Асимптотически точен при $\delta t \rightarrow 0$
- Получаем «Velocity-Verlet algorithm»
- По аналогии возможны схемы более высоких порядков

Почему это работает?

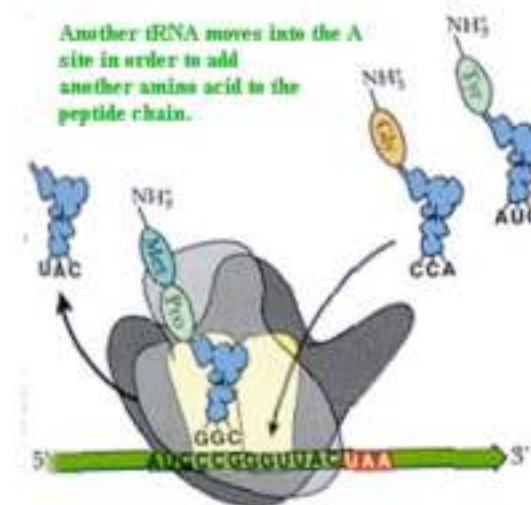
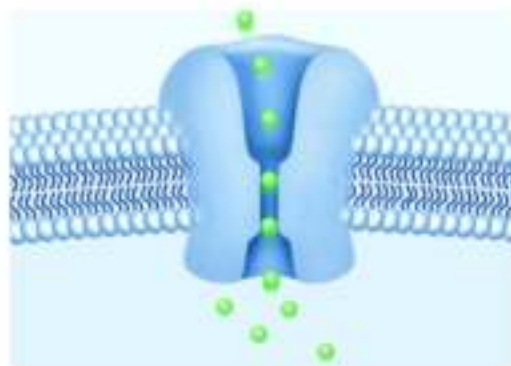
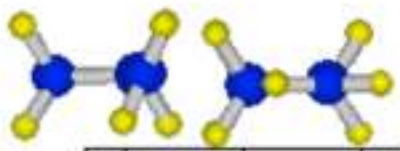
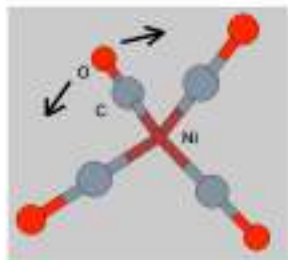
- “Velocity Verlet” для простого гармонического осциллятора
- Внешний круг - настоящий H
- Внутренний эллипс – «shadow» H



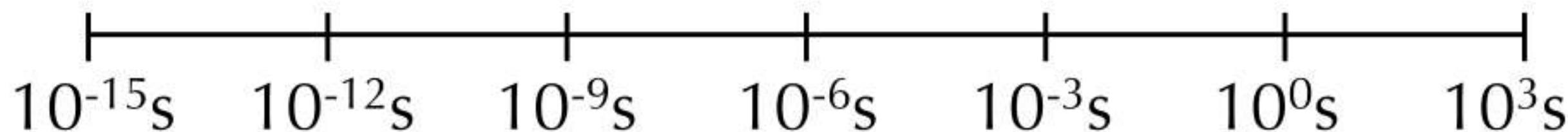
Как мы запускаем МД?

- Получаем структуру белка (координаты атомов)
- Вставляем в мембрану
- Добавляем растворитель вне мембраны
- Добавляем противоионы (для правильной концентрации или для общего нулевого заряда)
- Запускаем вычисления!

Времена



Biological Experiments

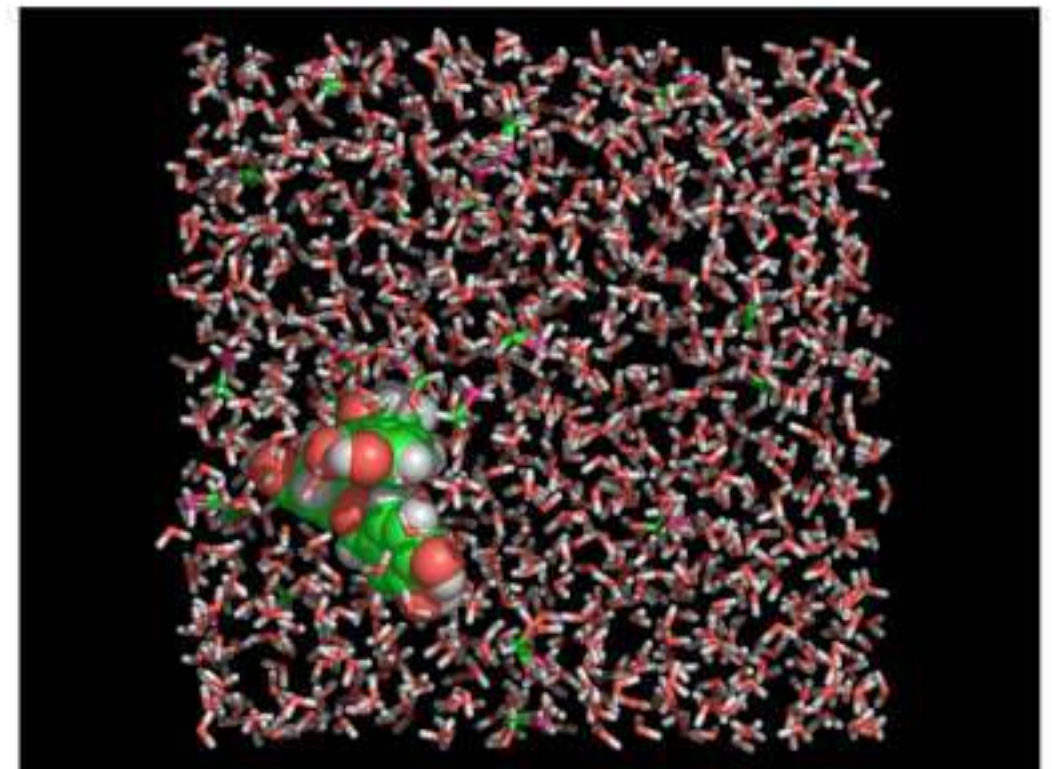


Мы сейчас здесь

А нужно быть здесь

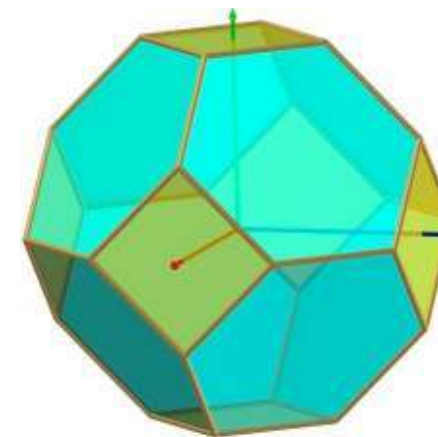
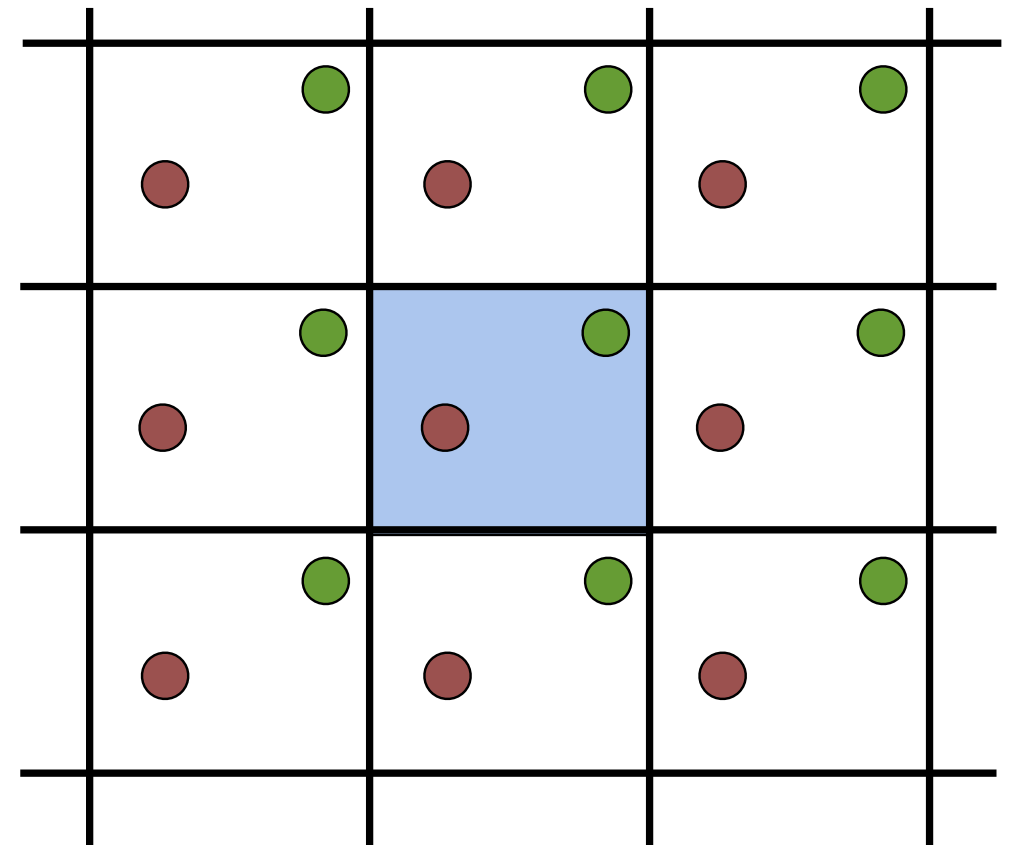
Граничные условия

- Вакуум – без растворителя
- Неявные модели растворителя (трудоемко, невысокая точность)
- Капля: Сферический слой воды вокруг белка
- Периодические Граничные условия (РВС) – вода справа перепрыгивает налево

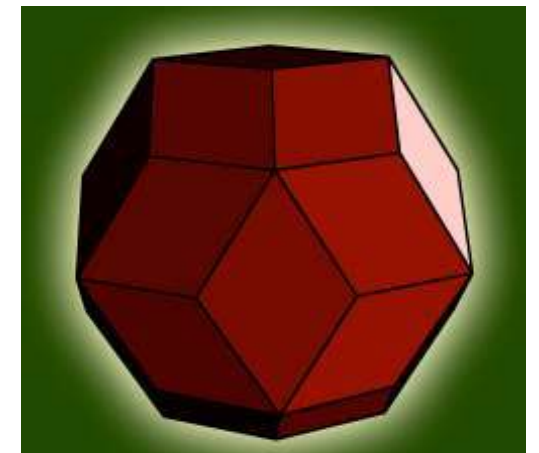


Формы периодических ячеек

- Кубическая
 - Самая широко используемая!
- Усеченный октаэдр (похож на сферу)
- Ромбический додекаэдр (для большинства сферических систем)
- Объем октаэдра 77% от объема куба, додекаэдр 71% при одинаковом периоде!



Усеченный
октаэдр

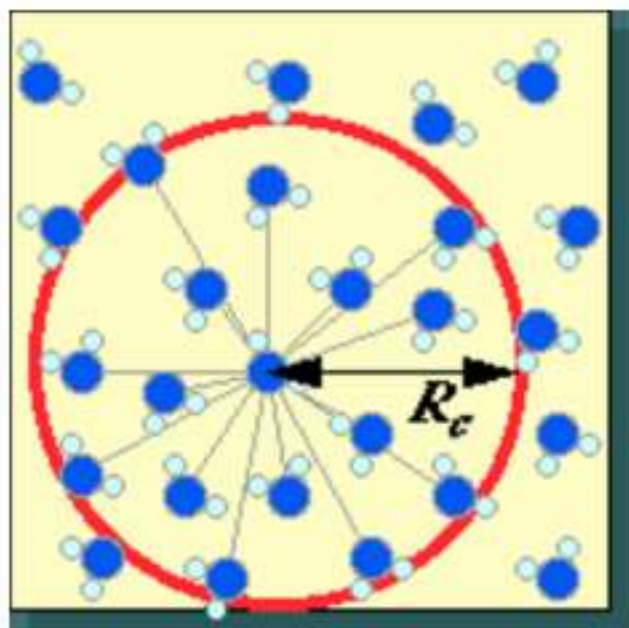


Ромбический
додекаэдр

Невалентные взаимодействия

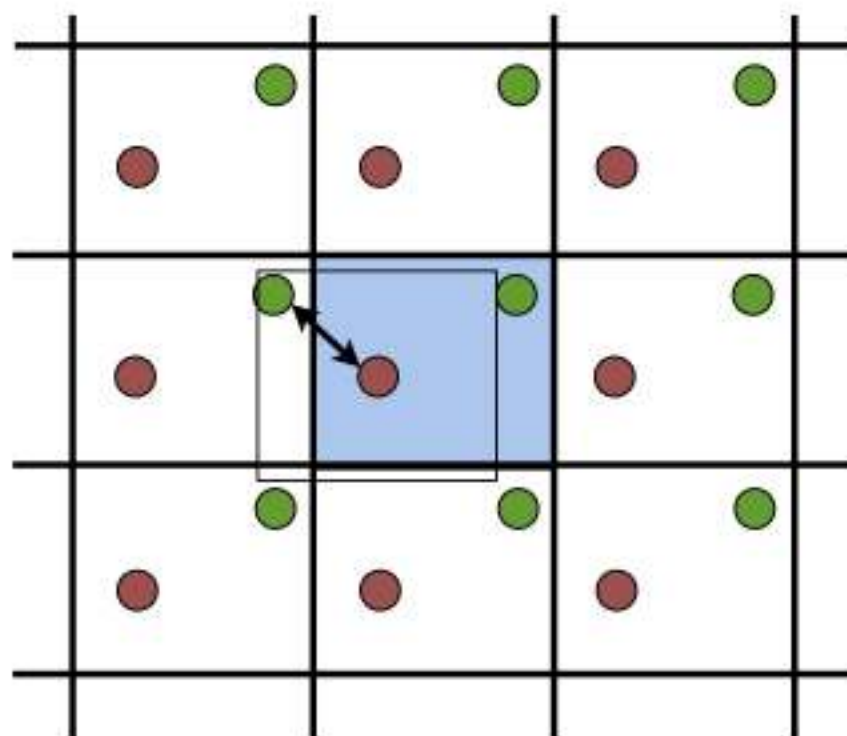
- Огромное количество взаимодействий после добавления растворителя
- Что происходит на границе?
 - Взаимодействия с атомами с другой стороны!
- Невозможно посчитать взаимодействия со всеми копиями
- Вводим обрезание потенциала и списки взаимодействий

Обрезание потенциала



Взаимодействия со всеми атомами, включая копии, в пределах расстояния обрезания потенциала

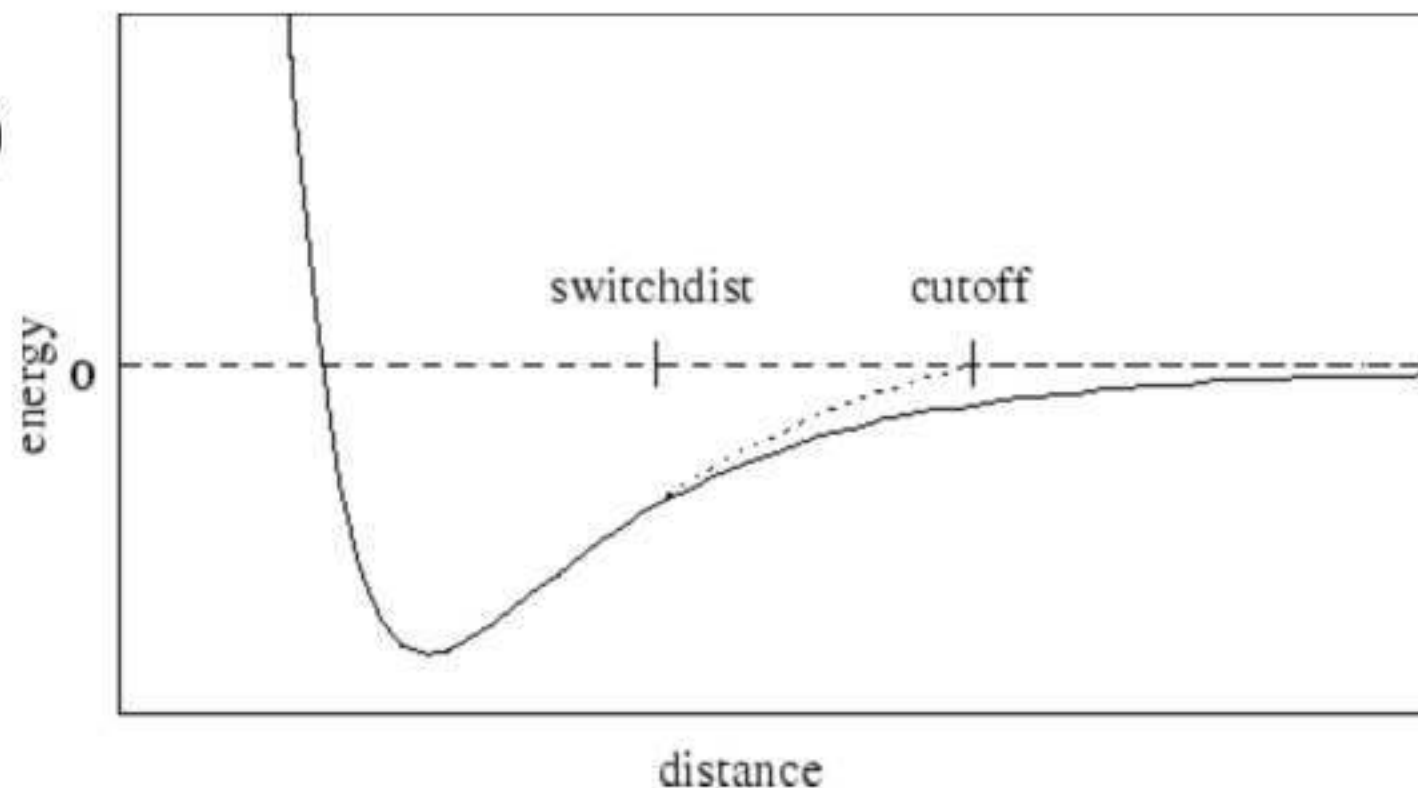
Считаем взаимодействия только с ближайшей копией атома: «*minimum image convention*»



Обрезание потенциала

$$E_{vdw} = \left(\frac{A}{|\vec{r}_{ij}|^{12}} - \frac{B}{|\vec{r}_{ij}|^6} \right) SW(|\vec{r}_{ij}|)$$

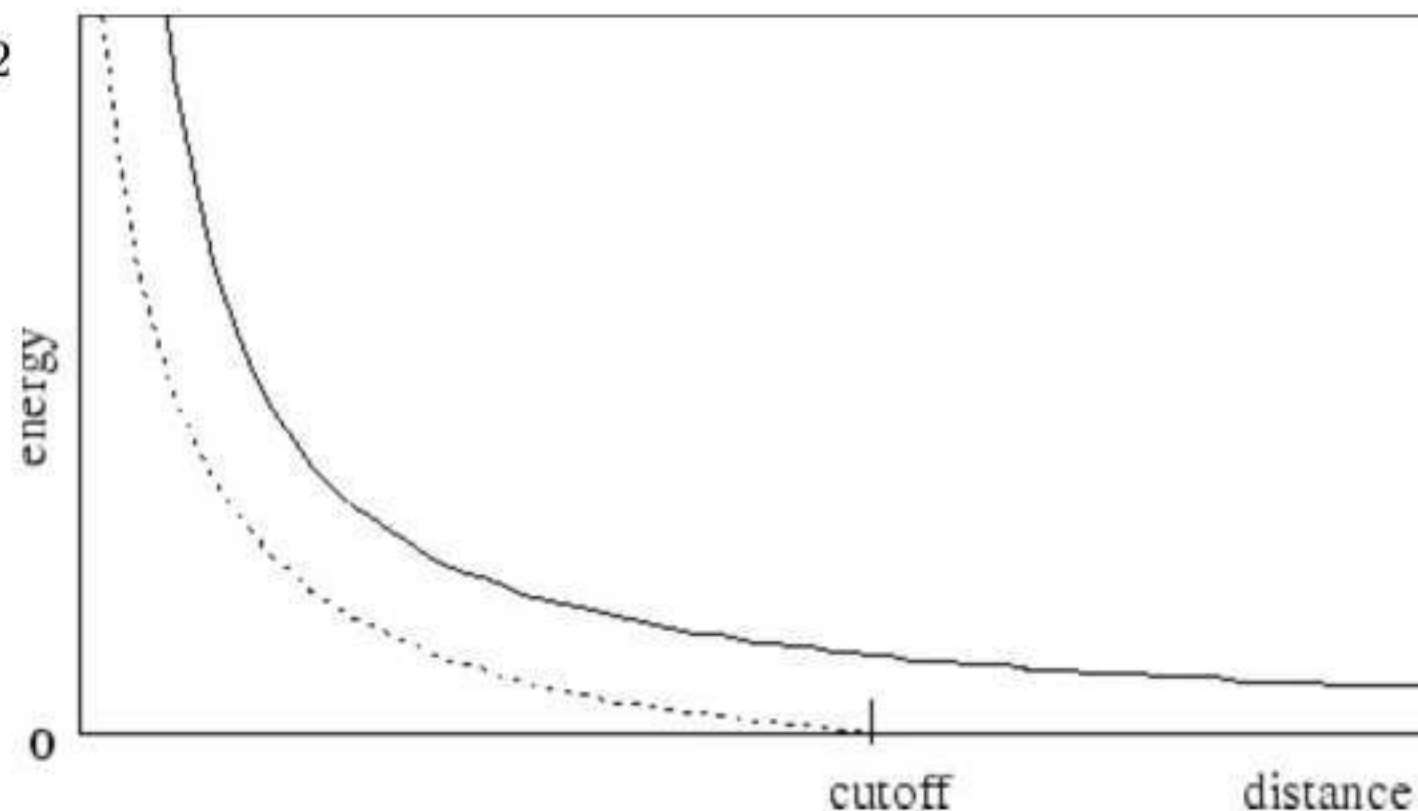
- Не можем обрезать просто так
- Вместо этого, мы вводим функцию переключения



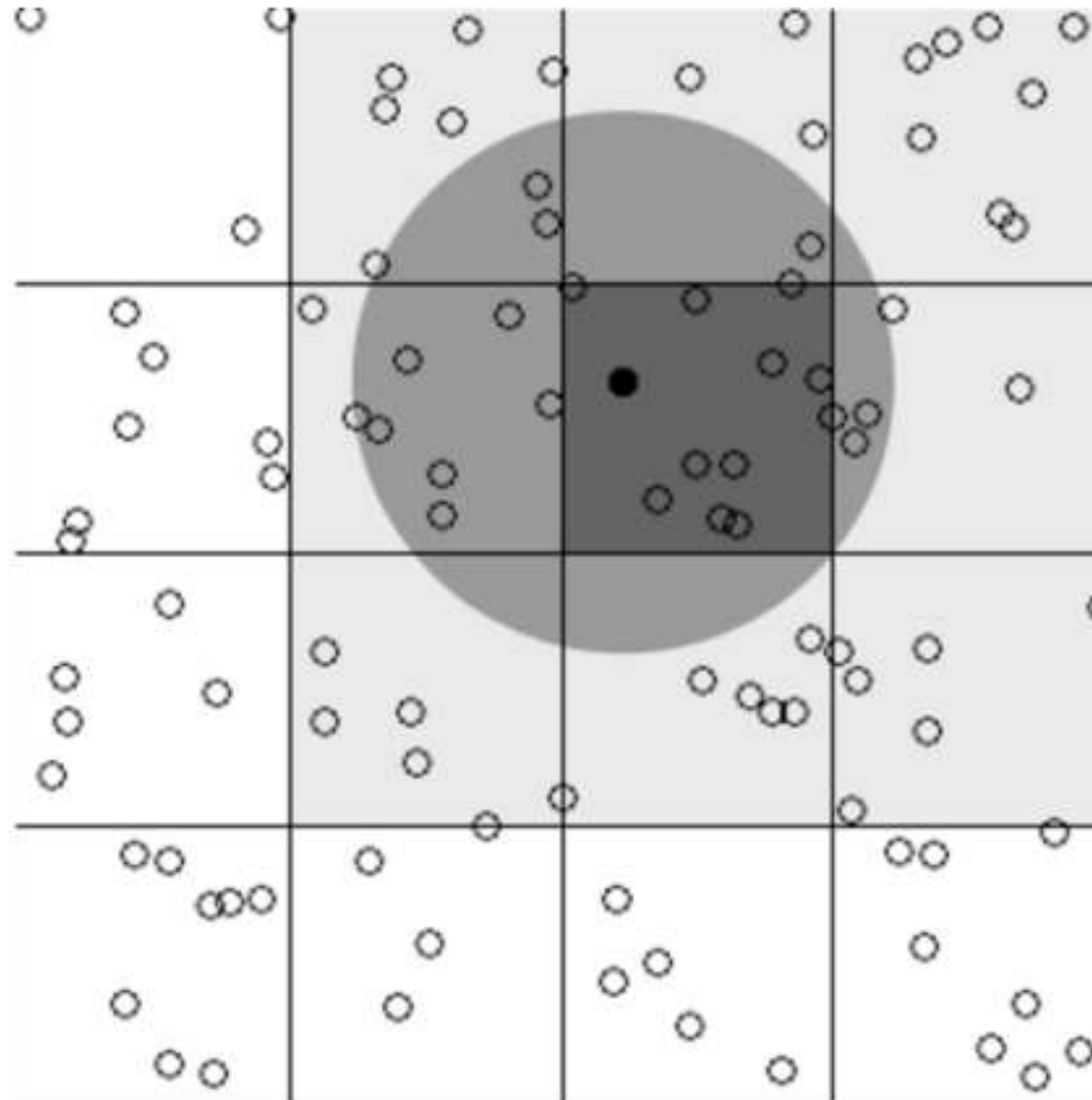
$$E_{elec} = \epsilon_{14} \frac{C q_i q_j}{\epsilon_0 |\vec{r}_{ij}|} \left(1 - \frac{|\vec{r}_{ij}|^2}{R_{off}^2} \right)^2$$

where $|\vec{r}_{ij}| < R_{off}$

$$E_{elec} = 0 \text{ where } |\vec{r}_{ij}| \geq R_{off}$$



Связанный список

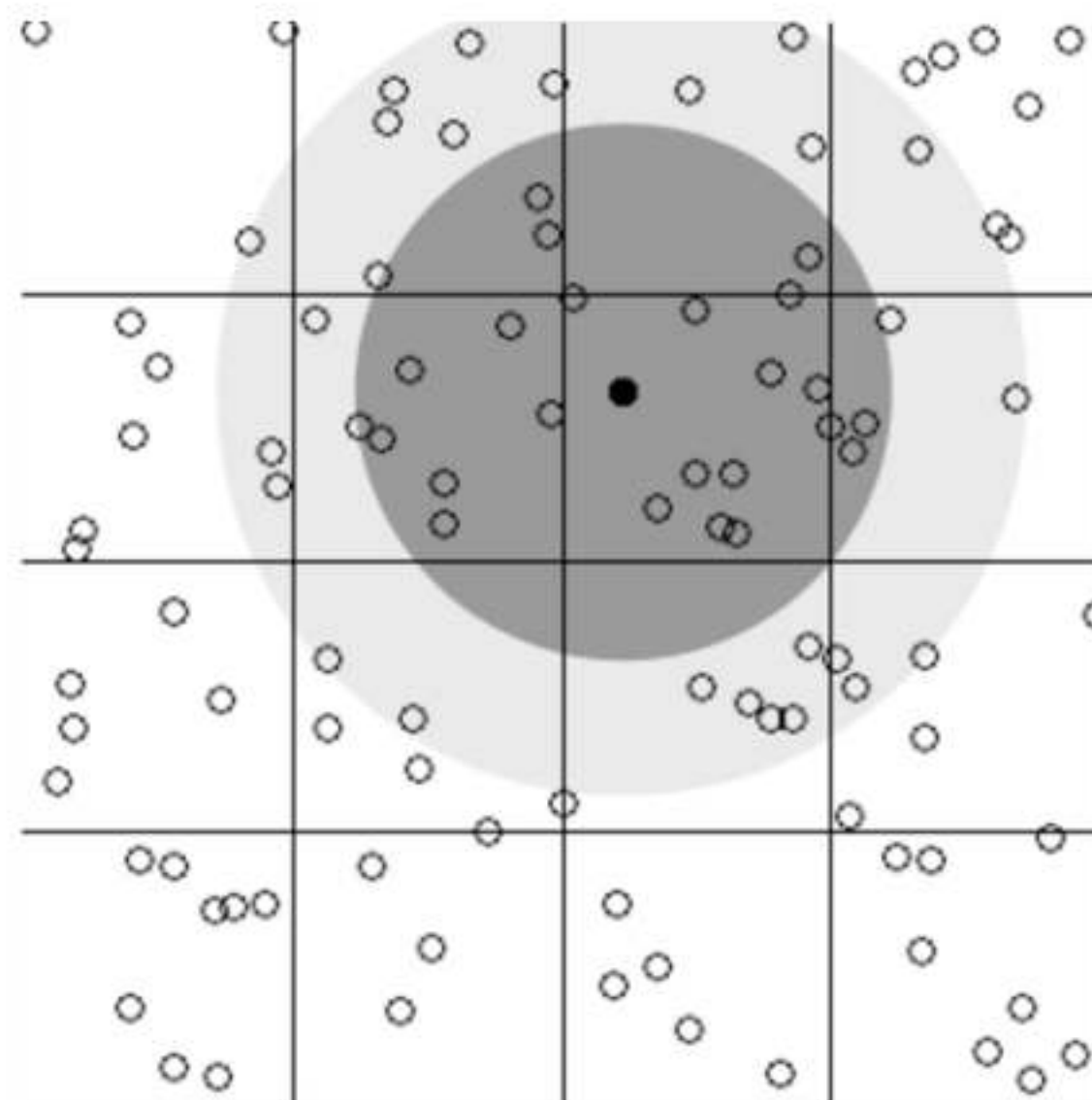


$O(N)$!

Для вычисления взаимодействий мы рассматриваем атомы из той же клетки (темно-серая) и из соседних клеток (светло-серые). Только атомы из обозначенного круга взаимодействуют с данным.

Список Верле

- $O(N^2)$!
- Обновляем лишь изредка, обычно каждые 10-20 шагов



Список взаимодействий содержит все атомы из светло-серой области. Только атомы из темно-серой области взаимодействуют с данным.

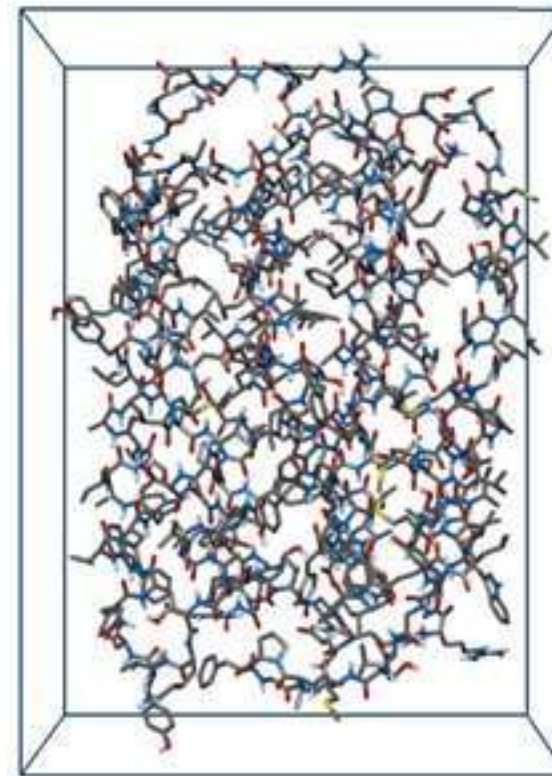
Ограничивающие Объемы

Input: AABB - root axis-aligned bounding box

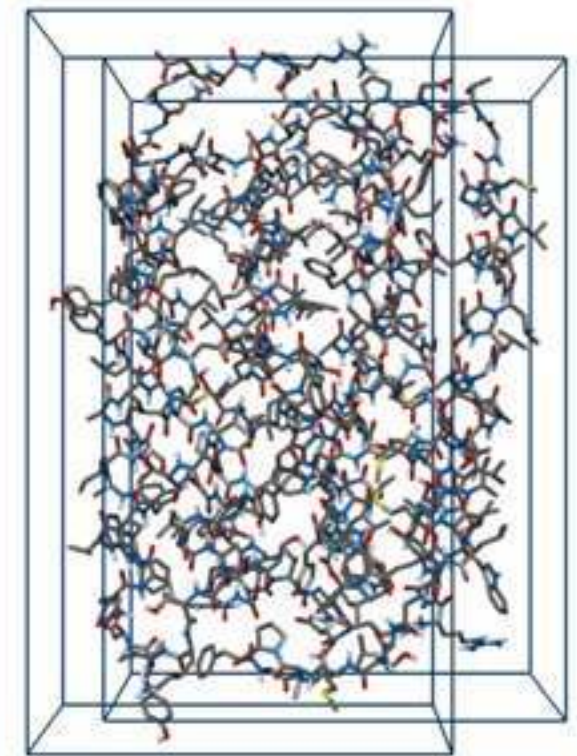
```
if AABB diameter < cutoff
  add all rigid body pairs of AABB
  into interaction list
return
else
  if AABB has children
    restart algorithm with AABB.leftChild as input
    restart algorithm with AABB.rightChild as input
  compare AABB.leftChild and AABB.rightChild
```

comparison loop (AABBa, AABBb):

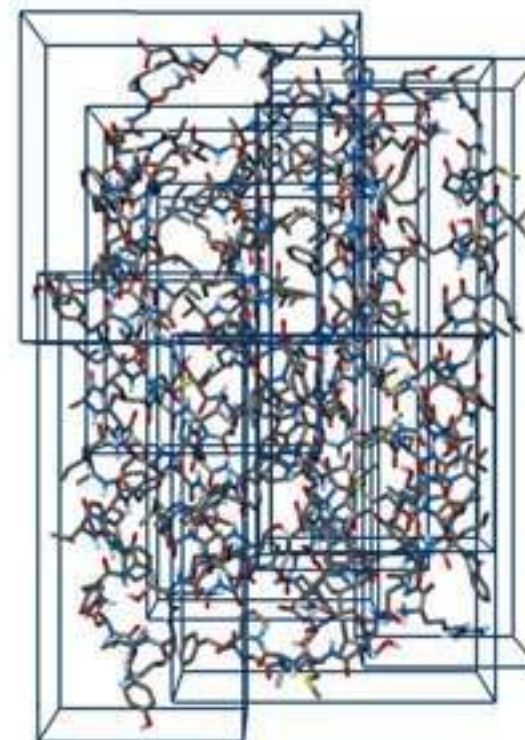
```
if distance between AABBa and AABBb < cutoff
  if AABBa and AABBb have no children
    add corresponding rigid body pair
    into interaction list
  else if diameter of AABBa + AABBb < cutoff
    add all corresponding rigid body pairs into
    interaction list
else
  subdivide boxes further and restart the loop
```



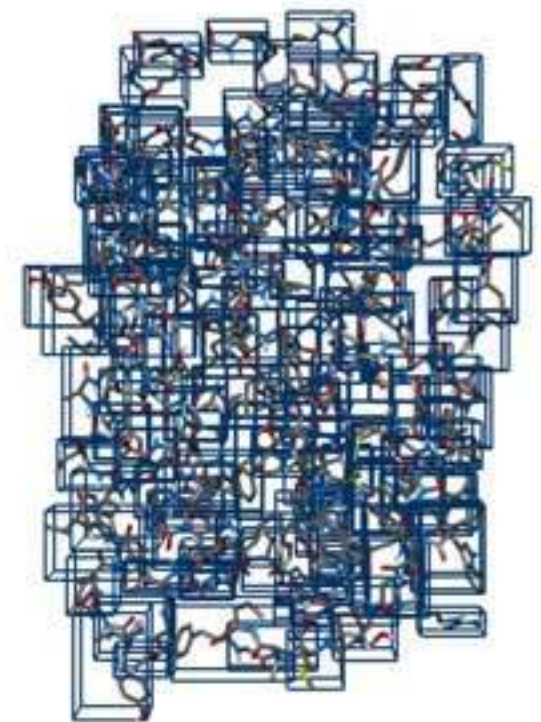
Level 1



Level 2

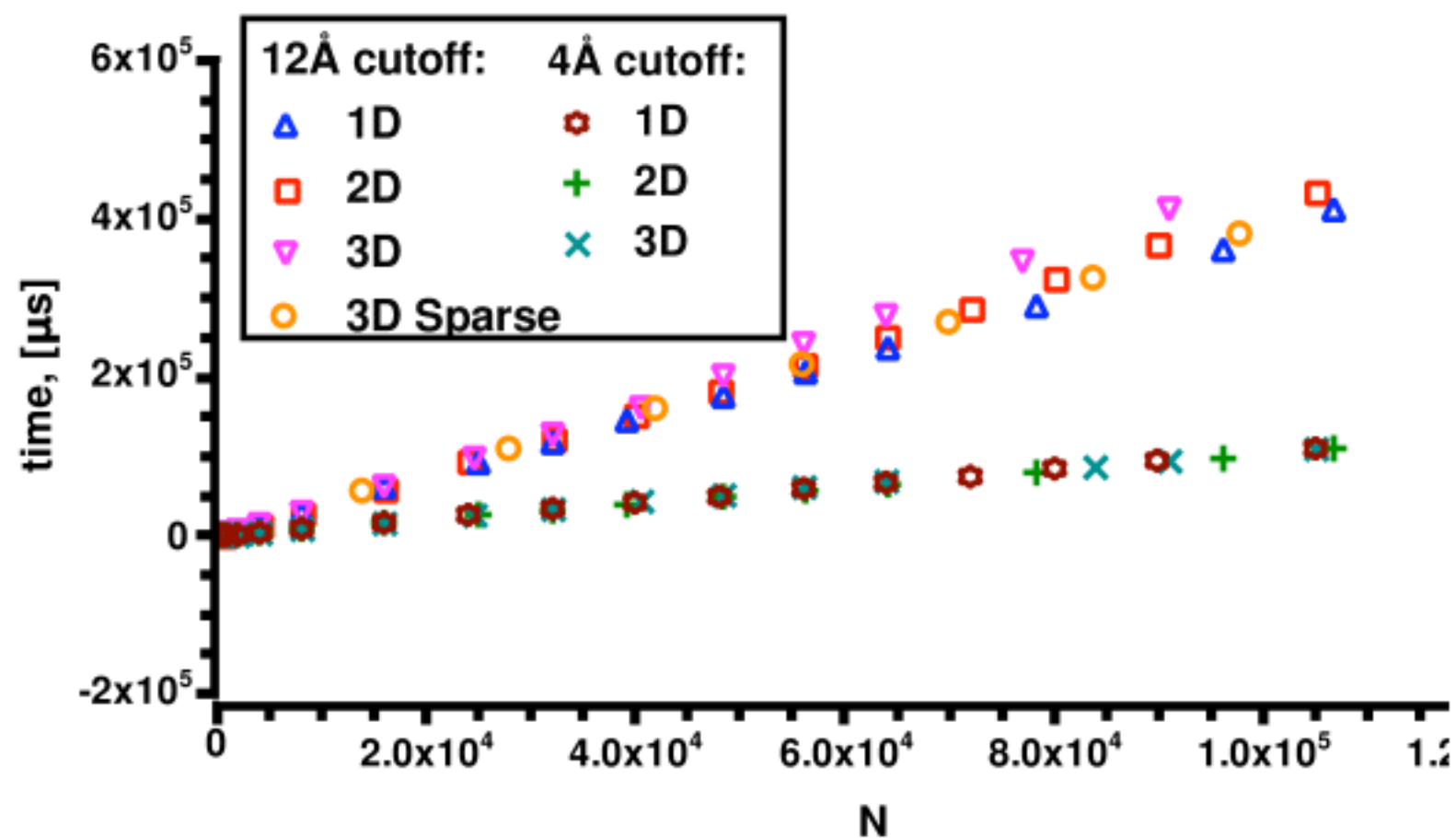
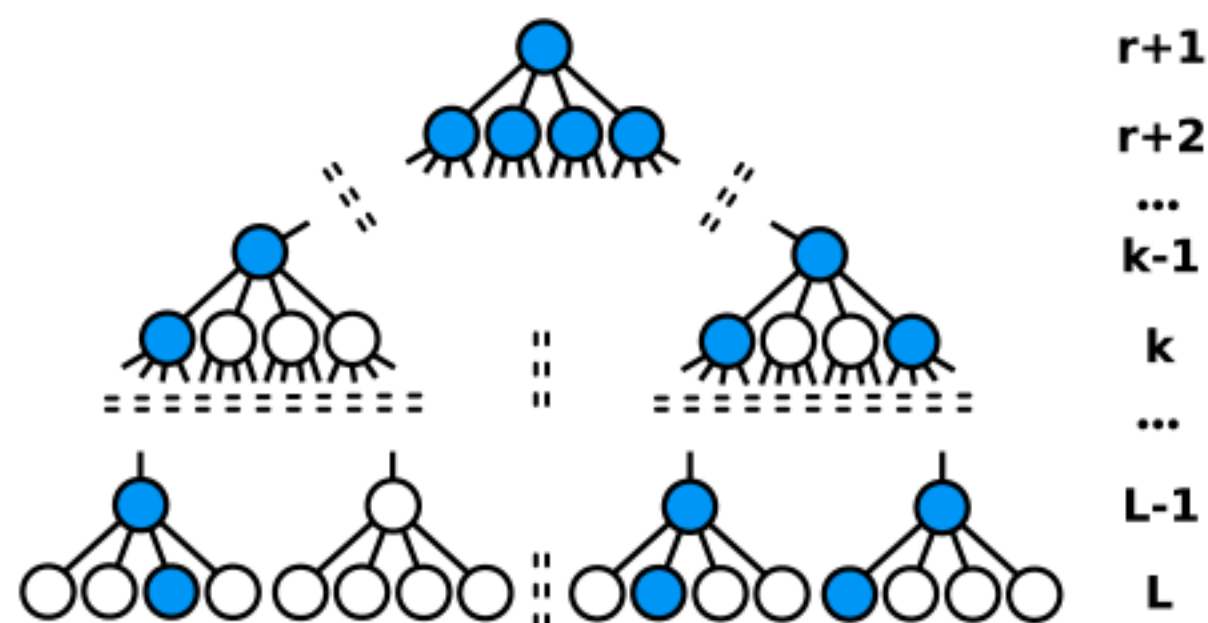


Level 4

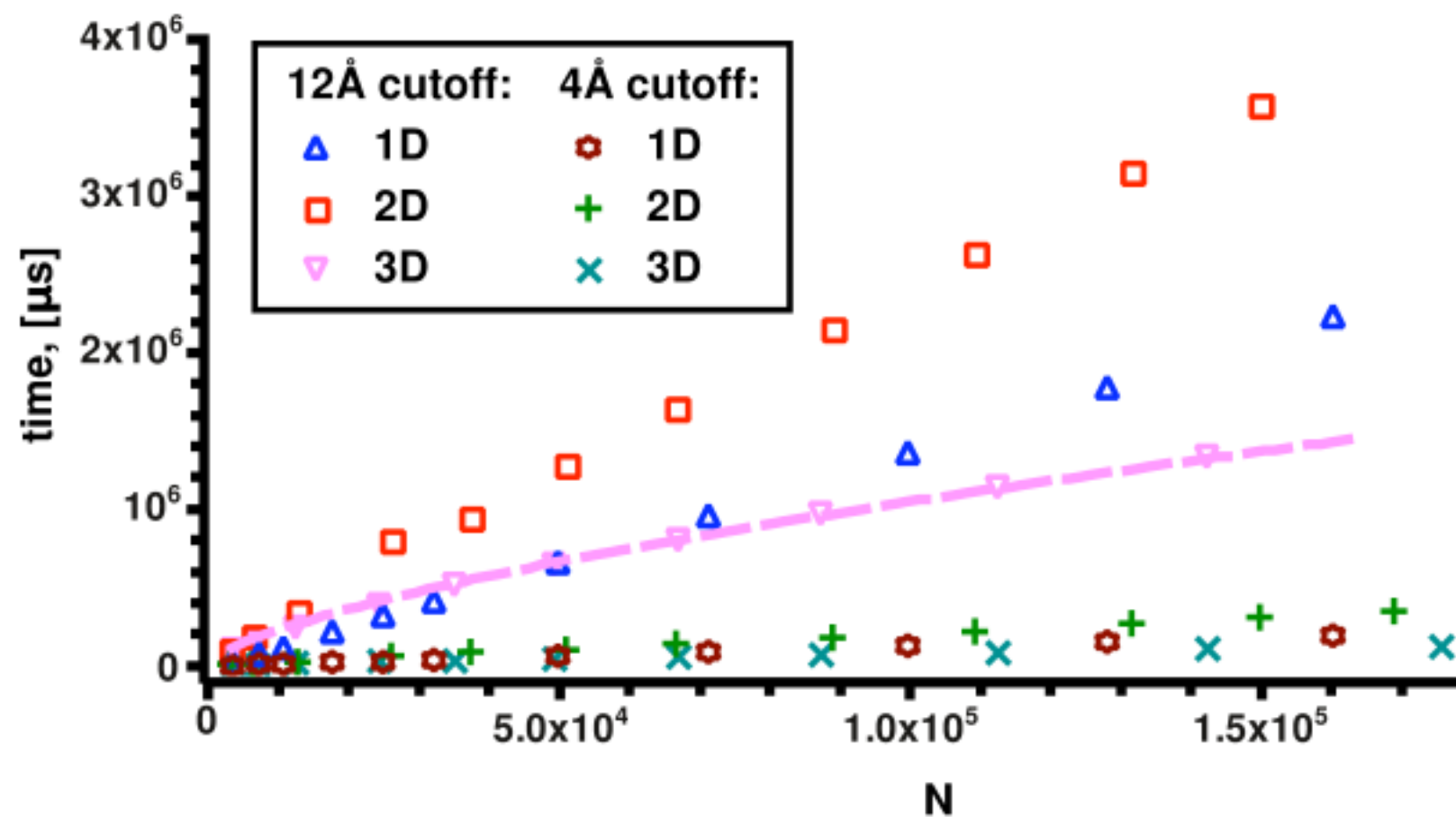
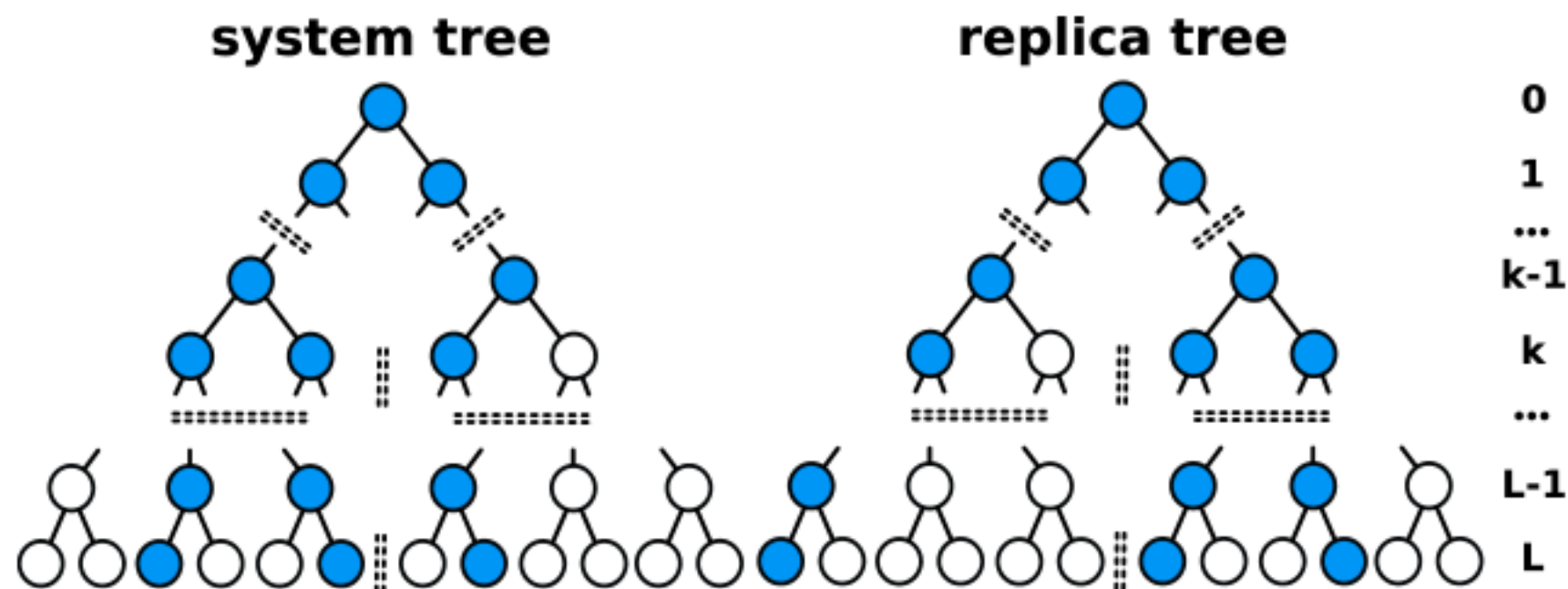


Level 8

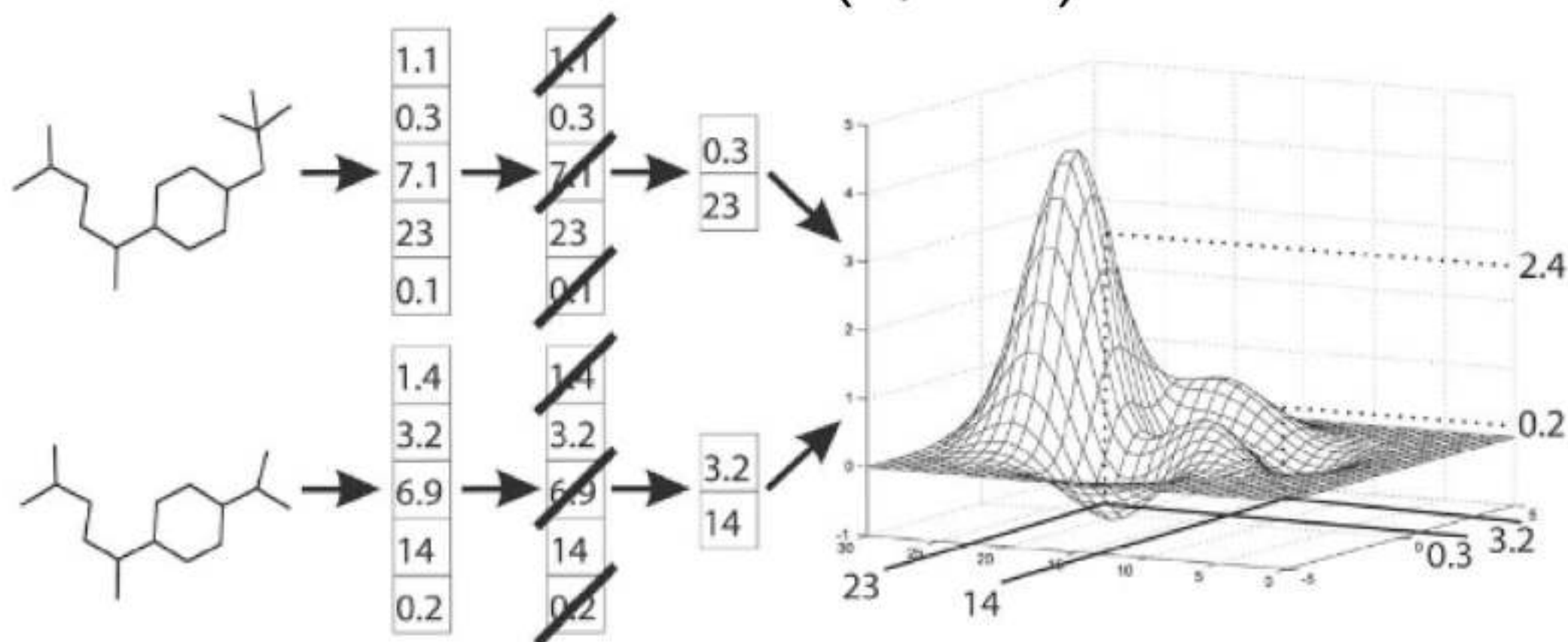
Ограничивающие Объемы



Ограничивающие Объемы



Поиск количественных соотношений структура-свойство (QSAR)



Основные шаги QSAR:

- Молекулярная структура кодируется при помощи численных дескрипторов
- Набор дескрипторов урезается так, чтобы остались наиболее информативные
- Активность выводится как функция выбранных дескрипторов

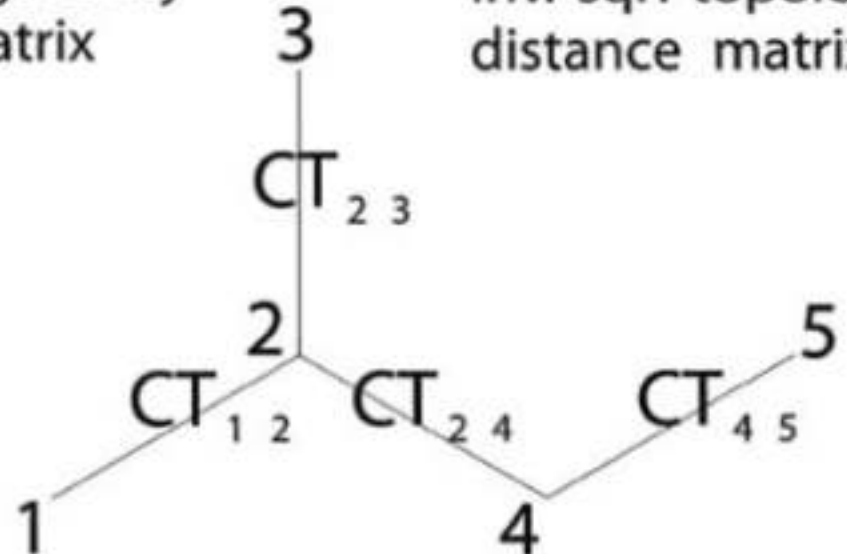
QSAR дескрипторы

- 2D
 - компонентные
 - электростатические и кванто-химические
 - топологические
 - геометрические
 - фрагментарные дескрипторы и “молекулярные отпечатки”
- 3D
 - зависящие от ориентации
 - не зависящие от ориентации
- Сравнение производительности 2D- и 3D-QSAR дескрипторов
 - обычно 3D работают намного лучше чем 2D

Пример: 2D QSAR дескрипторы

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 1 & 1/4 & 1/4 & 1/9 \\ 1 & 0 & 1 & 1 & 1/4 \\ 1/4 & 1 & 0 & 1/4 & 1/9 \\ 1/4 & 1 & 1/4 & 0 & 1 \\ 1/9 & 1/4 & 1/9 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1/4 \\ 1/2 & 3 & 1/2 & 1/2 & 11/9 \\ 1 & 0 & 1 & 1 & 1/4 \\ 10/9 & 1/4 & 10/9 & 2 & 1/4 \\ 1/4 & 1 & 1/4 & 0 & 1 \end{bmatrix}$$

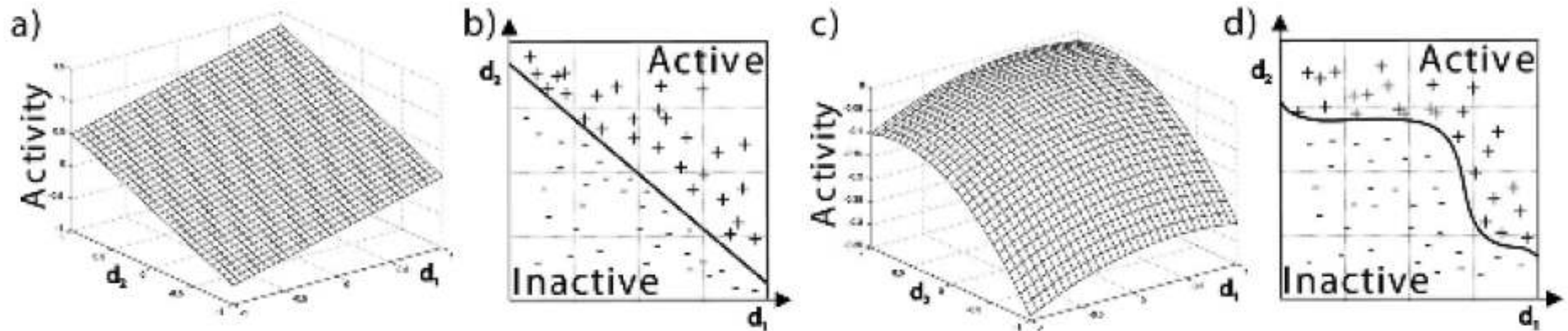
adjacency matrix inv. sqr. topological distance matrix matrix M and CT pairs



$$\begin{aligned}
 CT_{12} &= 0 - 1/2 = -1/2 \\
 CT_{23} &= 1/2 - 0 = +1/2 \\
 CT_{24} &= 1/2 - 1/4 = +1/4 \\
 CT_{45} &= 1/4 - 0 = +1/4 \\
 G_1 &= 1.5 \\
 J_1 &= G_1/4 = 0.375
 \end{aligned}$$

Пример топологических зарядовых индексов Галвеца (Galvez) первого порядка G_1 и J_1 для изопентана. Для определения зарядовых членов CT_{ij} как $M_{ij} - M_{ji}$ используется матричное произведение M_{ij} матрицы соседей с матрицей топологических связей, определенной как квадрат обратных расстояний по связям. Индексы G_k определены как алгебраическая сумма абсолютных значений зарядовых членов для пар атомов, отстоящих друг от друга на k связей. Индексы J_k получаются нормированием G_k на количество связей в молекуле.

QSAR отображение



Методы QSAR отображения:

- a) линейная регрессия, активность - функция двух дескрипторов, d_1 и d_2
- b) бинарная классификация с линейной границей между классами активных (+) и неактивных (-) веществ
- c) нелинейная регрессия
- d) нелинейная бинарная классификация

Литература

- Andrew R. Leach. “Molecular Modelling: Principles and Application”. Prentice Hall 2001.
- D.Allen. “Computer Simulation of Liquids”. Oxford: Clarendon Press 1987.
- Daan Frenkel and Berend Smit. “Understanding Molecular Simulation: From Algorithms to Applications”. Academic Press, Inc. 1996.