

Молекулярно–динамическое моделирование бактериальной рибосомы: методы и особенности

Г.И. Макаров, Т.М. Макарова

28 мая 2018

- ▶ Как готовить систему к моделированию?
- ▶ Какие методы используются?
- ▶ Какова производительность расчетов на СК „Ломоносов“?

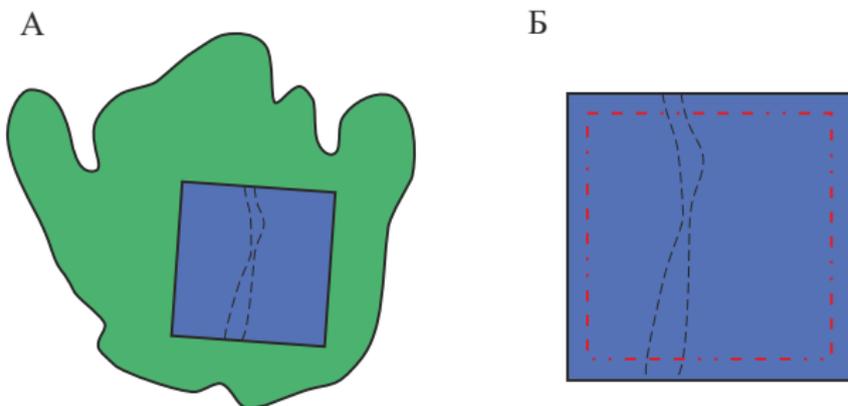
Подготовка расчета целой рибосомы

1. Формирование моделируемой системы, нейтрализация заряда ионами K^+ , построение топологии
2. Оптимизация координат ионов K^+ при неподвижных атомах биополимеров и ионах Mg^{2+} методом наискорейшего спуска
3. Добавление воды и дополнительных ионов Mg^{2+} , K^+ и Cl^-
4. Оптимизация координат воды и ионов методом наискорейшего спуска и короткой молекулярной динамикой при неподвижных атомах биополимера
5. Оптимизация координат всей системы методом наискорейшего спуска и короткой молекулярной динамикой при неподвижных C_α -атомах и атомах фосфора

Вода и ионы

- ▶ Ионы Mg^{2+} , разрешенные в исходной структуре, сохраняются
- ▶ Отрицательный заряд рРНК нейтрализуется ионами K^+ , устанавливаемыми вблизи отрицательно заряженных групп (doi: 10.1371/journal.pone.0038024)
- ▶ Дополнительно прибавляется столько ионов Mg^{2+} , K^+ и Cl^- , чтобы достигнуть в воде концентрации $MgCl_2$ в 7 мМ и KCl в 100 мМ. Это препятствует вымыванию ионов, координированных РНК.
- ▶ Используются ионы K^+ с оптимизированными параметрами (doi: 10.1021/jp8001614). Для ионов K^+ из Parm99SB переоценено взаимодействие с анионными группами (doi: 10.1093/nar/gkr639)
- ▶ Используется вода TIP4P с оптимизированными параметрами (doi: 10.1063/1.1683075), соответствующая используемым ионам K^+

Моделирование фрагмента рибосомы



Для моделирования рибосомных антибиотиков удобно использовать кубический фрагмент большой субъединицы рибосомы с ребром ~ 7 нм, который включает весь рибосомный туннель. Остатки биополимеров на гранях позиционно ограничены.

Это экономит вычислительные ресурсы: 200 нс – 67 часов вычислений на 8 узлах $2 \times \text{Intel Xeon E5630} + 2 \times \text{Tesla X2070}$.

Целая рибосома: 200 нс – 240 часов вычислений на 32 узлах $2 \times \text{Intel Xeon E5630} + 2 \times \text{Tesla X2070}$

Алгоритм расчета траектории

$$t = 0, \mathbf{r}(0) = (\mathbf{r}_1(0), \dots, \mathbf{r}_N(0)), \mathbf{v}(0) = (\mathbf{v}_1(0), \dots, \mathbf{v}_N(0))$$

$$t = \Delta t, \mathbf{r}_i(\Delta t) = \mathbf{r}_i(0) + \mathbf{v}_i(0) \Delta t$$

$$U = U_{\text{связи}} + U_{\text{углы}} + U_{\text{торсионы}} + U_{\text{несобств}} + U_{VdW} + U_{qq}$$

$$\mathbf{F}_i = -\frac{\partial U(\mathbf{r}_1(t), \dots, \mathbf{r}_N(t))}{\partial \mathbf{r}_i}$$

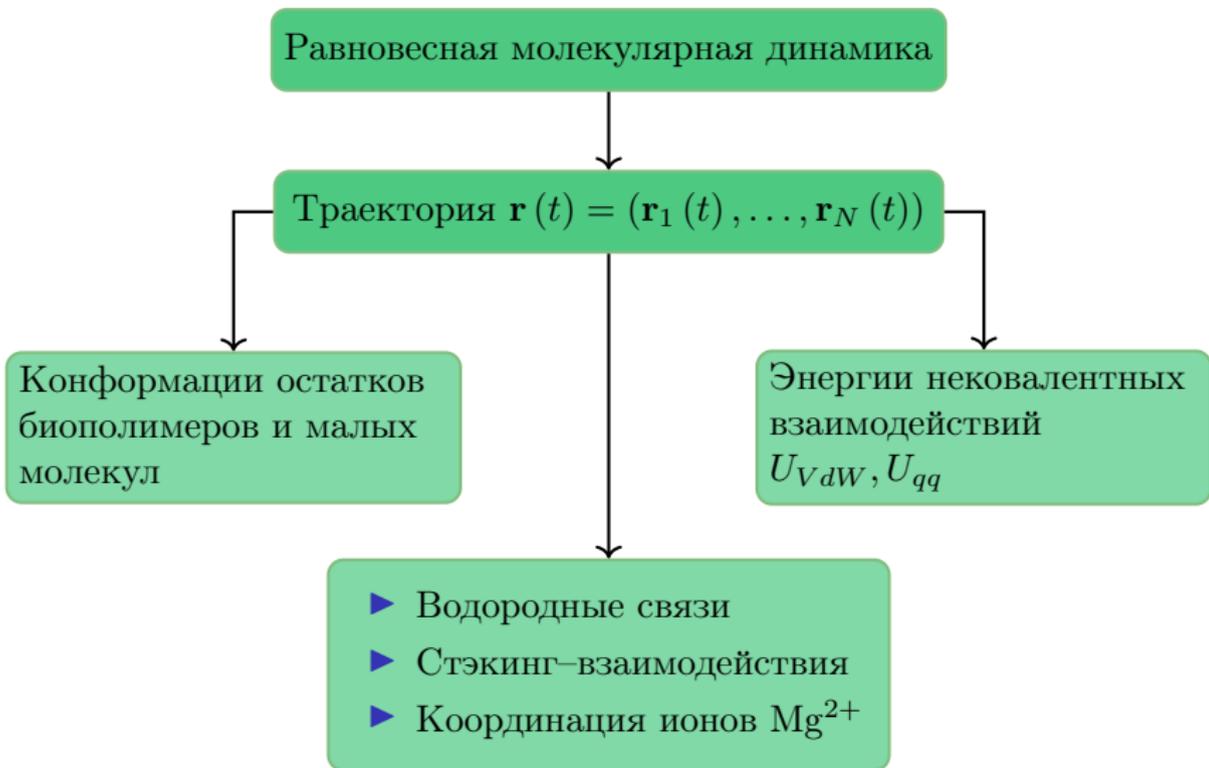
$\mathbf{r}(t), \mathbf{v}(t)$

$$\begin{aligned} \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t) \Delta t + \frac{1}{2} \mathbf{a}_i(t) \Delta t^2 \\ \mathbf{v}_i(t + \Delta t) &= \mathbf{v}_i(t) + \frac{1}{2} (\mathbf{a}_i(t) + \mathbf{a}_i(t + \Delta t)) \Delta t \end{aligned}$$

Условия моделирования

Шаг интегрирования	2 фс
Протяженность траектории	~ 200 нс
Температура	300 К
Термостат	Масштабирования скоростей со стохастическим членом
Баростат	Берендсена
Электростатические взаимодействия	Сеть частиц Эвальда
Дисперсионные взаимодействия	Короткодействующие

Что дает равновесная динамика?



Метадинамика

Моделируется *взвешенный* ансамбль

$$E_{\text{взвеш}}(\Gamma) = E(\Gamma) + V(\mathbf{X}) \quad (1)$$

Взвешивающий потенциал задается как сумма гауссиан

$$V(\mathbf{X}) = V(\mathbf{X}, t) = W \sum_{\substack{i \\ i\tau_G < t}} \exp \left\{ - \sum_j^K \frac{(x_j(\mathbf{r}) - x_j(i\tau_G))^2}{2\sigma_j^2} \right\} \quad (2)$$

$$\mathbf{X}(\mathbf{r}) = (x_1(\mathbf{r}), \dots, x_K(\mathbf{r})) \quad (3)$$

и подбирается автоматически: чем дольше система находится в области фазового пространства с $\mathbf{X} = \mathbf{x}$, тем больше для неё $V(\mathbf{X})$. Задержка системы говорит о попадании в энергетический минимум, так что в пределе

$$V(\mathbf{X}, t) \rightarrow -F(\mathbf{X}) \quad (4)$$

Но можно попасть в „нефизичные“ области фазового пространства.

Уравновешенная метадинамика

Пусть высота гауссиан W уменьшается в зависимости от величины добавленного $V(\mathbf{X}, t)$

$$W = W(\mathbf{X}, t) = w \exp \left\{ -\frac{V(\mathbf{X}, t)}{k_B \Delta T} \right\} \quad (5)$$

Уравнение 33 принимает вид

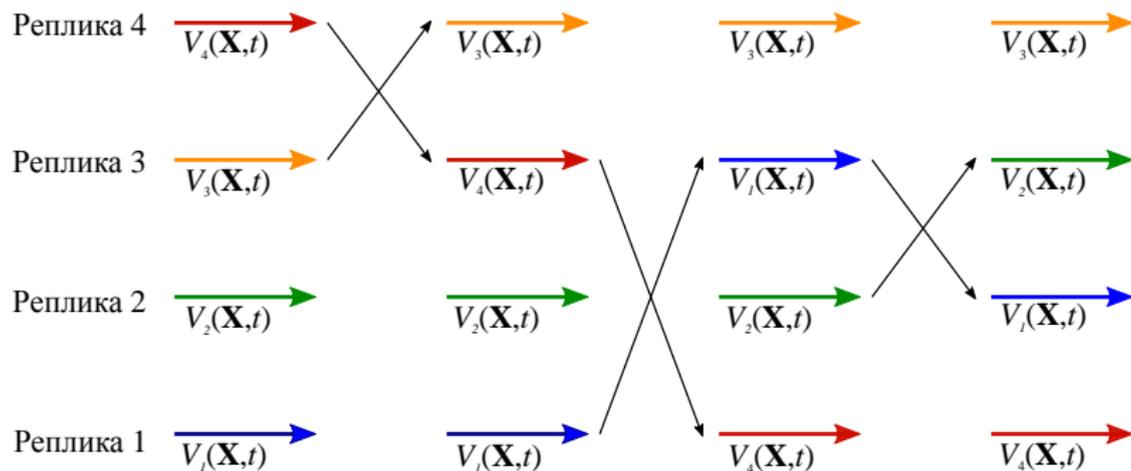
$$V(\mathbf{X}, t) = \sum_{i\tau_G < t} w \exp \left\{ -\frac{V(\mathbf{X}, i\tau_G)}{k_B \Delta T} \right\} \exp \left\{ -\sum_j^K \frac{(x_j(\mathbf{r}) - x_j(i\tau_G))^2}{2\sigma_j^2} \right\} \quad (6)$$

Величина ΔT определяет скорость насыщения $V(\mathbf{X}, t)$. Оценка свободной энергии

$$F(\mathbf{X}) = -\frac{T + \Delta T}{\Delta T} V(\mathbf{X}) \quad (7)$$

Метадинамика с обменом потенциалами

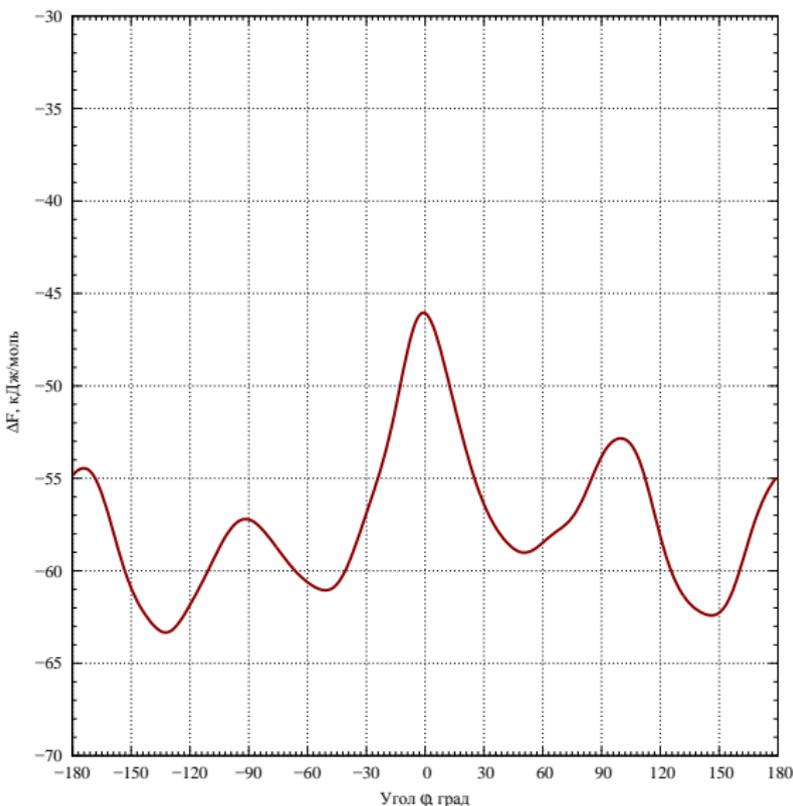
Регулярный обмен $V_i(\mathbf{X}, t)$ между копиями системы



Смысл в том, чтобы оперировать более простыми взвешивающими потенциалами при использовании многих обобщенных переменных.

Что дает метадинамика?

Изучение
множества
возможных
конформаций
системы с оценкой
их свободных
энергий



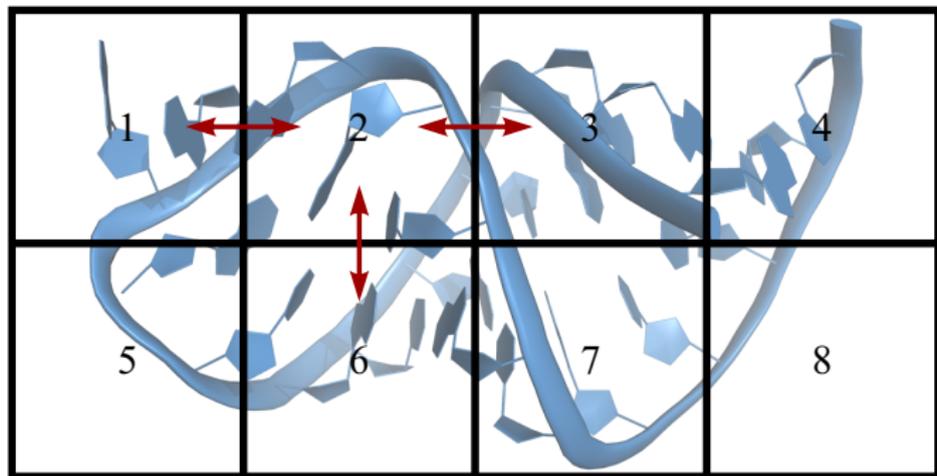
GROMACS

FAST. FLEXIBLE. FREE.



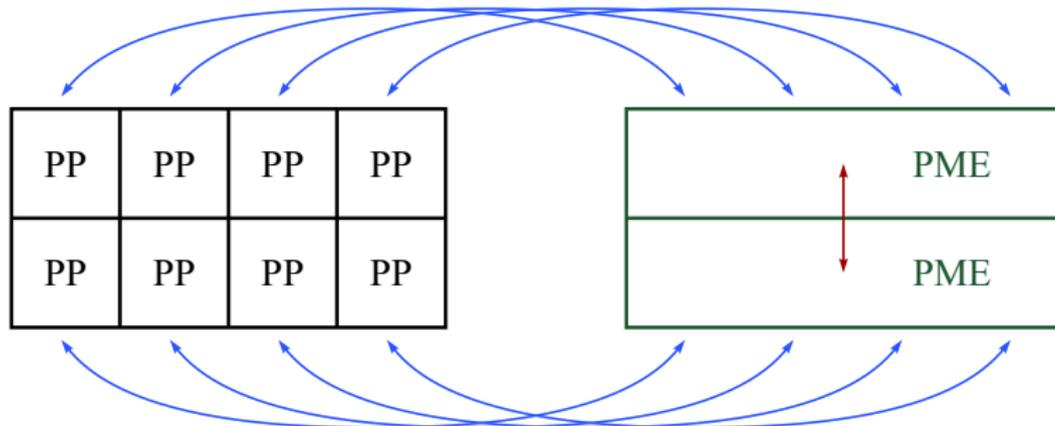
- ▶ GROMACS 5.1.4
- ▶ PLUMED 2.3
- ▶ Parm99SB (биополимеры) + GAFF (малые молекулы)
- ▶ Intel Compilers 13; Intel MPI 5.0.1; Intel MKL 11

Доменное разбиение



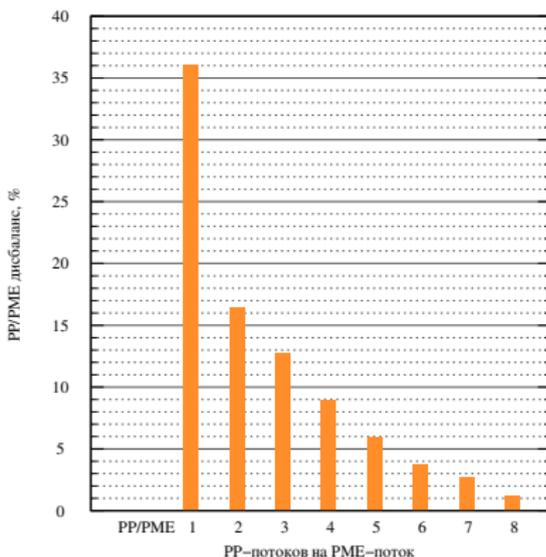
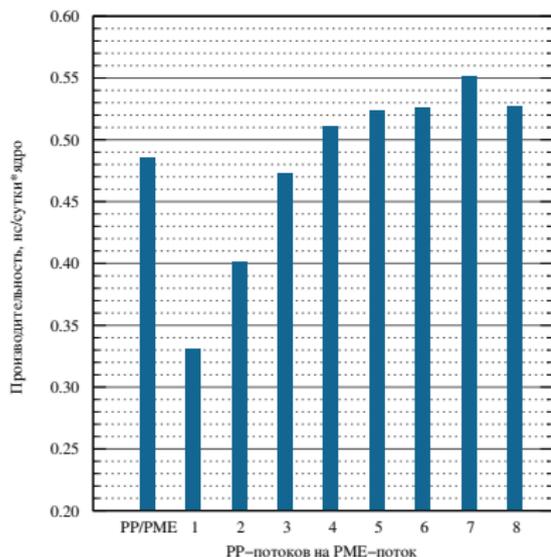
Параллелизм GROMACS основан на разбиении моделируемой системы на части (домены), каждая из которых обрабатывается отдельным потоком. Периодически „смежные“ потоки передают друг другу атомы и обмениваются информацией о координатах частиц и силах.

PP- и PME-потоки



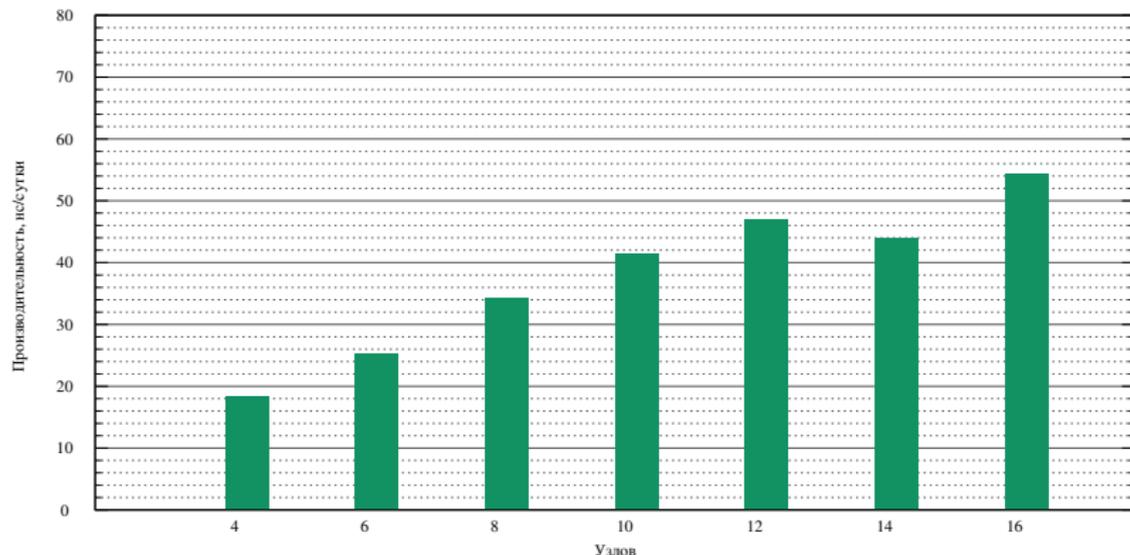
PP-потоки обрабатывают ковалентные взаимодействия и нековалентные взаимодействия между близкими частицами; PME-потоки обрабатывают дальнедействующие нековалентные взаимодействия. PME-потоков нужно меньше, чем PP-потоков.

PP- и PME-потоки: производительность



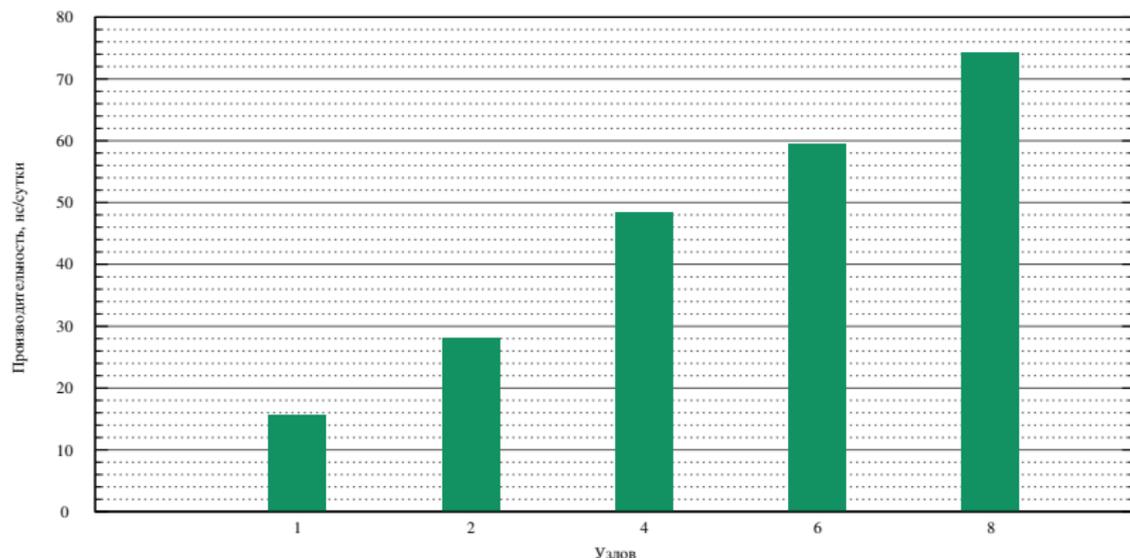
Кубический фрагмент рибосомы+вода T1P4P, 70853 атомов+14644 виртуальных сайтов. Узлы 2×Intel Xeon X5570. 8 MPI-потоков на узел, 1 OpenMP-поток на MPI-поток. Наивысшая производительность достигается при 7 PP-потоках на PME-поток. Чем больше отдельных молекул в системе, тем меньше нужно PP-потоков на PME-поток.

Производительность на CPU-узлах



Кубический фрагмент рибосомы+вода TIP4P, 70853 атомов+14644 виртуальных сайтов. Узлы 2×Intel Xeon X5570. 8 MPI-потоков на узел, 1 OpenMP-поток на MPI-поток. 7 PP-потоков на 1 PME-поток.

Производительность на CPU/GPU-узлах



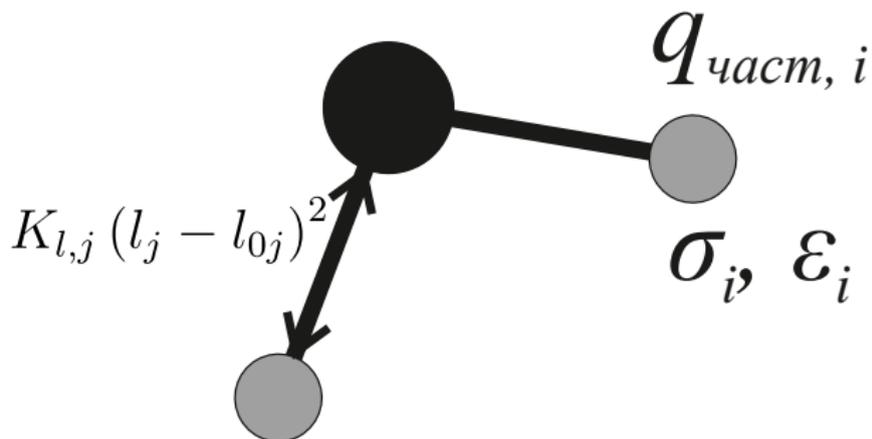
Кубический фрагмент рибосомы+вода TIP4P, 70853 атомов+14644 виртуальных сайтов. Узлы 2×Intel Xeon E5630+2×Tesla X2070. 2 MPI-потока на узел, 4 OpenMP-потоков на MPI-поток. PP- и PME-домены совпадают и считаются одним и тем же MPI-потоком.

Благодарности

- ▶ НИВЦ МГУ за доступ к СК „Ломоносов“
- ▶ А.А. Богданову
- ▶ А.В. Головину
- ▶ Н.В. Сумбатьян и кафедре химии природных соединений Химического Факультета МГУ

Приложение

Молекулярная механика — атом и молекула



Атом — материальная точка с массой, зарядом и Ван-дер-Ваальсовыми параметрами.

Молекула — совокупность связанных атомов. Связи не перестраиваются сами собой.

Потенциальная энергия

$$U = U_{\text{связи}} + U_{\text{углы}} + U_{\text{торсионы}} + U_{\text{несобств}} + U_{VdW} + U_{qq} \quad (8)$$

Ковалентные взаимодействия Нековалентные взаимодействия

$$U_{\text{связи}} = \sum_j K_{l,j} (l_j - l_{0j})^2$$

$$U_{\text{углы}} = \sum_j K_{\theta,j} (\theta_j - \theta_{0j})^2$$

$$U_{\text{торс}} = \sum_j K_{\varphi,j} (1 \pm \cos n_j \varphi_j)^2$$

$$U_{\text{несоб}} = \sum_j K_{\psi,j} (1 - \cos 2\psi_j)^2$$

$$U_{VdW} = \sum_{i < j} \sum_j 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

$$U_{Coul} = \sum_{i < j} \sum_j \frac{q_i q_j}{\varepsilon r_{ij}}$$

$$\mathbf{F}_i = - \frac{\partial U(\mathbf{r}_1(t), \dots, \mathbf{r}_N(t))}{\partial \mathbf{r}_i} \quad (9)$$

Алгоритм расчета траектории

$$t = 0, \mathbf{r}(0) = (\mathbf{r}_1(0), \dots, \mathbf{r}_N(0)), \mathbf{v}(0) = (\mathbf{v}_1(0), \dots, \mathbf{v}_N(0))$$

$$t = \Delta t, \mathbf{r}_i(\Delta t) = \mathbf{r}_i(0) + \mathbf{v}_i(0) \Delta t$$

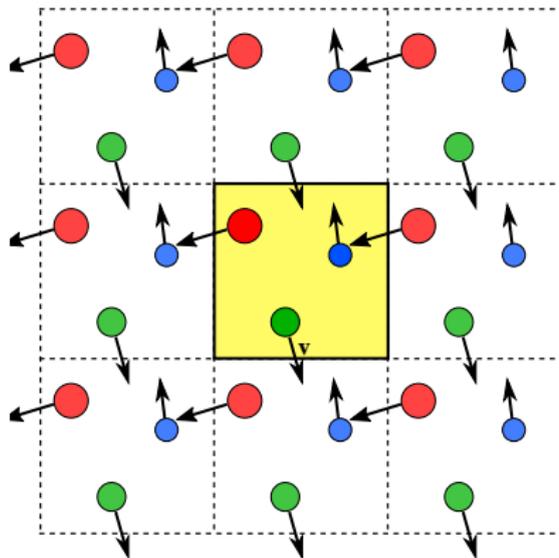
$$U = U_{\text{связи}} + U_{\text{углы}} + U_{\text{торсионы}} + U_{\text{несобств}} + U_{VdW} + U_{qq}$$

$$\mathbf{F}_i = -\frac{\partial U(\mathbf{r}_1(t), \dots, \mathbf{r}_N(t))}{\partial \mathbf{r}_i}$$

$\mathbf{r}(t), \mathbf{v}(t)$

$$\begin{aligned} \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t) \Delta t + \frac{1}{2} \mathbf{a}_i(t) \Delta t^2 \\ \mathbf{v}_i(t + \Delta t) &= \mathbf{v}_i(t) + \frac{1}{2} (\mathbf{a}_i(t) + \mathbf{a}_i(t + \Delta t)) \Delta t \end{aligned}$$

Периодические граничные условия

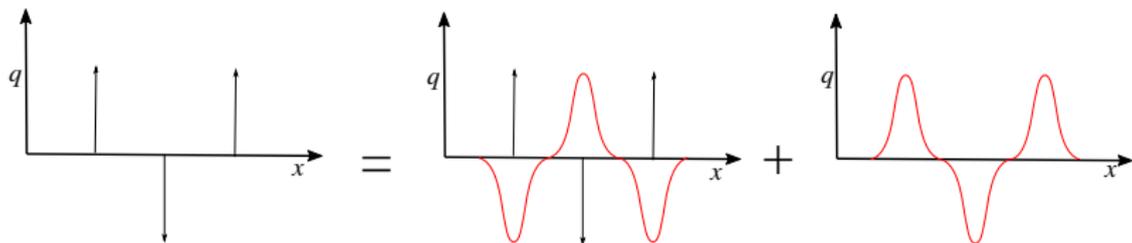


Предполагается, что пространство целиком заполнено копиями ячейки — *периодическими образами*, в которых происходит всё то же, что и в ячейке.

Частица, выходящая из моделируемой ячейки, входит в неё через противоположную грань

Дальнодействующая электростатика — суммирование по Эвальду

Нужно учитывать влияние периодических образов ячейки!



$$U_{Coul} = \frac{1}{2V} \sum_{\mathbf{k} \neq 0} \frac{4\pi}{k^2} |\rho(\mathbf{k})| \exp\left\{-\frac{k^2}{4\alpha}\right\} - \sqrt{\frac{\alpha}{\pi}} \sum_i q_i^2 + \frac{1}{2} \sum_{i \neq j} \frac{q_i q_j \operatorname{erfc}(\sqrt{\alpha r_{ij}})}{r_{ij}} \quad (10)$$

$\sqrt{\frac{2}{\alpha}}$ — ширина экранирующей гауссианы, $\rho(\mathbf{k})$ — Фурье-образ экранирующей плотности заряда $\rho(\mathbf{r})$

Термостатирование

По теореме о равнораспределении:

$$\left\langle p_k \frac{\partial E}{\partial p_k} \right\rangle = \langle p_k \dot{q}_k \rangle = \langle 2K_k \rangle = k_B T \quad (11)$$

где K есть кинетическая энергия, получаем величину температуры T и мгновенной температуры $T(t)$:

$$T = \frac{2 \langle K \rangle}{3Nk_B}, \quad T(t) = \frac{2 \langle K(t) \rangle}{3Nk_B} \quad (12)$$

Для привязки мгновенной температуры к заданной T_{ref} можно использовать термостат масштабирования скоростей. С периодом привязки $\tau \geq \Delta t$ скорости умножаются на фактор S :

$$S = \sqrt{\frac{T_{ref}}{T(t)}} \quad (13)$$

Термостатирование

Лучшие результаты дает термостат масштабирования скоростей с добавочным стохастическим членом (dW). Кинетическую энергию корректируют на величину

$$dK = (K_{ref} - K(t)) \frac{dt}{\tau} + 2\sqrt{\frac{K_{ref}K(t)}{3N}} \frac{dW}{\sqrt{\tau}} \quad (14)$$

Позволяет избежать перераспределения энергии от высокочастотных степеней свободы к низкочастотным — артефакта „летающей льдинки“

Баростатирование

По теореме о равнораспределении:

$$\frac{1}{3} \left\langle \mathbf{r}_i \frac{\partial E}{\partial \mathbf{r}_i} \right\rangle = \frac{1}{3} \langle -\mathbf{r}_i \dot{\mathbf{p}}_i \rangle = -\frac{1}{3} \langle \mathbf{F}_i \mathbf{r}_i \rangle = k_B T \quad (15)$$

Суммируем это равенство по всем N частицам в системе и разделим работу сил на работу внешних и внутренних сил:

$$- N k_B T = \frac{1}{3} \left\langle \sum_{i=1}^N \mathbf{F}_i \mathbf{r}_i \right\rangle = \left\langle \sum_{i=1}^N \mathbf{F}_{i, \text{внешн}} \mathbf{r}_i \right\rangle + \left\langle \sum_{i=1}^N \mathbf{F}_{i, \text{внутр}} \mathbf{r}_i \right\rangle \quad (16)$$

Баростатирование

Работа внешних сил равна произведению давления на объем, усреднение по работе внутренних сил называется внутренним вириалом Ξ :

$$\begin{aligned} -PV &= \frac{1}{3} \left\langle \sum_{i=1}^N \mathbf{F}_{i, \text{внутр}} \mathbf{r}_i \right\rangle \\ \Xi &= \frac{1}{3} \left\langle \sum_{i=1}^N \mathbf{F}_{i, \text{внутр}} \mathbf{r}_i \right\rangle = \frac{1}{3} \left\langle \sum_i \sum_{i < j} \mathbf{F}_{ij} \mathbf{r}_{ji} \right\rangle \end{aligned} \quad (17)$$

Преобразуя, получаем выражение для давления:

$$P = \frac{1}{3V} (2 \langle K \rangle - \Xi) \quad (18)$$

Мгновенное давление приводят к заданному, масштабируя ячейку.

Проблема локального минимума

Энергия системы в момент $t = 0$:

$$E(0) = K(\mathbf{v}(0)) + U(\mathbf{r}(0)) \quad (19)$$

Согласно уравнению 9, система движется по градиенту потенциальной энергии. В момент температурной привязки $t = \tau$, считая, что $E(\tau) = E(0)$:

$$E(\tau) = K(\mathbf{v}(\tau)) + U(\mathbf{r}(\tau)) \quad (20)$$

причем потенциальная энергия понизилась за счет перераспределения в кинетическую:

$$K(\mathbf{v}(\tau)) > K(\mathbf{v}(0)), \quad U(\mathbf{r}(\tau)) < U(\mathbf{r}(0)) \quad (21)$$

После температурной привязки

$$K_{\text{привязанная}}(\mathbf{v}(\tau)) = K(\mathbf{v}(0))$$
$$E_{\text{привязанная}}(\tau) < E(0), \quad U(\mathbf{r}(\tau)) < U(\mathbf{r}(0)) \quad (22)$$

При термостатировании из системы изымается избыток потенциальной энергии, что стабилизирует её в локальном энергетическом минимуме!

Свободная энергия и РМФ

Пусть положение системы в фазовом пространстве описывается вектором

$$\Gamma = (\mathbf{r}_1, \dots, \mathbf{r}_N, \mathbf{v}_1, \dots, \mathbf{v}_N) = (\mathbf{r}, \mathbf{v}) \quad (23)$$

Статистическая сумма Z и свободная энергия Гельмгольца F

$$Z = \int \exp \left\{ -\frac{E(\Gamma)}{k_B T} \right\} d\Gamma, \quad F = -k_B T \ln Z \quad (24)$$

Вероятность встретить в ансамбле систему, для которой переменная $\mathbf{X} = \mathbf{x}$

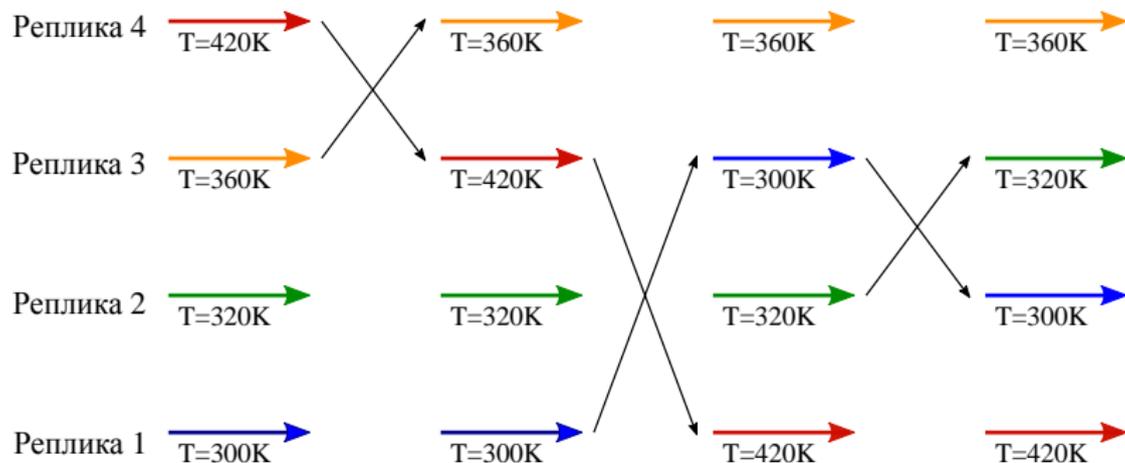
$$P(\mathbf{x}) = \frac{1}{Z} \int \exp \left\{ -\frac{E(\Gamma)}{k_B T} \right\} \delta(\mathbf{X} - \mathbf{x}) d\Gamma = \frac{Z(\mathbf{x})}{Z} \quad (25)$$

откуда после преобразований получаем

$$F(\mathbf{x}) - F = -k_B T \ln P(\mathbf{x}) \rightarrow \Delta F(\mathbf{x}) = -k_B T \ln N(\mathbf{x}) \quad (26)$$

МД с обменом репликами

Регулярный обмен температур между копиями системы



Критерий Метрополиса — вероятность обмена репликами:

$$w(\mathbf{X} \rightarrow \mathbf{X}') = w(\mathbf{x}_m^{[i(m)]} | \mathbf{x}_n^{[j(n)]}) = \begin{cases} 1, & \Delta \leq 0 \\ e^{-\Delta}, & \Delta > 0 \end{cases} \quad (27)$$

$$\Delta = \left(\frac{1}{k_B T_n} - \frac{1}{k_B T_m} \right) \left(U(\mathbf{r}^{[i(m)]}) - U(\mathbf{r}^{[j(n)]}) \right) \quad (28)$$

Зонтичная выборка

При моделировании к энергии системы добавляется *взвешивающий потенциал* $V(\mathbf{X})$, частично компенсирующий ландшафт свободной энергии $F(\mathbf{X})$

$$E_{\text{взвеш}}(\Gamma) = E(\Gamma) + V(\mathbf{X}) \quad (29)$$

Используя уравнение 25, получают свободную энергию

$$\Delta F(\mathbf{x}) = -k_B T \ln N_{\text{взвеш}}(\mathbf{x}) - V(\mathbf{x}) + A \quad (30)$$

$$A = -k_B T \ln \left\langle \exp \left\{ -\frac{V(\mathbf{X})}{k_B T} \right\} \right\rangle \quad (31)$$

Если $V(x)$ покрывает всю интересующую область значений \mathbf{X} , то A можно отбросить. Результаты расчетов с несколькими разными $V_i(\mathbf{X})$ сводят воедино, вычисляя A_i с помощью алгоритма WHAM (Weighted Histogram Analysis Method). Очевидно, наилучший $V(\mathbf{X})$ таков, что

$$V(\mathbf{X}) = -F(\mathbf{X}) \quad (32)$$

Взвешивающий потенциал задается как сумма гауссиан

$$V(\mathbf{X}) = V(\mathbf{X}, t) = W \sum_{i: i\tau_G < t} \exp \left\{ - \sum_j^K \frac{(x_j(\mathbf{r}) - x_j(i\tau_G))^2}{2\sigma_j^2} \right\} \quad (33)$$

$$\mathbf{X}(\mathbf{r}) = (x_1(\mathbf{r}), \dots, x_K(\mathbf{r})) \quad (34)$$

и подбирается автоматически: чем дальше система находится в области фазового пространства с $\mathbf{X} = \mathbf{x}$, тем больше для неё $V(\mathbf{X})$. Задержка системы говорит о попадании в энергетический минимум, так что в пределе

$$V(\mathbf{X}, t) \rightarrow -F(\mathbf{X}) \quad (35)$$

Но можно попасть в „нефизичные“ области фазового пространства.

Уравновешенная метадинамика

Пусть высота гауссиан W уменьшается в зависимости от величины добавленного $V(\mathbf{X}, t)$

$$W = W(\mathbf{X}, t) = w \exp \left\{ -\frac{V(\mathbf{X}, t)}{k_B \Delta T} \right\} \quad (36)$$

Уравнение 33 принимает вид

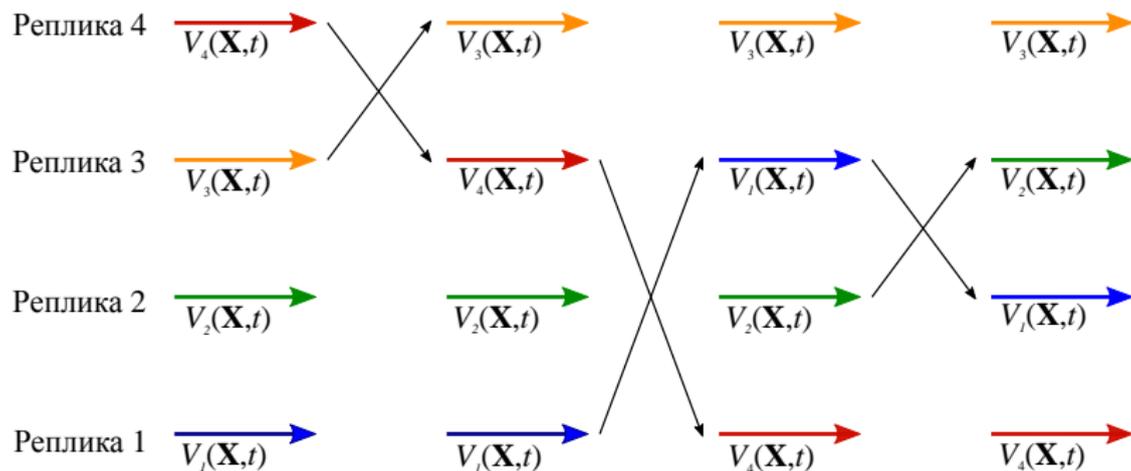
$$V(\mathbf{X}, t) = \sum_{i\tau_G < t} w \exp \left\{ -\frac{V(\mathbf{X}, i\tau_G)}{k_B \Delta T} \right\} \exp \left\{ -\sum_j^K \frac{(x_j(\mathbf{r}) - x_j(i\tau_G))^2}{2\sigma_j^2} \right\} \quad (37)$$

Величина ΔT определяет скорость насыщения $V(\mathbf{X}, t)$. Оценка свободной энергии

$$F(\mathbf{X}) = -\frac{T + \Delta T}{\Delta T} V(\mathbf{X}) \quad (38)$$

Метадинамика с обменом потенциалами

Регулярный обмен $V_i(\mathbf{X}, t)$ между копиями системы



Смысл в том, чтобы оперировать более простыми взвешивающими потенциалами при использовании многих обобщенных переменных.

Управляемая динамика

Систему тянут по исследуемой области фазового пространства, используя модифицированный гамильтониан

$$H_\lambda(t) = H + V_\lambda(t), \quad V_\lambda(t) = \frac{k}{2}(x - \lambda(t))^2 \quad (39)$$

и оценивают работу по перемещению

$$\frac{dW}{dt} = \frac{\partial H_\lambda(t)}{\partial t} = \frac{\partial V_\lambda(t)}{\partial t} = -k(x - \lambda(t)) \frac{\partial \lambda(t)}{\partial t} = -vk(x - \lambda(t)) \quad (40)$$

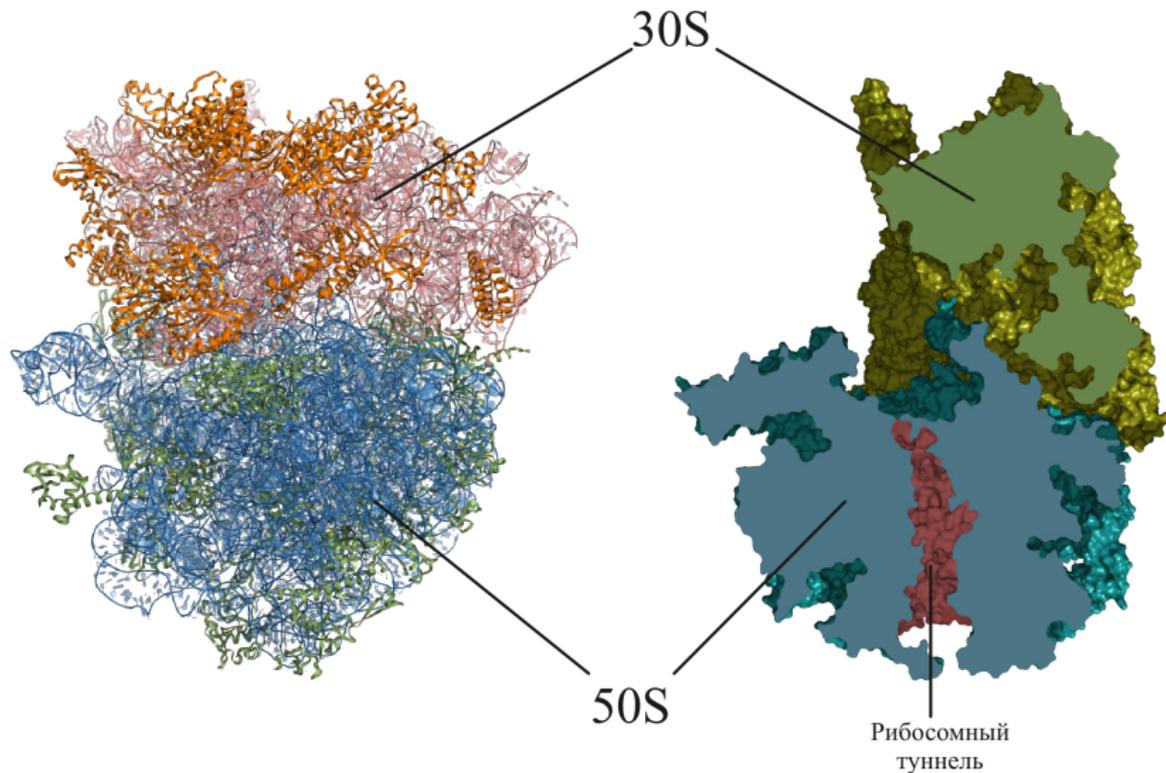
$$W = \int_0^t dW = \int_0^t \frac{\partial H_\lambda(t')}{\partial t'} dt' = \int_0^t -vk(x - x_0 - vt') dt' \quad (41)$$

По равенству Ярынского

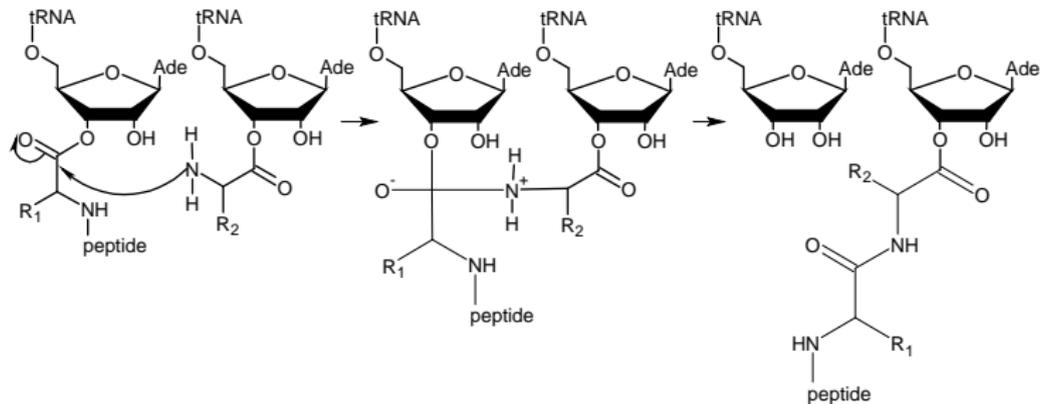
$$\left\langle \exp \left\{ -\frac{W}{k_B T} \right\} \right\rangle = \exp \left\{ -\frac{\Delta F_{1 \rightarrow 2}}{k_B T} \right\} \quad (42)$$

Рибосома

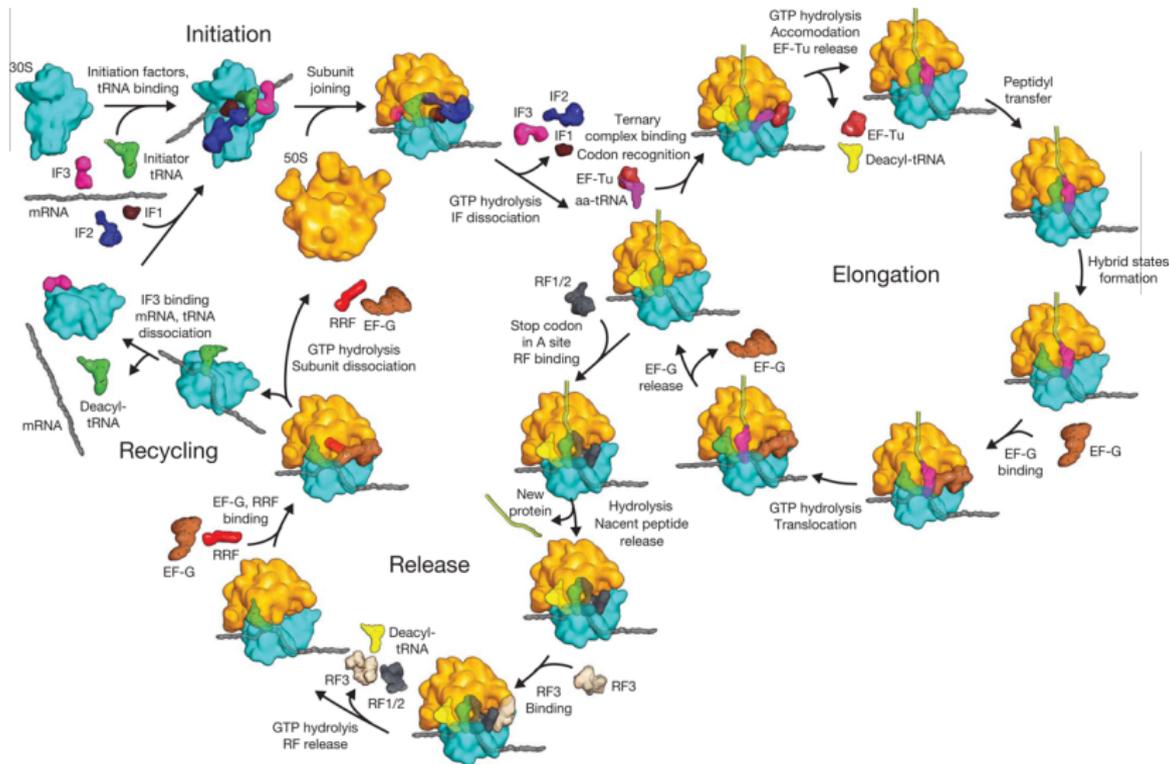
Строение рибосомы



Пептидилтрансферазная реакция (ПТР)



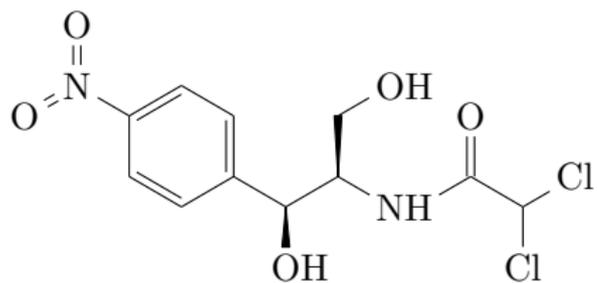
Цикл трансляції у прокаріот



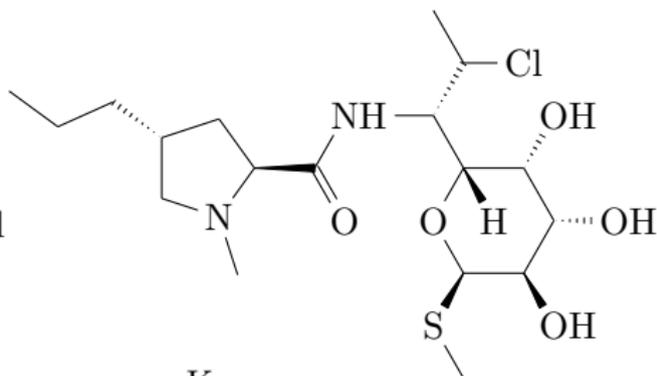
Рибосомные антибиотики

Семейство	Мишень	Механизм действия
Аминогликозиды	30S	Нарушение распознавания тРНК
Тетрациклины	30S	Подавление связывания аминоацил-тРНК
Оксазолидиноны	50S	Подавление ПТР
Амфениколы	50S	Подавление ПТР
Плевомутилины	50S	Подавление ПТР
Линкозамиды	50S	Подавление ПТР
Стрептограмины	50S	Подавление ПТР
Макролиды	50S	Подавление ПТР
Фузидиевая кислота	EF-G	Подавление диссоциации EF-G от рибосомы

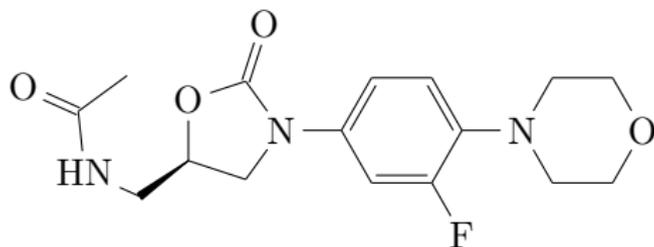
Туннельные антибиотики



Хлорамфеникол



Клиндамицин



Линезолид

