

# Введение в организацию белковой структуры и молекулярное моделирование

- Основы организации белковой структуры
- Молекулярная Динамика (MD)
- Броуновская Динамика (BD)

Сентябрь, 2006

Научный Университет Токио, Тадаси  
Андо

# Проблематика

самоорганизации

структуры белка

“Предсказание трехмерной структуры белка по его аминокислотной последовательности.”

“Каким образом белковая молекула самостоятельно организуется в свою третичную структуру?”



С середины прошлого столетия на этот вопрос еще не дан ответ

# свои третичные структуры спонтанно

В растворах белковые молекулы самостоятельно организуются в трехмерные формы



Конечная трехмерная форма белка зависит от уровня минимальной свободной энергии системы «белок – растворитель». (Постулат Анфинсена)



Если мы сможем точно вычислить энергию системы, тогда можно установить трехмерную структуру белка!

# Парадокс Левинтала

Можем предположить, что для каждой аминокислоты существует три варианта ее состояния в белке ( $\alpha$ -спираль,  $\beta$ -слой и клубок). Если белок состоит из 100 аминокислот (аминокислотных остатков), тогда общее количество вариантов состояний будет:

$$3^{100} = 515377520732011331036461129765621272702107522001 \\ \doteq 5 \times 10^{47}.$$

Если требуется только 100 пс ( $10^{-10}$  с) для перехода из одного состояния в другое, тогда на перебор всех вариантов уйдет

$$5 \times 10^{47} \times 10^{-10} \text{ с} \doteq 1.6 \times 10^{30} \text{ лет.}$$

А, сворачивание аминокислотной последовательности в третичную структуру происходит за время, измеряемое миллисекундами. Таким образом, белки самоорганизуются не через случайный подбор вариантов, а используют более изощренный путь.

**Мы хотим проследить процесс сворачивания белка в третичную форму, используя технологию молекулярного моделирования.**

# Почему “Самоорганизация структуры белка” так важна?

- Белки играют важные роли в живых организмах.
- Некоторые белки напрямую связаны с заболеваниями. Информация о структуре белка необходима, чтобы объяснить и определить его функцию, а также, чтобы создать молекулу, которая бы связывалась с этим белком, если это необходимо в стратегии лечения.
- На сегодня, полностью расшифрованы геномы (полные наборы генов) разных организмов и мы понимаем, что функции многих генов нам до сих пор неизвестны, а часть связана с разными заболеваниями.
- Таким образом, знание процесса самоорганизации белка поможет нам исследовать функции этих генов, и создать эффективные лекарства против связанных с ними заболеваний.
- В добавление к этому, это понимание покажет нам путь к созданию белков, у которых будут совсем новые функции, например, как новые наномшины.

# Почему проблема

## “Самоорганизации структуры белка” так сложна?

С точки зрения компьютерного моделирования,

1. Сложно моделировать весь процесс самоорганизации белковой молекулы на уровне атомов, даже используя самые современные компьютеры.
2. Пока не ясно, достаточно ли точность параметров и значений функций энергии для

моделирования процесса самоорганизации белка и ДНК. ... , позвольте мне пересказать диалог с Фрэнсисом в 1975 (который получил Нобелевскую премию за открытие структуры ДНК). Крик (Spick) утверждал, что «очень сложно предположить научную проблему, которая не будет решена за следующие 20 лет ... кроме модели работы мозга и процесса самоорганизации белка». Хотя Крик был более заинтересован в функциях мозга, он отметил обе проблемы, потому что они являются сложными из-за того, что в них участвует масса взаимодействий в трехмерном пространстве. (Levitt M, “Through the breach.” *Curr. Opin. Struct. Biol.* 1996, 1, 193-194)

# Молекулярная Динамика (MD)

В модели молекулярной динамики, мы имитируем движение атомов как функцию от времени на основе **уравнений движений Ньютона**. Уравнение для системы из  $N$  атомов выглядит так

$$m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = \mathbf{F}_i(t), \quad (i = 1, 2, \dots, N). \quad (1)$$

Где,  $\mathbf{r}_i$  и  $m_i$  это положение и масса атома  $i$ , а  $\mathbf{F}_i(t)$  это функция силы на атом  $i$  по времени  $t$ .  $\mathbf{F}_i(t)$  описывается так:

$$\mathbf{F}_i = -\nabla_i V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N), \quad (2)$$

где  $V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$  потенциальная энергия системы, которая зависит от положения  $N$  атомов в системе.  $\nabla_i$  это

$$\nabla_i = \mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} + \mathbf{k} \frac{\partial}{\partial z} \quad (3)$$

# Интегрирование с использованием метода конечной разности

Положения в моменты времени  $(t + \Delta t)$  и  $(t - \Delta t)$  могут быть выражены через формулы Тэйлора для момента времени  $t$ ,

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i(t)\Delta t + \frac{1}{2}\ddot{\mathbf{r}}_i(t)\Delta t^2 + \frac{1}{6}\dddot{\mathbf{r}}_i(t)\Delta t^3 + O(\Delta t^4), \quad (4a)$$

$$\mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) - \dot{\mathbf{r}}_i(t)\Delta t + \frac{1}{2}\ddot{\mathbf{r}}_i(t)\Delta t^2 - \frac{1}{6}\dddot{\mathbf{r}}_i(t)\Delta t^3 + O(\Delta t^4) \quad (4b)$$

Сумма двух выражений:

$$\mathbf{r}_i(t + \Delta t) + \mathbf{r}_i(t - \Delta t) = 2\mathbf{r}_i(t) + \ddot{\mathbf{r}}_i(t)\Delta t^2 + O(\Delta t^4) \quad (5)$$

Используя формулу (1), получаем:

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{1}{m_i}\mathbf{F}_i(t)\Delta t^2 + O(\Delta t^4) \quad (6)$$

Мы должны последовательно рассчитывать по формуле (6), чтобы получить траектории атомов в системе. (алгоритм Верлета).



процессе

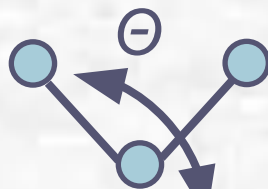
# самоорганизация и белковой молекулы

- Электростатические взаимодействия
- Силы Ван дер Ваальса
- Водородные связи
- Гидрофобные взаимодействия  
(Гидрофобные молекулы в водном растворе соединяются друг с другом, так как молекулы воды их отталкивают. Это как разделение в слое масло/вода. Присутствие воды для этих взаимодействий является важным фактором.)

# Функции от составляющих сил, рассчитываемые в Молекулярном Моделировании



Энергия сжатия-  
растяжения связи



Угловое

взаимодействие



Межплоскостное  
взаимодействие

$$V_{\text{total}} = \sum_{\text{bonds}} K_b (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi - \gamma)]$$

$$+ \sum_{\text{Hbonds}} \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right)$$

Водородная

связь



$$+ \sum_{\text{van der Waals } i, j \text{ pairs}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{\text{electrostatic } i, j \text{ pairs}} \frac{q_i q_j}{\epsilon r_{ij}}$$

Сила Ван дер

Ваальса



Электростатиче

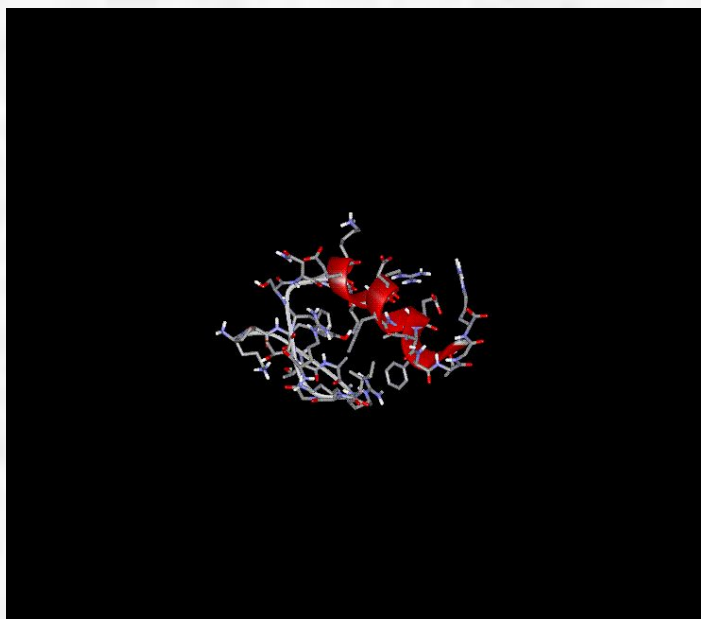
ская сила



**Самая  
тяжелая  
расчетная  
часть.**

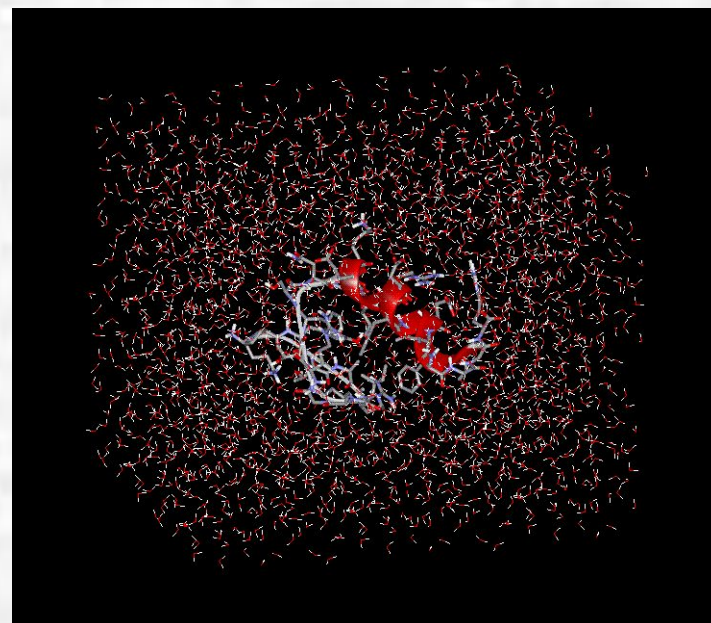
# Система для Моделирования с помощью Молекулярной Динамики

Без молекул воды



Кол-во атомов:  
304

С молекулами воды



Кол-во атомов: 304 +  
**7,377** = 7,681

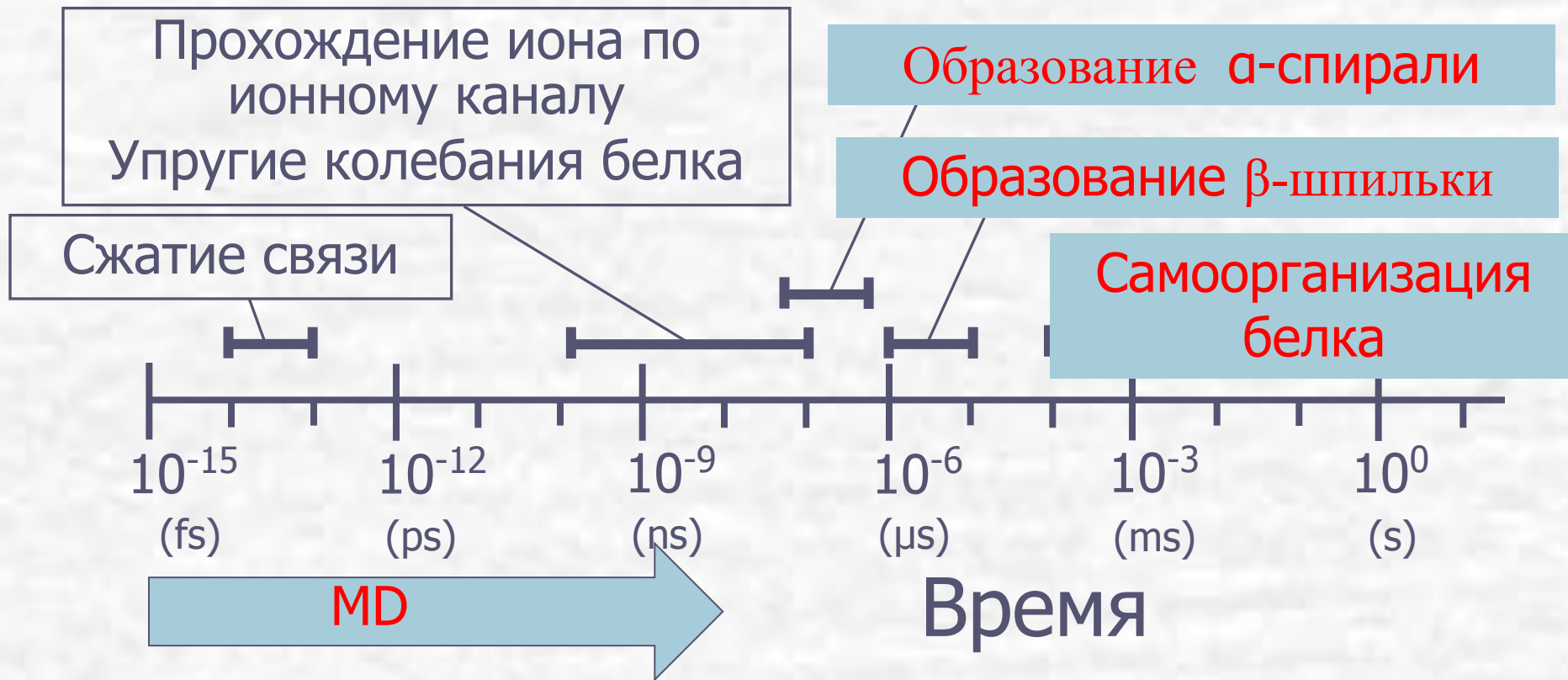
# MD требует колоссальных вычислительных мощностей

- **Квант времени в методе MD ( $\Delta t$ ) ограничен 1 фсек ( $10^{-15}$  сек).**
  - ← Размер  $\Delta t$  должен составлять приблизительно одну десятую от длительности самого быстрого изменения в системе. При моделировании белков, самое быстрое изменение, это изменение энергии сжатия-растяжения связи у легких атомов (например, O-H, C-H), и время этого изменения составляет примерно  $10^{-14}$  сек, поэтому  $\Delta t$  обычно берется около 1 фсек.
- **При биомолекулярном MD моделировании должно учитываться большое количество молекул воды.**
  - ← Количество атомных пар, учитываемых в расчетах взаимодействий (не считая химические связи, а только силы Ван дер Ваальса и электростатические силы) увеличивается по формуле  $N^2$  (где  $N$  это количество атомов).



Сложно моделировать длительный процесс. **Обычно, выполняется расчет на несколько десятков наносекунд.**

# Шкала времени изменения состояний белка и MD

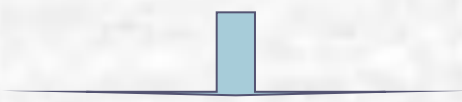


Все еще трудно моделировать весь процесс самоорганизации (сворачивание) белка, используя традиционный MD метод.

# Гораздо быстрее, Гораздо больше!

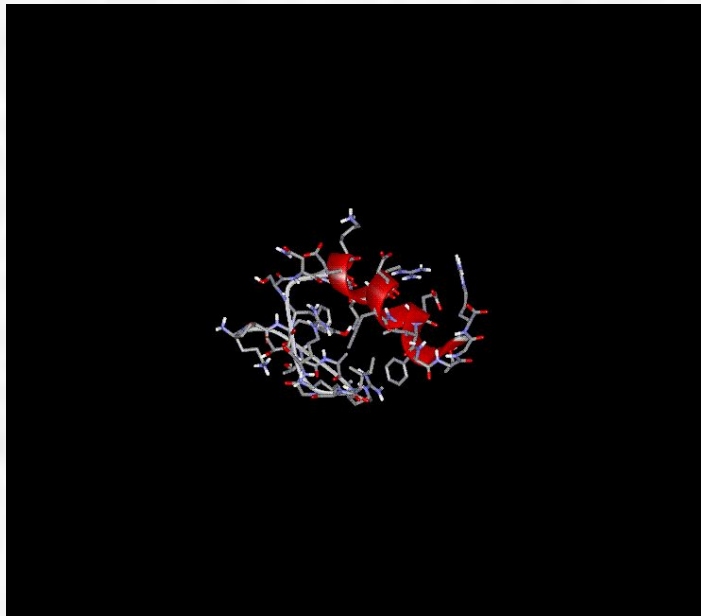
- Специальные компьютеры
  - Расчет свободных взаимодействий (не химических связей) обеспечивается с использованием специальных процессоров, разработанных для этих целей.
  - Например;
    - MDM (Машина Молекулярной Динамики) или MD-Grape: RIKEN
    - MD Engine: Taisho Pharmaceutical Co., and Fuji Xerox Co.
- Распараллеливание
  - Задача делится на несколько подзадачей и они обрабатываются каждая на своем процессоре. Сегодня, почти все MD программы для биомолекулярного моделирования (например: AMBER, CHARMM, GROMOS, NAMD, MARBLE, и т.д.) могут работать на параллельных машинах.

# Броуновская Динамика (BD)

- Динамический вклад растворителя представлен в виде рассеянного случайного воздействия (Открытие Эйнштейна в 1905). Поэтому, **молекулы воды могут не рассчитываться как каждая.**
  - С того момента как был создан BD алгоритм, с учетом того, что растворителя гораздо больше и эта энергия сглаживается за очень короткое время, **стали возможны расчеты более длинных временных периодов.**
- 
- BD метод является более подходящим для продолжительного моделирования.

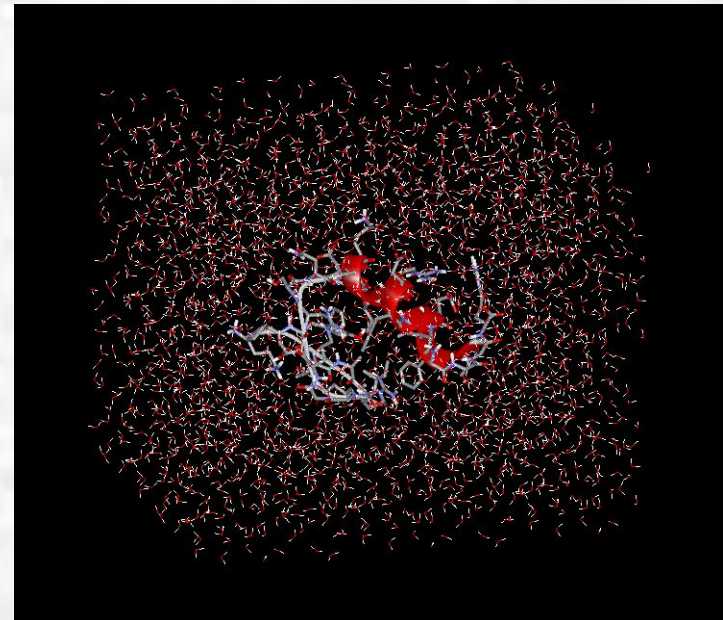
# Система для BD Моделирования

Без молекул воды



Кол-во атомов:  
304

С молекулами воды



Кол-во атомов: 304 +  
**7,377** = 7,681



# Алгоритм BD

Уравнение Ланжевина может быть выражено так:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = -\zeta_i \frac{d\mathbf{r}_i}{dt} + \mathbf{F}_i + \mathbf{R}_i \quad (7)$$

Здесь,  $\mathbf{r}_i$  и  $m_i$  отражают соответственно положение и массу атома  $i$ .  $\zeta_i$  это коэффициент трения, определяемый законом Стокса, то есть  $\zeta_i = 6\pi a_i^{\text{Stokes}} \eta$ , где  $a_i^{\text{Stokes}}$  это Стоксовский радиус атома  $i$ , а  $\eta$  это вязкость воды.  $\mathbf{F}_i$  это систематическое воздействие на атом  $i$ .  $\mathbf{R}_i$  случайное воздействие на атом  $i$ , равное 0  $\langle \mathbf{R}_i(t) \rangle = 0$ , а вариации  $\langle \mathbf{R}_i(t) \mathbf{R}_j(t) \rangle = 6\zeta_i kT \delta_{ij} \delta(t)$ ; это выводится из свойств растворителя.

Для задемпфированного предела, левую часть уравнения (7) мы приравниваем 0,

$$\zeta_i \frac{d\mathbf{r}_i}{dt} = \mathbf{F}_i + \mathbf{R}_i \quad (8)$$

Интегрированное выражение уравнения 8 называется Броуновская Динамика;

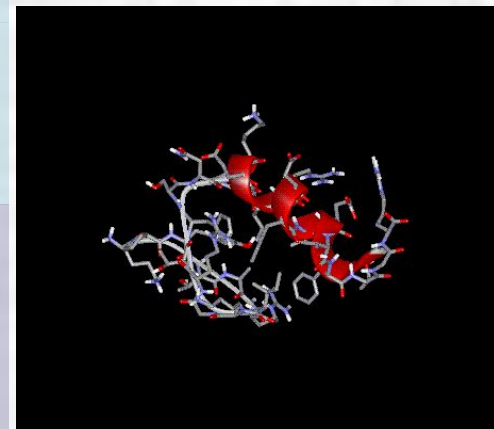
$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \frac{\mathbf{F}_i(t)}{\zeta_i} \Delta t + \sqrt{\frac{2k_B T}{\zeta_i}} \Delta t \boldsymbol{\omega}_i \quad (9)$$

где  $\Delta t$  это шаг времени, а  $\boldsymbol{\omega}_i$  это случайный вектор помех, получающийся из распределения Гаусса.

# Вычислительное время BD

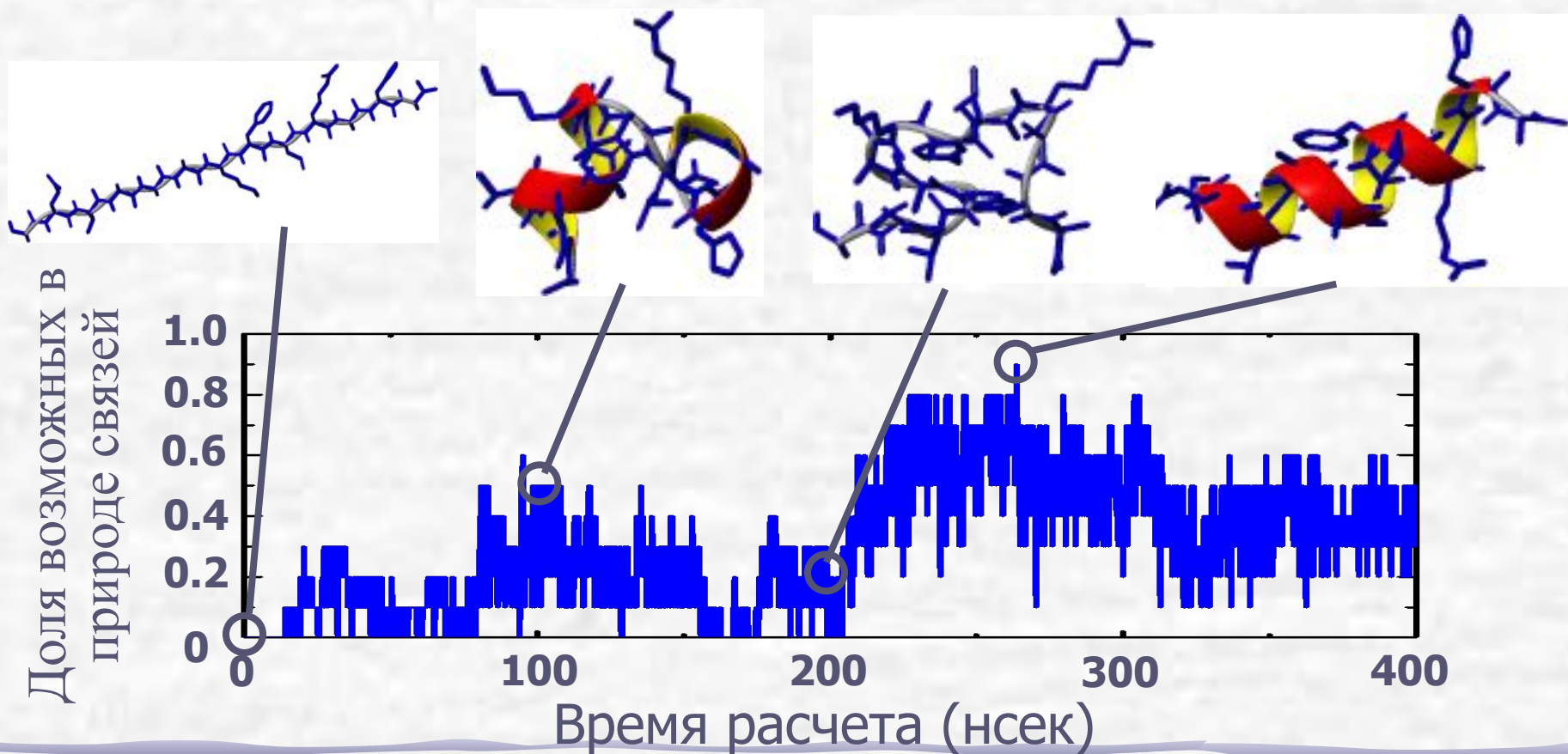
Вычислительное время, необходимое для моделирования 1 нсек изменений полипептида

Алгоритм	Компьютер	Кол-во атомов	Время (сек)	Эффективность
MD	Pentium 4 2.8 GHz	7,681	2,057	1.00
BD	Pentium 4 2.8 GHz	304	38.8	53.0
<b>BD +MTS<sup>†</sup></b>	<b>Pentium 4 2.8 GHz</b>	<b>304</b>	<b>12.8</b>	<b>161</b>
<b>BD +MTS<sup>†</sup></b>	<b>IBM Regatta 8 CPU</b>	<b>304</b>	<b>3.4</b>	<b>605</b>

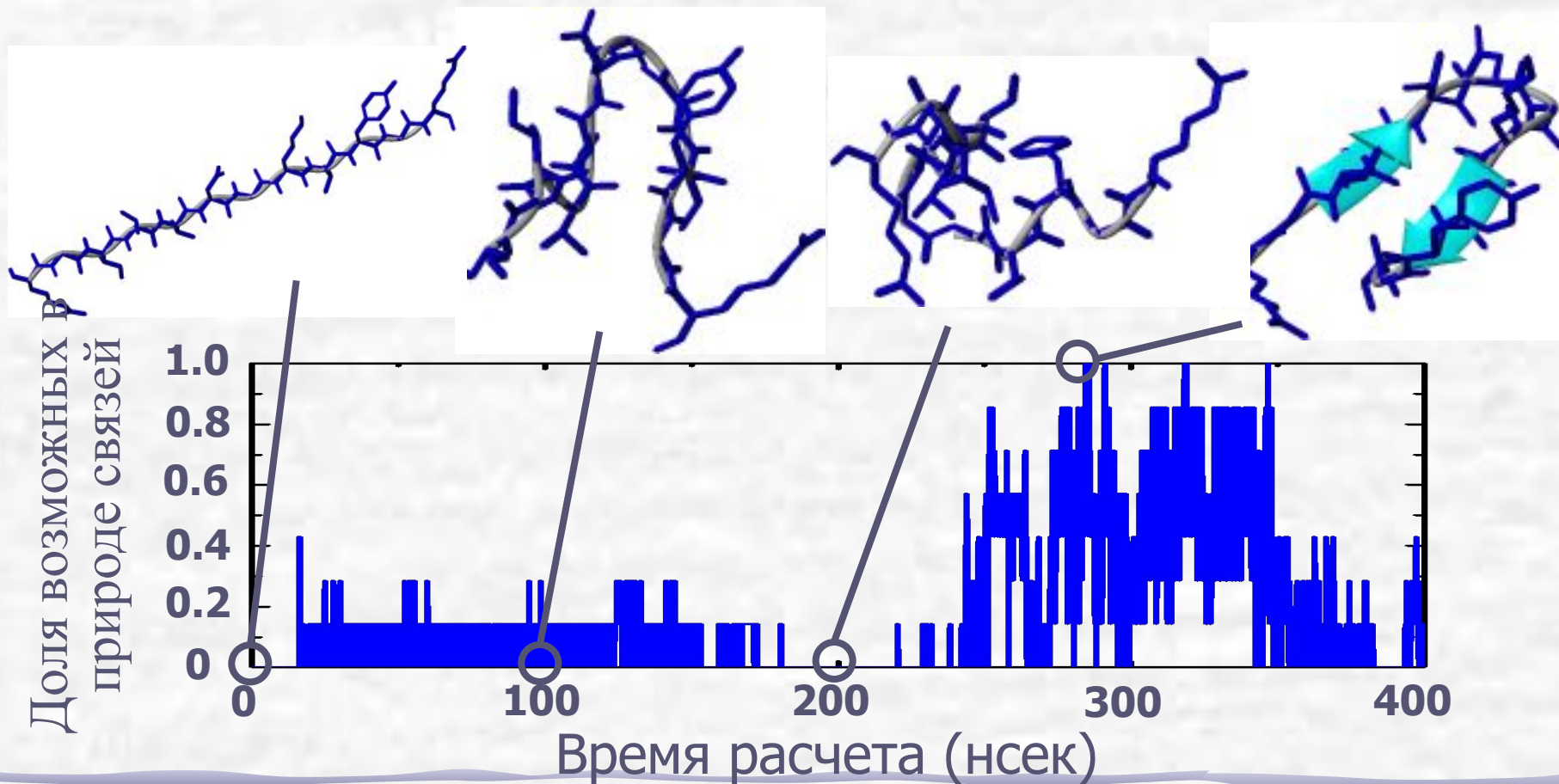


<sup>†</sup>MTS(Multiple time step) алгоритм: Этот метод позволяет уменьшить необходимость расчетов самой тяжелой части уравнения, расчет энергии свободных взаимодействий).

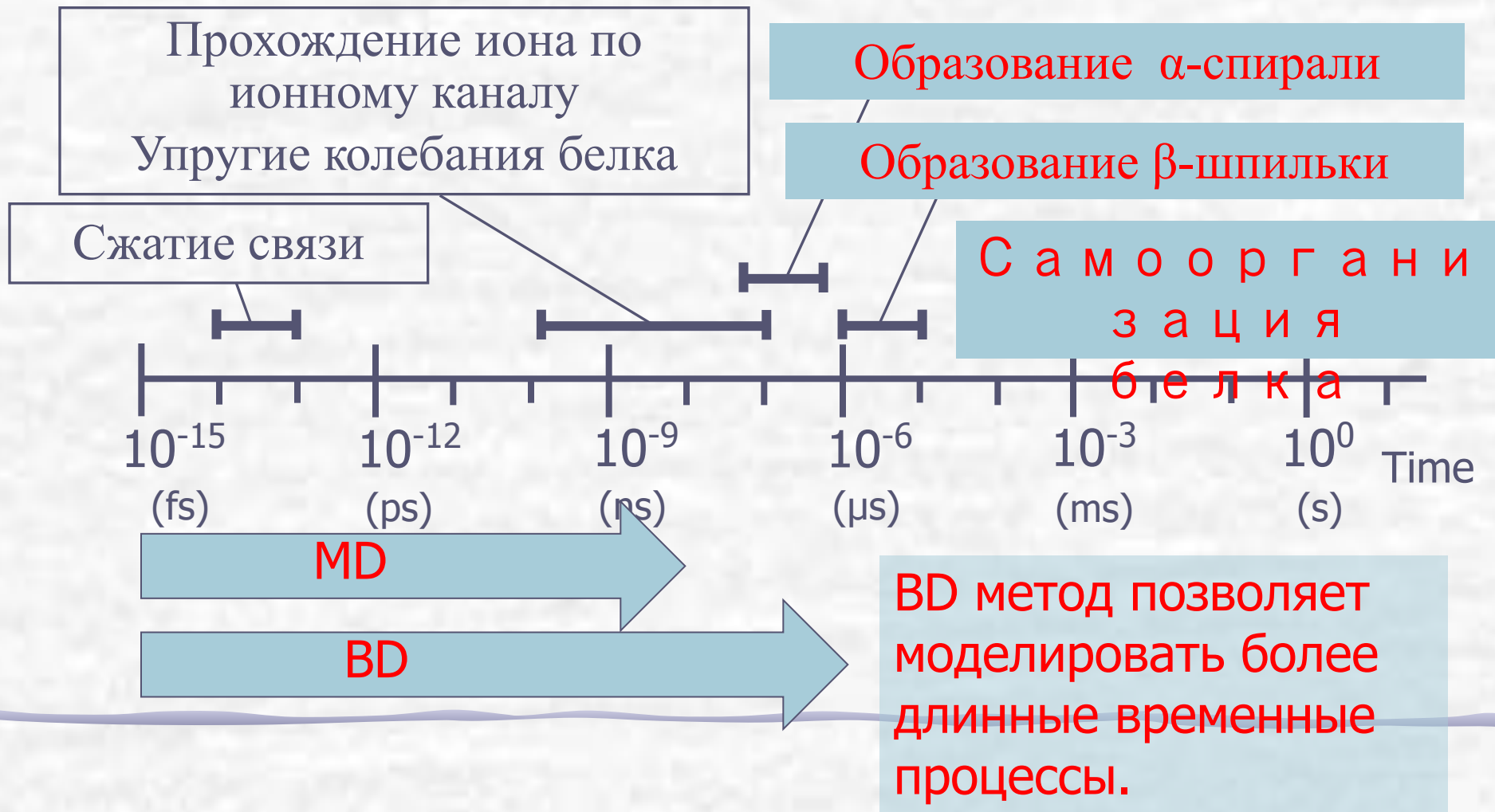
# Моделирование образования $\alpha$ -спирали с помощью BD



# Моделирование образования $\beta$ -шпильки с помощью BD



# Шкала времени изменения состояний белка и BD



# Заключение

- Проблема сворачивания молекулы белка является одной из исторических проблем в биологии. Решение проблемы откроет двери в новый период генетической биологии.
- В MD методе, Ньютоновские уравнения движения атомов системы используются в методе конечной разности.
- В MD методе, шаг времени ограничен примерно одной фсек, и является важным детальным расчет воздействия молекул воды. В связи с этим процесс моделирования длинных периодов является сложным.
- С другой стороны, процесс самоорганизации белка совершается за период от мсек до одной секунды.
- Создание алгоритмов распараллеливания и специально созданные компьютеры позволяют моделировать большие модели и гораздо быстрее.
- VD метод является более подходящим для моделирования длинных периодов.