# Структурная Биоинформатика

## Лекция 8. Моделирование самосборки белка

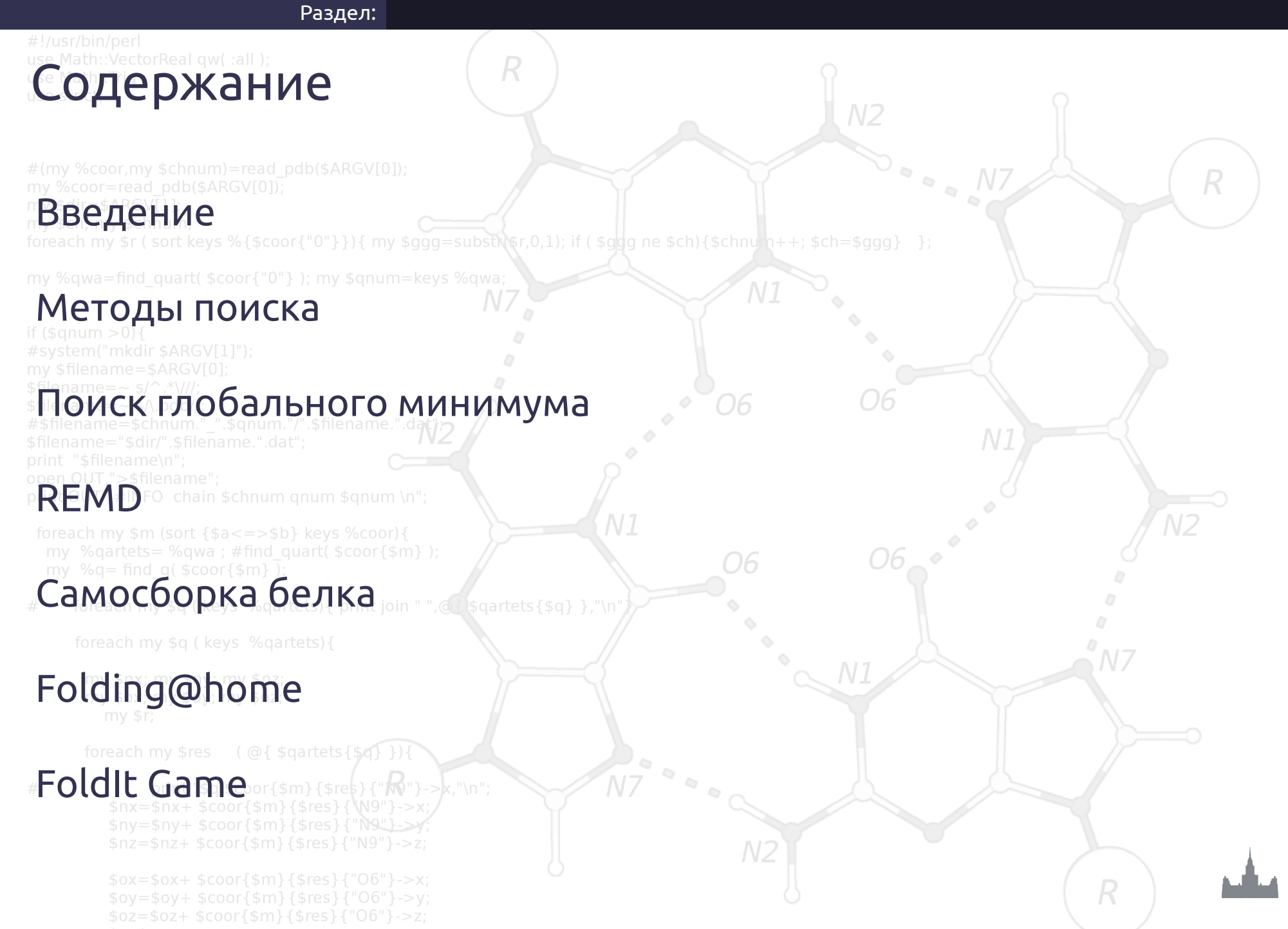
Головин А.В. [1]

[1] МГУ им М.В. Ломоносова, Факультет Биоинженерии и Биоинформатики

Москва, 2012

# Содержание





# Конформации

- В 1950 году Бартон показал, что химические и физические свойства определяются конформациями молекул.
- Это справедливо и для биополимеров.
- Важны не только оптимальные конформации.

# Систематические методы: Поиск по сетке

- Находим торсионные углы способные вращаться.
- Генерируем ротамеры по каждому углу с небольшим шагом.
- Минимизируем энергию для каждого ротамера.

$$N_{conf} = \prod_{i=1}^{N} \frac{360}{\alpha_i}$$

Если у нас пять вращаемых связей и инкремент 30 то количество конформаций 248832

# Поиск по сетке: подходы

- Необязательно рассматривать все торсионные углы, только те, которые образуются между “строительными блоками”
- Важно иметь базу фрагментов и уметь точно вычленять фрагменты из целевой молекулы.
- Первое допущение: фрагменты в молекуле не зависимы
- Второе допущение: набор заготовок должен покрывать все конформации фрагмента в составе молекулы

# Методы случайного поиска

- Основная идея это случайное изменение значений торсионных углов.
- В ходе каждой итерации надо минимизировать энергию.
- В случае с циклическими молекулами снова применяется подход с “псевдо-ациклическими” конформациями.
- Поиск сравнимый с Монте-Карло заканчивается при условии, что генерируемые конформации уже встречались.

# Дистанционная геометрия

- Конформацию молекулы можно представить как по парные расстояния между всеми атомами.
- В молекуле с N атомами количество расстояний $N\frac{(N-1)}{2}$.
- Есть удобная особенность: верхняя граница для некоторых расстояний.
- Тут становится удобно использовать матрицы для оптимизации случайных значений.

# Поиск конформаций на основе моделирования (МД или МК)

- В принципе оба метода могут быть использованы для поиска конформаций
- МК методы имеют ряд трудностей при работе с гибкими молекулами. Напоминанию, что МК методы не используют минимизацию энергии.
- Методы МД используют при высоких температурах. Конформации с минимальной энергией подвергают процедуре минимизации энергии.

# Выбор метода для поиска конформаций

## Поиск конформеров циклогептадекана $C_{17}H_{34}$

| Метод | Количество конформеров |
|---|---|
| Систематический поиск | 211 |
| Случайный коор. поиск | 222 |
| Случайный торс. поиск | 249 |
| Дистанционные ограничения | 176 |
| МД | 169 |

# Поиск глобального минимума

- Эволюционные алгоритмы
- Методы моделирования "отжига"
- На ряду с поиском глобального минимума эти методы также используются пр докинге, молекулярном дизайне, QSAR и определении фармакофора.

# Эволюционные алгоритмы

**Три основных метода**

- Генетический алгоритм
- Эволюционное программирование
- Эволюционные стратегии

Общее: создаётся набор вероятных решений проблемы. Решения становятся хорошими или плохими и эволюционируют улучшаясь. Мера “хорошо” это внутренняя энергия молекулы определяемая из молекулярной механики.

# Моделирование отжига

- Отжиг — это процесс при котором расплавленное вещество медленно остужают и оно кристаллизуется в идеальный кристалл.
- Идеальный кристалл и есть искомый минимум энергии
- Моделирование отжига направлено именно на достижение глобального минимума с при моделировании охлаждения системы.

# Моделирование отжига

- Важно тщательно контролировать температуру
- Не факт что метод может привести к глобальному минимуму, но если несколько запусков привели к одной структуре, то можно это предположить.

# Конформации белка из ограниченной МД

- Суть идеи состоит в том, что в молекулярную динамику вводятся потенциалы, которые увеличивают энергию если конформация не соответствует экспериментальным данным.
- Например: проверяем в ходе оптимизации геометрии R-фактор структуры и если он отличается от экспериментального, то увеличиваем энергию.

$$E_{tot} = U_{pot} + E_{sf}$$

$$E_{sf} = S \sum ||F_{obs}| - |F_{calc}||^2$$

# Метод обмена репликами (REMD)

- Основная идея: запустить параллельно несколько счётов с разными температурами.
- Мы можем выбрать правило когда производить обмен конформациями.
- Если мы проводим обмен когда потенциальная энергия одной из реплик ниже чем других, то это похоже на моделирование отжига.
- Такой подход часто используется для моделирования самосборки.

# REMD

- Цель метода это ускорить сканирование (sampling) конформационного пространства.
- Применимо к переходам через значимые энергетические барьеры.
- В Gromacs обмен между репликами происходит случайно по условию:

$$P(1 \leftrightarrow 2) = \min\left(1, \exp\left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2}\right)(U_1 - U_2)\right]\right)$$

А скорости маштабируются: $(T_1/T_2)^{\pm 0.5}$

# REMD: количество реплик

Разница между температурами (репликами):

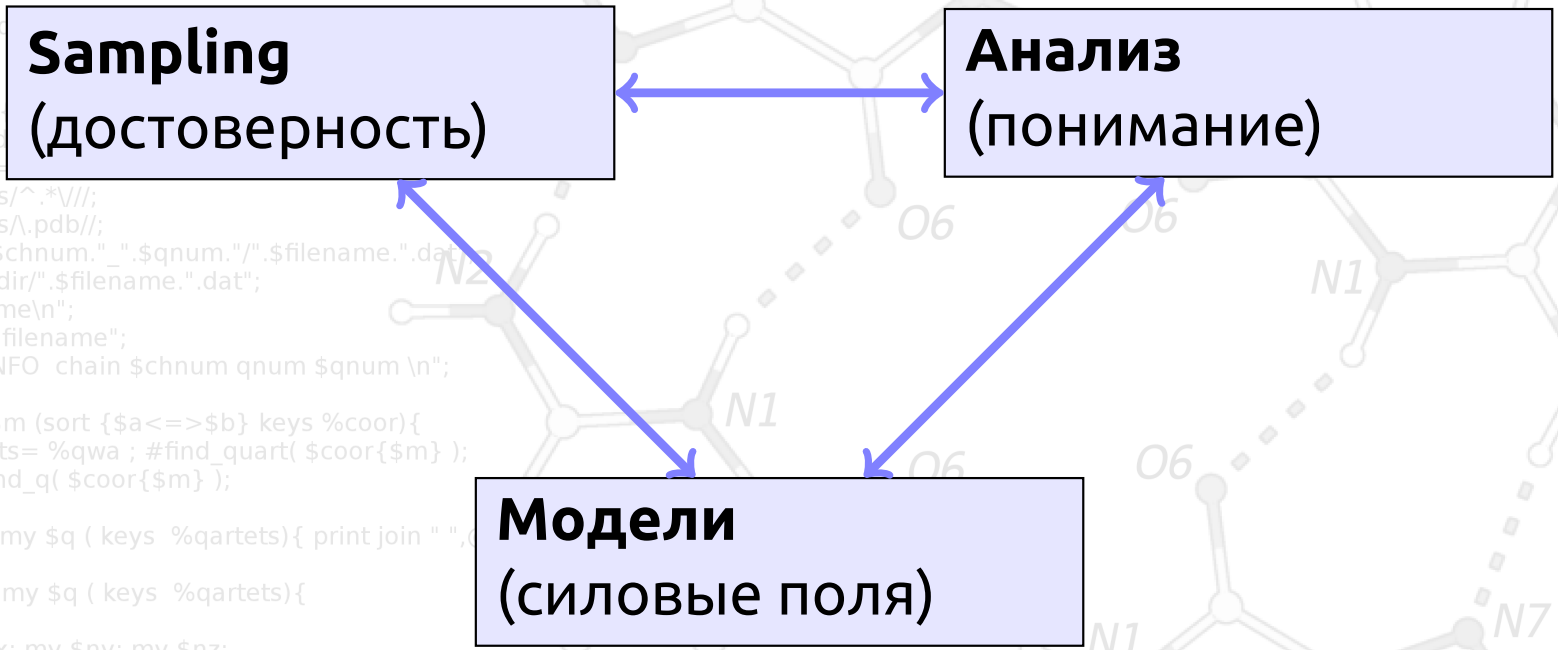
$$U_1 - U_2 = N_{df}\frac{c}{2}k_B(T_1 - T_2)$$

где $N_{df}$ это количестве степеней свободы и "c" это величина от 1 до 2 для системы белок вода.
Если $T_2 = (1+\epsilon)T_1$ тогда вероятность обмена:

$$P(1 \leftrightarrow 2) = \exp\left(-\frac{\epsilon^2 c\, N_{df}}{2(1+\epsilon)}\right) \approx \exp\left(-\epsilon^2\frac{c}{2}N_{df}\right)$$

Таким образом для вероятности обмена $e^{-2} \approx 0.135$ получаем $\epsilon \approx 2/\sqrt{c\, N_{df}}$.
И если мы контролируем длину связей, то: $N_{df} \approx 2\, N_{atoms}$
и при $c$ = 2 надо использовать: $\epsilon = 1/\sqrt{N_{atoms}}$.

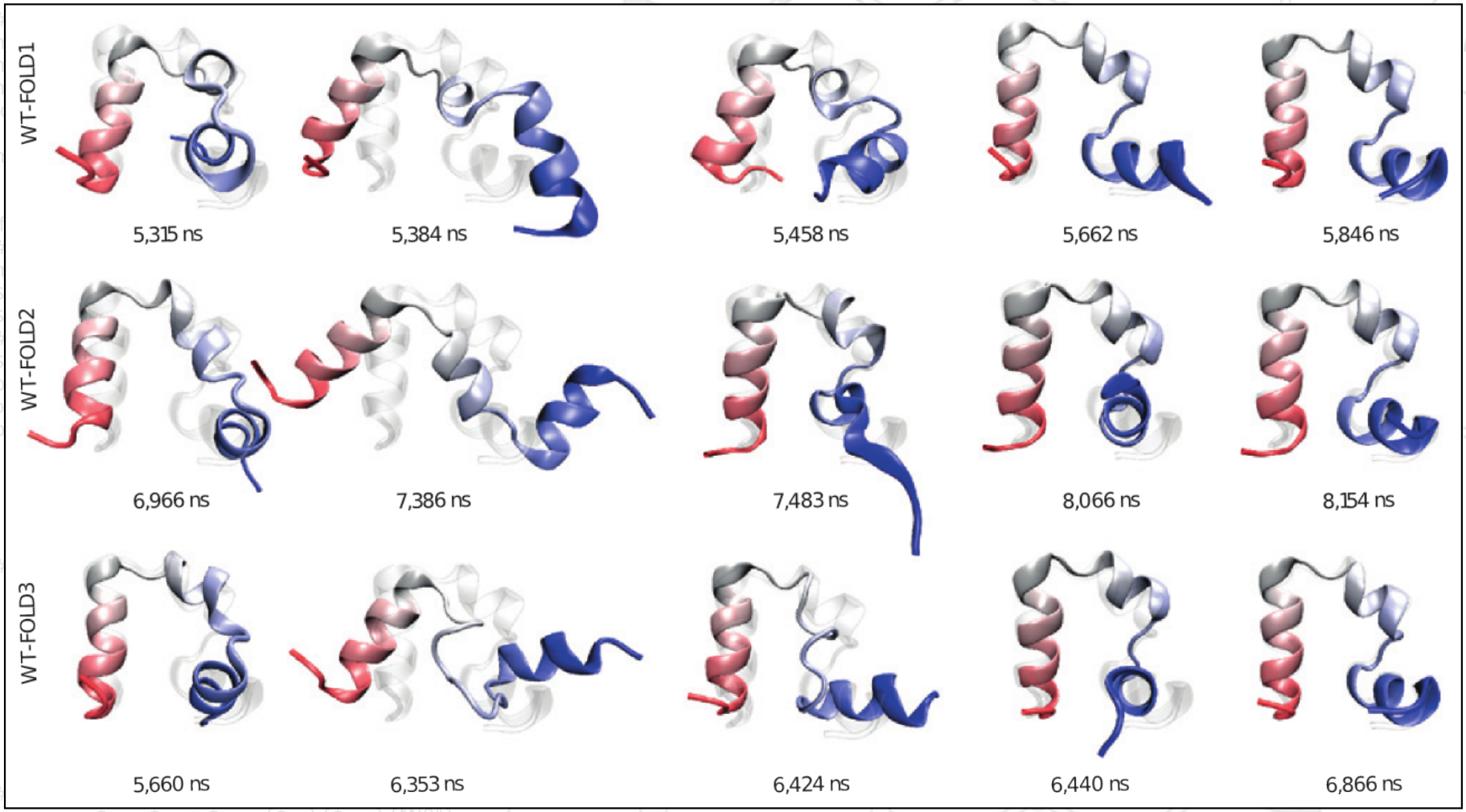
# Вызовы молекулярного моделирования



# Моделирование самосборки белка

- Как иметь дело с большими временами?
  Самые "быстрые" белки собираются за 10-100 $\mu$с, моделирование должно быть на порядок длинее.
- На сколько хороши наши силовые поля?
  Сможем ли мы получать нативное состояние белка без знания структуры.
  Будут ли предсказанные структуры иметь правильные параметры: $\Delta G$, $K_f$.
- Сможем ли мы изучать сборку моделированием?
  Мехнаизмы, Теория …

# МД в моделировании самосборки белка




DOI: 10.1038/NPHYS1713

# МД в моделировании самосборки белка

- Традиционный подход: несколько длинных траекторий
- Альтернатива: методы стохастического кинетического сканирования.

Фолдинг это стохастический процесс с экспоненциальной кинетикой,т.е. количество молекул,которые собрались:

$$f(t) = M[1 - exp^{(-kt)}]$$

для малых времён:

$$f(t) \sim Mkt \quad M \sim 10,000 procs, k \sim 1/10,000 ns, t \sim 20 ns/proc$$

ожидается, что можно увидеть 20 раз сборку белка.

- Это эффективно.
- Эргодично

# Folding@home

# Folding@home

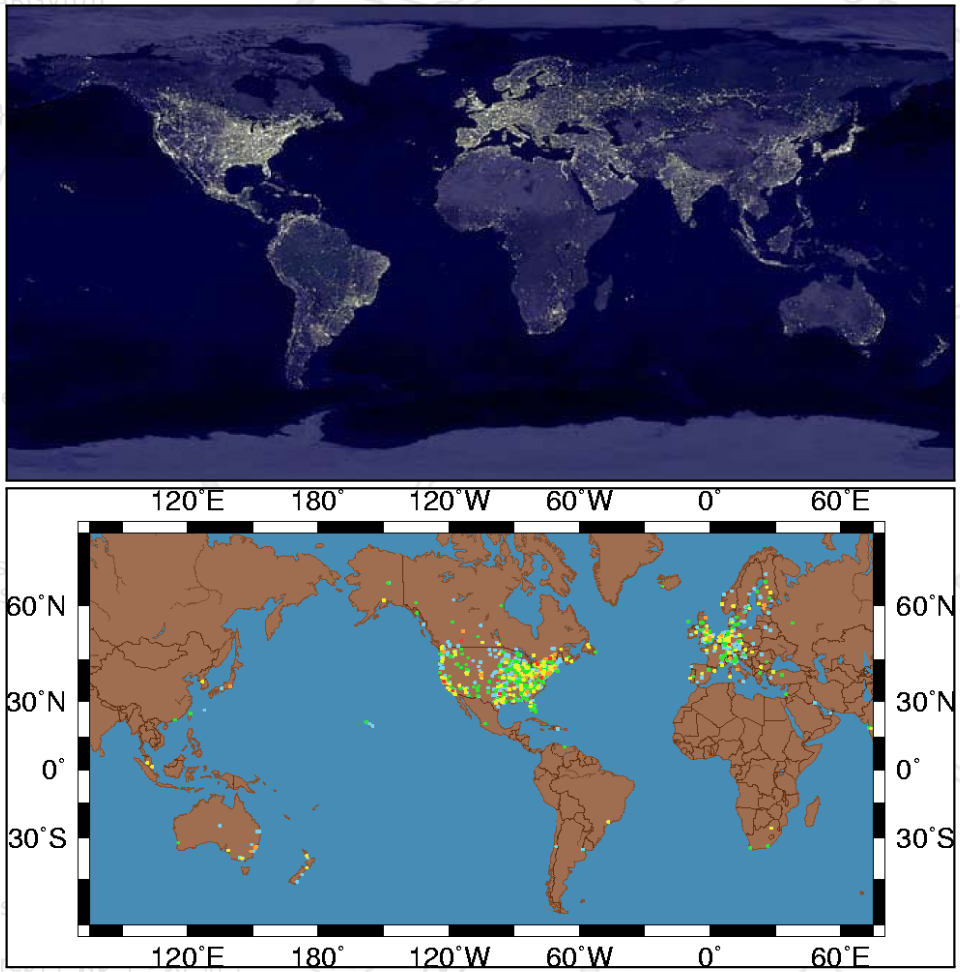



Illustration from Stefan Larson

# Результаты

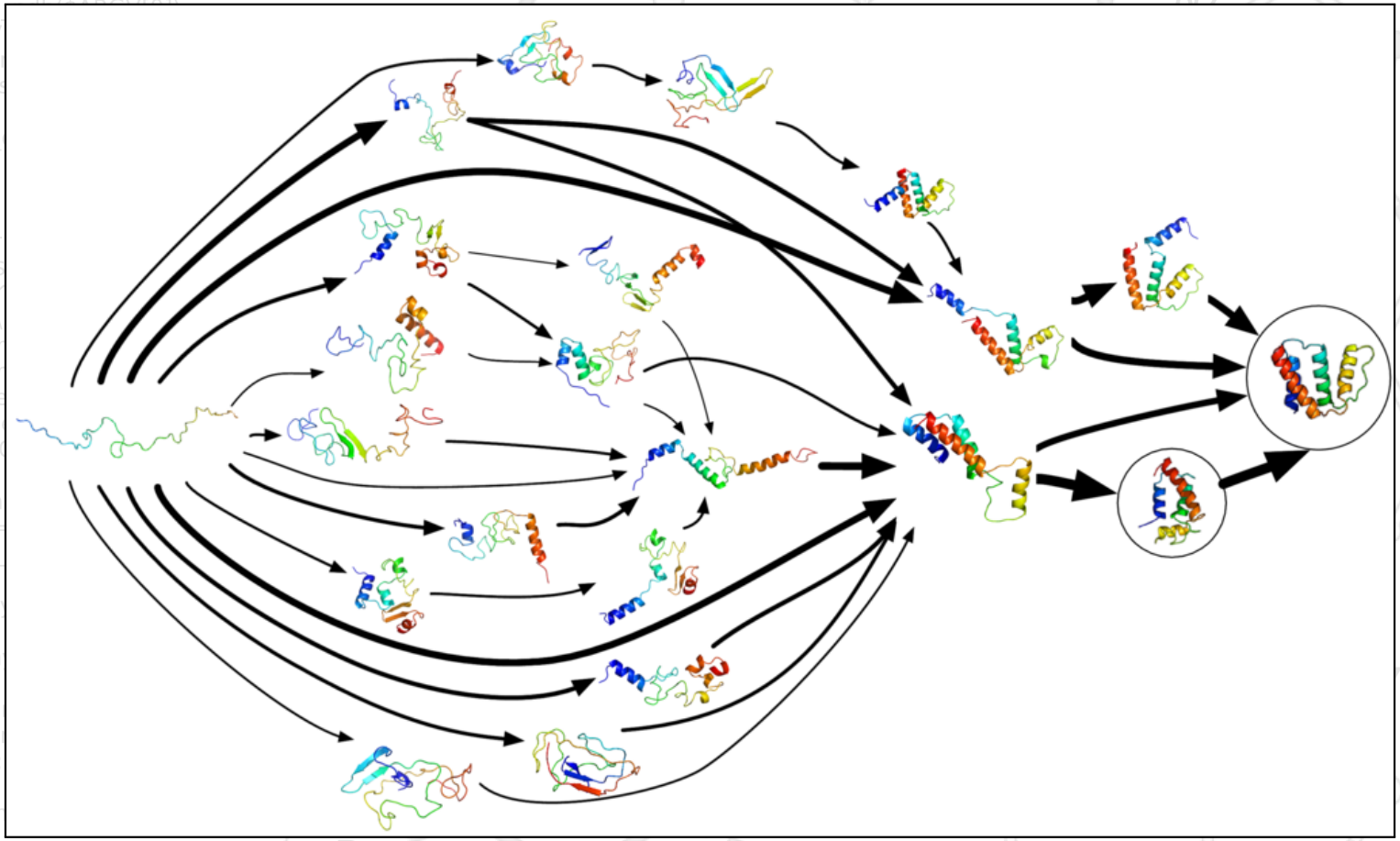


Illustration from Wikipedia

# Результаты и ответы

Ответы на наши вопросы:

- Выборка достаточна.
- Силовые поля работают
  Сборка происходит правильно и с нужной скоростью
- Механизм
  Вероятно, что самосборка это не универсальный механизм и индивидуален для каждого белка

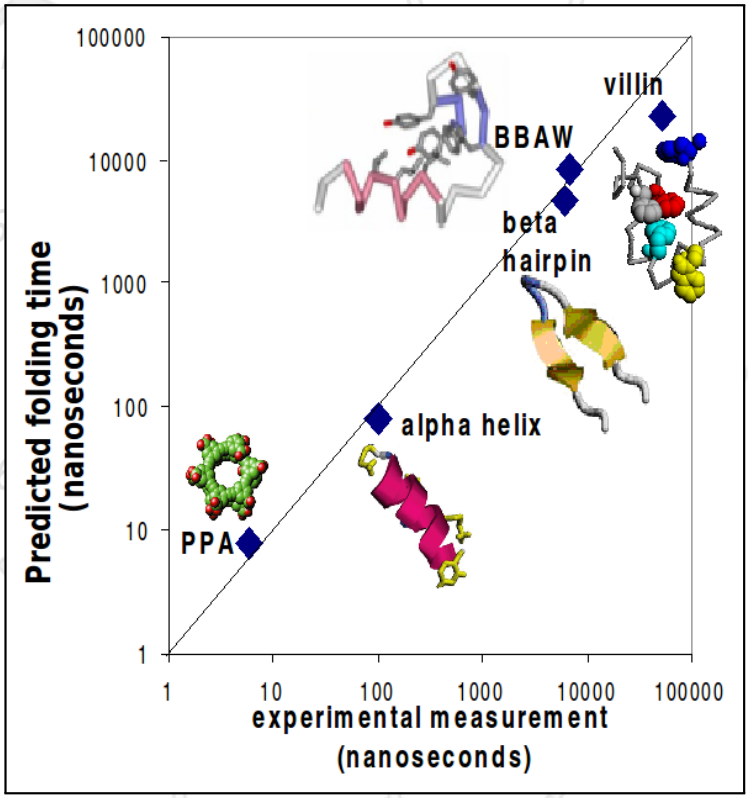



Illustration from Stefan Larson

# FoldIt Game

Основная идея:

- Использование **crowdsourcing** для поиска оптимальной геометрии.
- Пользователь может изменять торсионные углы и добиваться оптимальной энергии.
- В ходе игры для прохождения уровня пользователь должен достичь нужного уровня энергии для структуры
- Создатели программы собираются утилизировать способы придуманные людьми для разработки новых алгоритмов.

# FoldIt

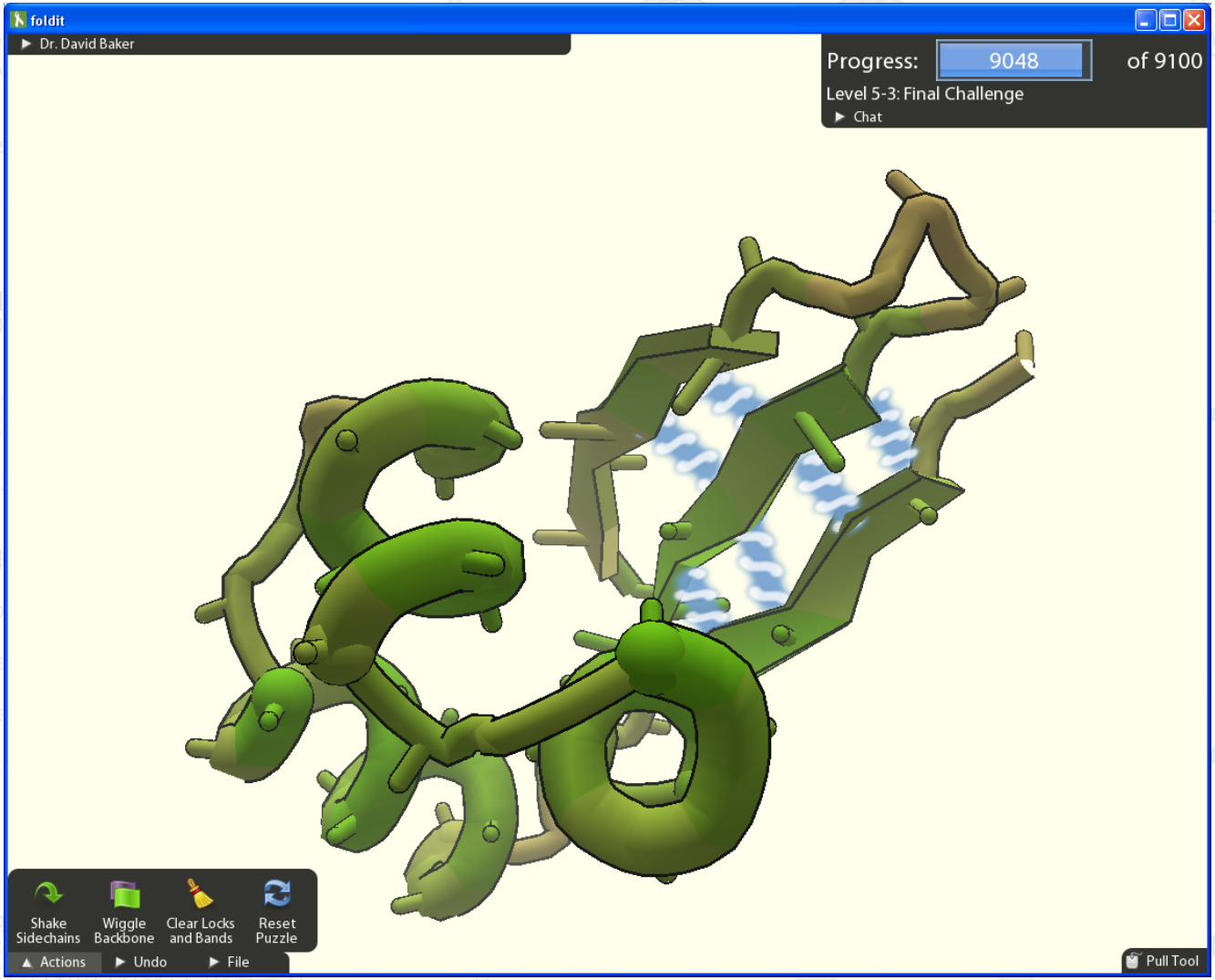
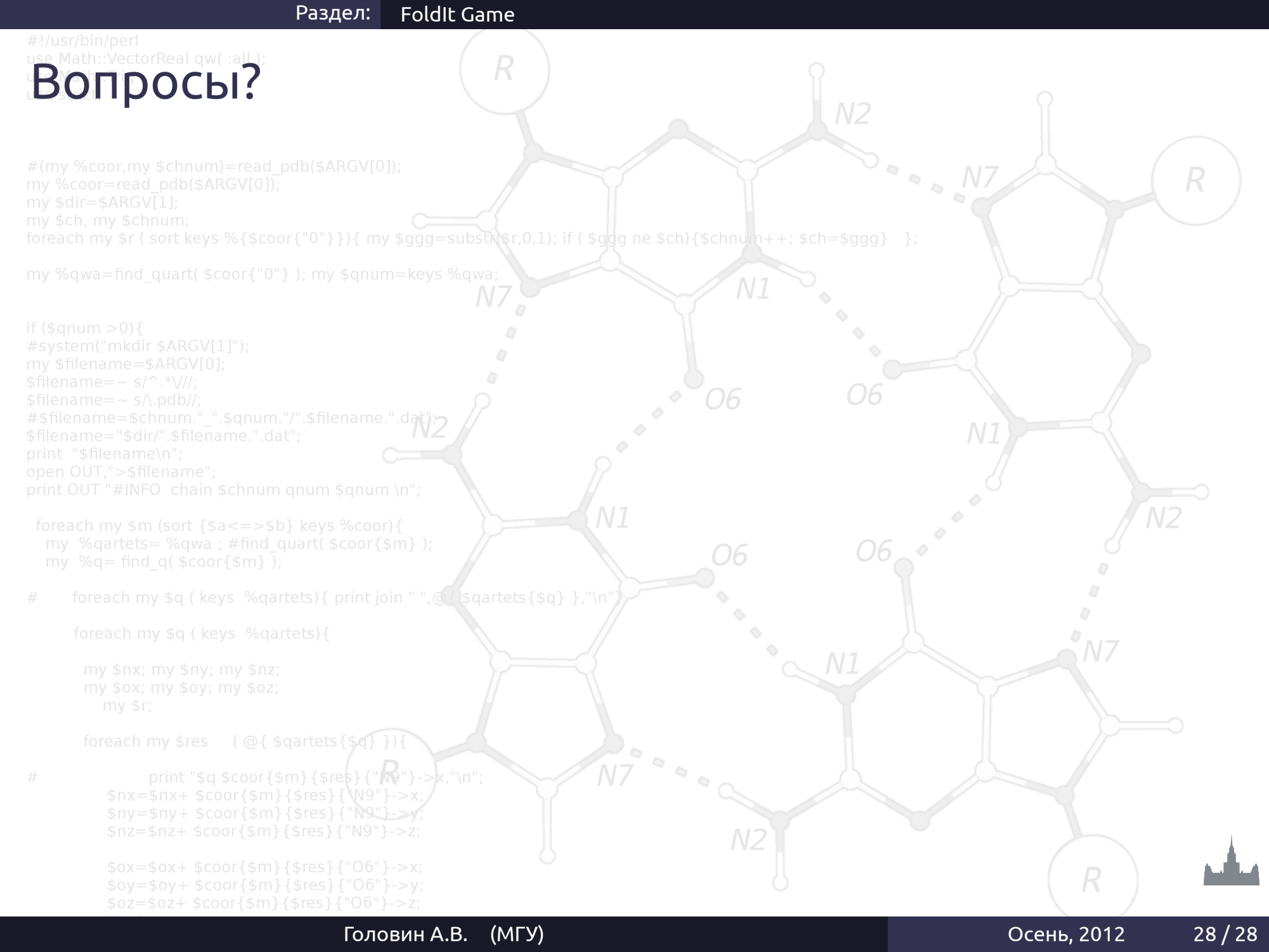



Illustration from Wikipedia

# Вопросы?


#!/usr/bin/perl
use Math::VectorReal qw( :all );
#(my %coor,my $chnum)=read_pdb($ARGV[0]);
my %coor=read_pdb($ARGV[0]);
my $dir=$ARGV[1];
my $ch, my $chnum;
foreach my $r ( sort keys %{$coor{"0"}}){ my $ggg=substr($r,0,1); if ( $ggg ne $ch){$chnum++; $ch=$ggg} };
my %qwa=find_quart( $coor{"0"} ); my $qnum=keys %qwa;
if ($qnum >0){
#system("mkdir $ARGV[1]");
my $filename=$ARGV[0];
$filename=~ s/^.*\///;
$filename=~ s/\.pdb//;
#$filename=$chnum."_".$qnum."/".$filename.".dat";
$filename="$dir/".$filename.".dat";
print "$filename\n";
open OUT,">$filename";
print OUT "#INFO chain $chnum qnum $qnum \n";
foreach my $m (sort {$a<=>$b} keys %coor){
my %qartets= %qwa ; #find_quart( $coor{$m} );
my %q= find_q( $coor{$m} );
# foreach my $q ( keys %qartets){ print join " ",@{ $qartets{$q} },"\n"}
foreach my $q ( keys %qartets){
my $nx; my $ny; my $nz;
my $ox; my $oy; my $oz;
my $r;
foreach my $res ( @{ $qartets{$q} }){
# print "$q $coor{$m}{$res}{"N9"}->x,"\n";
$nx=$nx+ $coor{$m}{$res}{"N9"}->x;
$ny=$ny+ $coor{$m}{$res}{"N9"}->y;
$nz=$nz+ $coor{$m}{$res}{"N9"}->z;
$ox=$ox+ $coor{$m}{$res}{"O6"}->x;
$oy=$oy+ $coor{$m}{$res}{"O6"}->y;
$oz=$oz+ $coor{$m}{$res}{"O6"}->z;
R
N2
N7
R
N7
N1
O6
O6
N2
N1
N1
N2
O6
O6
N1
N7
N7
N2
R