

Вероятностные методы планирования движения и их применение к анализу поведения протеинов

Дипломная работа студента 545 группы

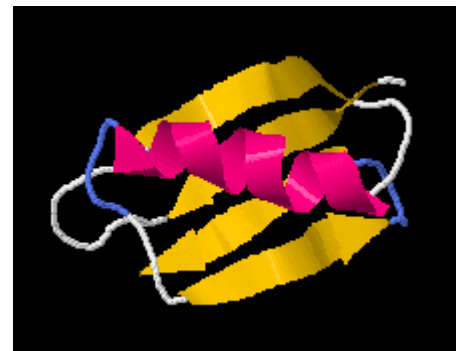
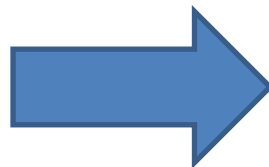
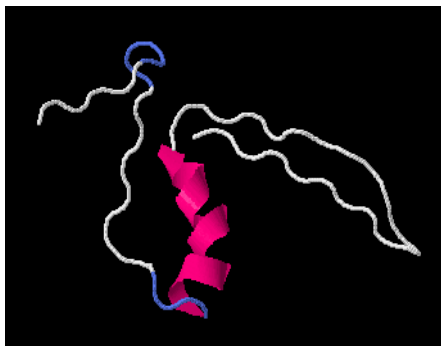
Катышева Алексея Александровича

Научный руководитель: к.ф.-м.н. К. В. Вяткина

Рецензент: д. ф.-м. н. Е. О. Степанов

Анализ поведения протеинов

- Белки (протеины) синтезируются в клетках и выполняют важные для организма функции
- Прежде чем выполнять свои функции каждый белок принимает свою естественную конфигурацию – происходит так называемый фолдинг белка



Проблемы

- Возможные последствия неправильного преобразования белка (мисфолдинга):
 - болезнь Альцгеймера
 - болезнь Хантингтона
 - коровье бешенство
 - ...

Известные подходы

- Молекулярная динамика
- Симуляция Монте-Карло
- Приближенные методы построения конкретной траектории
- Методы планирования движения

Методы планирования движения

- Изначально применялись в робототехнике
- Задача – нахождение эффективного перехода системы из одного состояния в другое
- Конфигурационное пространство
 - поиск пути из одной точки в другую
 - препятствия – запрещенные области

Вероятностная карта дорог

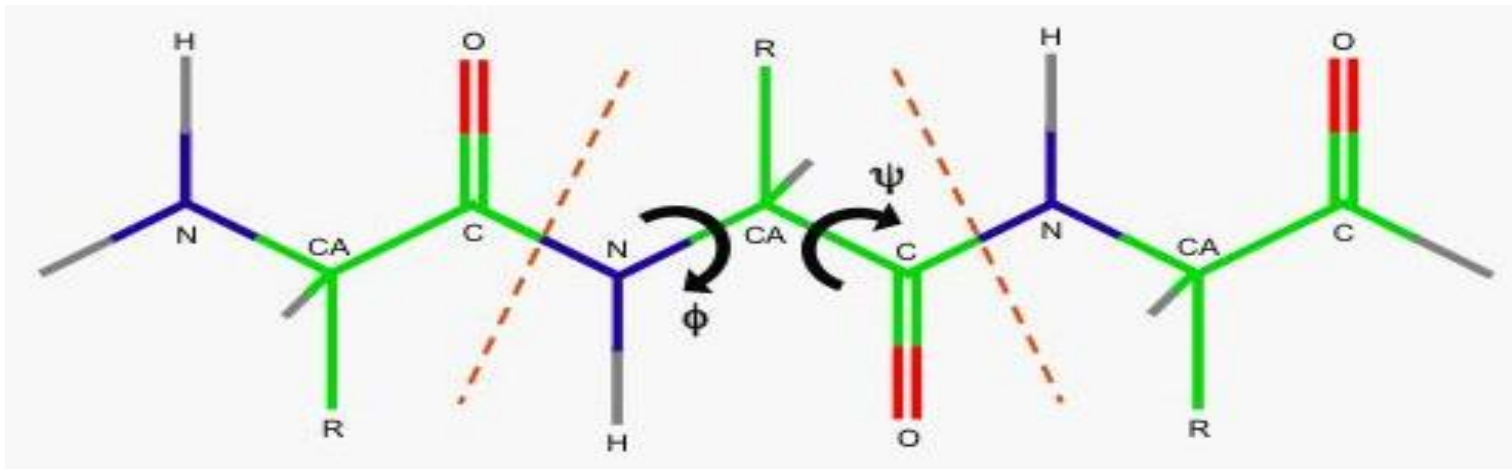
- Probabilistic RoadMap (PRM)
- Стэнфордский Университет и Университет Утрехта (1996)
- Эффективен для конфигурационных пространств высокой размерности

Цели работы

- Реализовать метод построения вероятностной карты дорог для задачи планирования движения протеинов.
- Оценить применимость полученной реализации

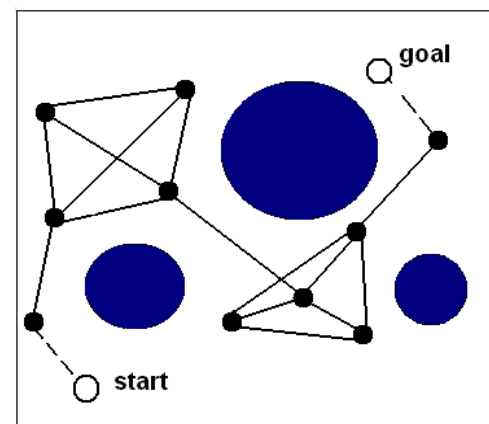
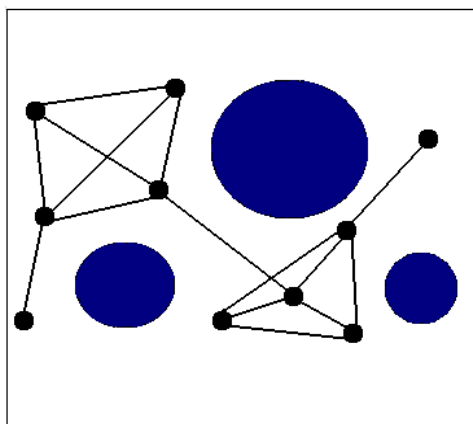
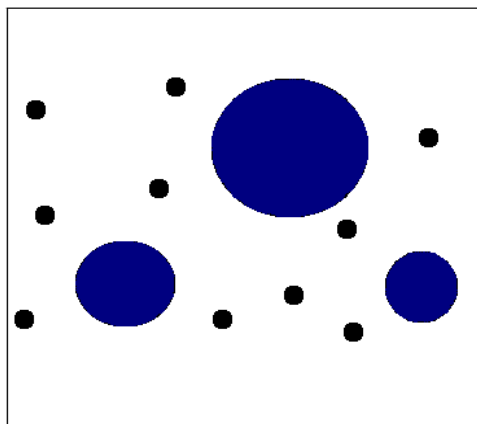
Применимость к протеинам

- Каждая цепочка протеина представляет собой последовательность аминокислот
- Каждая аминокислота – 2 степени свободы



Основные этапы

1. Генерация будущих вершин графа в допустимых областях
2. Соединение вершин ребрами (проверка внутренних точек отрезка)
3. Ответ на запрос

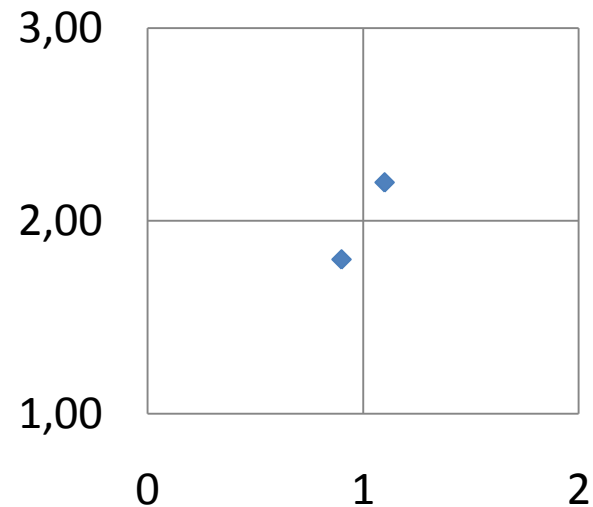
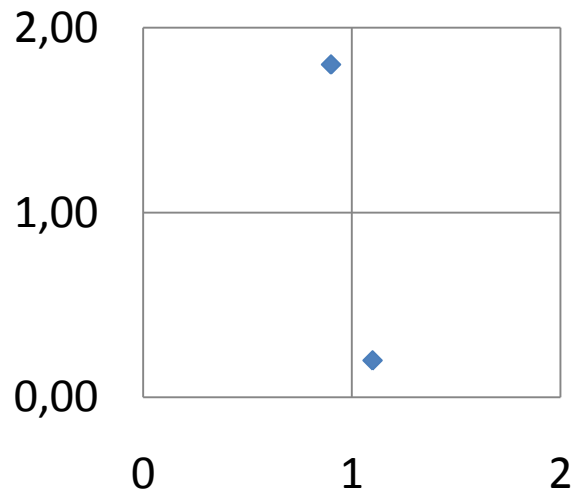


Особенности реализации

- Работа с файлами PDB (Protein Data Bank)
 - банк 3D-структур белков и нуклеиновых кислот
 - постоянно пополняется и обновляется
- Сохранение полученной карты дорог в файл
 - построение занимает несколько часов

Особенности реализации

- Конфигурационное пространство
– работа с циклическими координатами



точки $(1.1\pi, 0.2\pi)$ и $(0.9\pi, 1.8\pi)$

Особенности реализации

- Библиотека для различных геометрических преобразований в 3D пространстве
 - быстрое перестроение структуры белка по координатам точки в пространстве
- Поиск кратчайших путей в графе
 - Алгоритм Дейкстры для разреженных графов, $O(E \log V)$

Результаты работы программы

- Молекула белка GB1 (иммуноглобулин)
- 56 аминокислот -> 110 степеней свободы

	#V	#E	#MaxCC	Query Time	Acc.
PRM	2000	55000	1994	5	98
AS	2000	50000	1998	50	
PRM	10000	470000	9992	25	99
AS	10000	260000	9997	394	

- AS - N. M. Amato, G. Song. Using motion planning to study protein folding pathways (2002)

Результаты

- Разработана программная реализация метода построения PRM для решения задачи планирования поведения протеинов
- Полученная реализация оформлена в виде консольного приложения
- Показана применимость полученной реализации