

Проекты

Порозов Ю., к.м.н.

porozov@ifc.cnr.it

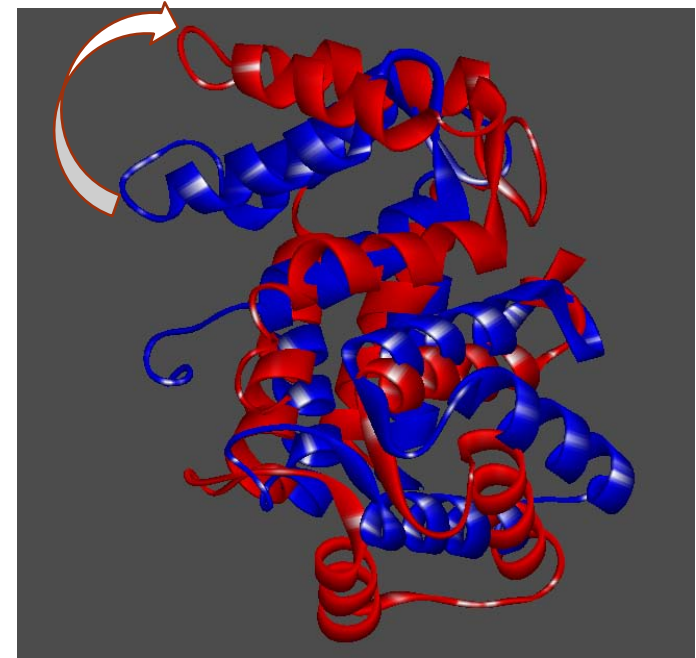
PDB-by-RMSD

В PDB хранятся записи о третичных структурах белков. На 15.09 этих записей 75801. Из них чуть больше 9000 - структуры, разрешенные при помощи NMR. Сайт PDB позволяет выбирать записи из базы данных по различным признакам, среди которых метод, разрешение, дата получения, ключевые слова, PDB ID, особенности структуры, авторы и т.д. В то же время часто бывают нужны записи из PDB (структуры NMR), в которых конформации находятся на строго определенном расстоянии.

Цель: разработка инструмента для получения записей из PDB, в которых конформации находятся на определенном расстоянии друг от друга (метрика расстояния root mean square deviation, RMSD).

Задачи: изучить методы получения записей PDB через веб и API. Освоить манипуляции с PDB файлами - считывание, получение информации о координатах атомов по каждой модели в файле. Научиться вычислять RMSD между конформациями. Разработать способ выравнивания всех конформаций в файле PDB по трём точкам. Обеспечить случайный поиск по PDB по заданным параметрам (min и max) RMSD между конформациями.

Full-atom root mean square deviation



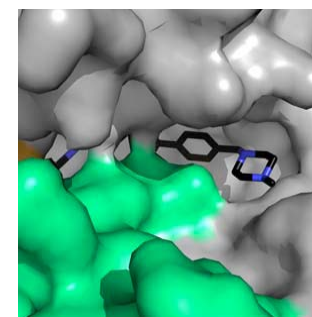
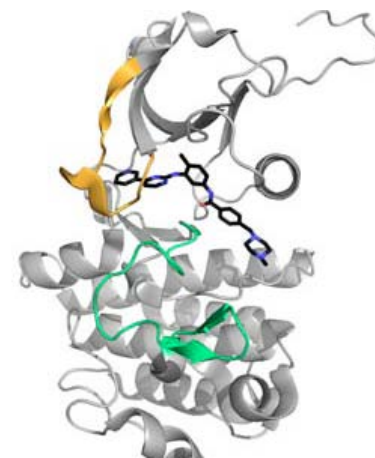
$$\text{RMSD}(v,w) = \sqrt{\frac{1}{n} \sum_{i=1}^n (v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2}$$

Таргетная и персонифицированная терапия – медицина будущего. CML-AutoDock.

Мутации белка BCR-ABL являются ключевым фактором, определяющим эффективность терапии при хроническом миелоидном лейкозе (CML).

Цель: Разработать прототип экспертной системы для выбора адекватного препарата в зависимости от имеющихся мутаций в BCR-ABL - шаг в направлении персонифицированной терапии.

Задачи: изучение способов взаимодействия с ПО AutoDock (Python, скрипты) из других программ. Создание прототипа БД с информацией об известных мутациях, примененных препаратах (основных всего три), эффективности терапии и результатами моделирования в AutoDock. Функционал должен предусматривать занесение в БД новых мутаций врачом с последующим автоматическим запуском протокола докинга на AutoDock и возвращением результатов моделирования.

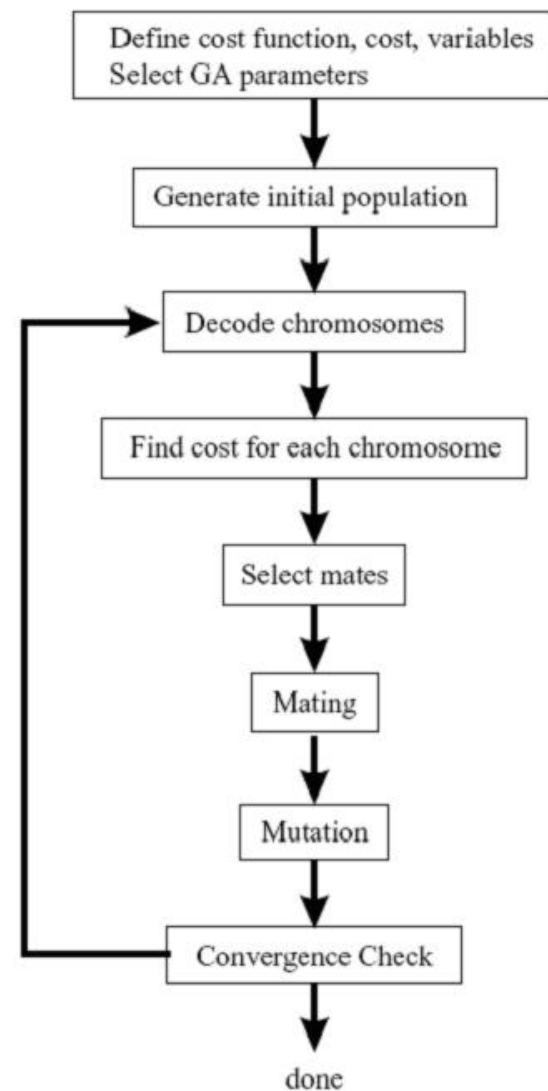


ProtGen – глобальная оптимизация движений белка

ProtGen. Моделирование движений, которые проделывают белки*, является ключевым вопросом как для понимания функции белков, так и для разработки методов модификации их функций. Существуют методы, которые быстро находят локальный минимум некоторого функционала, который определяет траекторию возможного движения. Но важной является и такая оптимизация функционала, при которой будет найден глобальный минимум, отражающий наиболее оптимальную траекторию.

Цель: разработать метод\ПО, реализующее парадигму эволюционных вычислений (а именно генетический алгоритм) для глобальной оптимизации функционала стоимости движения упрощенной модели белка.

* - в данном случае белок – это цепь, в которой разрешены повороты между звеньями. Движения определяются вектором этих поворотов.



HumanORI. Машинное обучение для поиска сайтов начала репликации в ДНК человека.

Репликация ДНК - важнейший клеточный процесс, определяющий рост и размножение. Репликация начинается в строго определенных участках ДНК. До сих пор не предложено способа предсказания этих участков в ДНК человека.

Цель: Разработать приложение, позволяющее сохранять в локальной БД информацию о многочисленных факторах (tracks), ассоциированных с определенным диапазоном "адресов" (нуклеотидов) из проекта ENCODE.

Задачи: освоить методы получения информации из ENCODE - FTP-доступ, API, Table Browser. Разработать ПО, позволяющее локально сохранять запрашиваемые треки из ENCODE и обеспечивающее базовые манипуляции с сохраненными данными - поиск, выборку, объединение, экспорт и пр.

