

МГУ им. Ломоносова, факультет биоинженерии и биоинформатики

# Обзор метода моделирования структур биополимеров Rosetta

## Оглавление

Введение .....	2
Описание основ подхода, используемого в Rosetta .....	3
Энергетическая функция.....	3
De novo предсказания структур с помощью Rosetta .....	7
Выбор фрагментов .....	7
Сборка фрагментов .....	7
Примеры успешных применений .....	9
1) <i>Решение фазовой проблемы</i> .....	9
2) <i>Белковый докинг</i> .....	10
3) <i>Дизайн белков</i> .....	11
4) <i>Дизайн ферментов</i> .....	12
Заключение.....	14
Список использованной литературы.....	15

## Введение

Белки – линейные последовательности аминокислот, которые складываются в уникальные трехмерные структуры, выполняющие различные биологические функции. Определение этих структур – крайне важная медицинская и научная задача, решение которой помогает как в понимании функций и механизмов действия белка, так и в моделировании лекарств или новых, более эффективных белков. На текущий момент, самыми аккуратными и точными методами определения трехмерных структур белков являются рентген структурный анализ и спектроскопия ядерно магнитного резонанса. К несчастью, эти два метода дорогостоящи и требуют большого времени. С другой стороны, методы секвенирования гораздо более продуктивны, и для многих белков известна их последовательность. Общепринятым считается, что трехмерная структура белка зависит только от его аминокислотной последовательности, а значит, зная первичную структуру белка, мы потенциально можем предсказать его третичную структуру.

Данная работа посвящена одному из подходов к решению таких задач – Rosetta.

## Описание основ подхода, используемого в Rosetta

Основные методы предсказания структур белков *de novo* базируются на предположении, что все наблюдаемые конформации и укладки белков обладают наименьшей энергией. Таким образом, проблема предсказания структуры, по сути, является проблемой для нахождения укладки с наименьшей энергией для заданной последовательности. Очевидно, что одинаково важными являются как сам алгоритм поиска структуры с наименьшей энергией, так и функция, которая задает эту энергию. Первым делом остановимся на энергетической функции.

### Энергетическая функция

Одним из показателей свернутости структуры является максимально плотная упаковка гидрофобных участков молекулы внутри белка и образование водородных связей. За первое ответственен эффект гидрофобных взаимодействий, за второй – Ван-дер-Ваальсовские взаимодействия. Так же не стоит забывать, что при сворачивании структуры разрушаются связи воды с полярными участками молекулы и эта потеря в энергии должна быть компенсирована образованием новых водородных связей. Хорошая энергетическая функция должна, так или иначе, отражать эти факторы стабильности молекулы.

В Розетте за упаковку отвечают атом-атом взаимодействия, вычисляемые с помощью потенциала Леонарда-Джонса, «implicit solvation model» – для описания гидрофобных эффектов и эффекта десольватирования остатков при их упаковке, а также потенциал для водородных связей в явном виде. Отдельные исследования [1-3] показывают, что явное задание функции для водородных связей (что не характерно для большинства моделей) дает преимущество в более аккуратной ориентации остатков при взаимодействии. Для сравнения различных энергетических функций можно ознакомиться с работой [4]. Если вдаваться в детали, то в Розетте существует две энергетических функции: простая и сложная. Первая применяется для построения общего вида молекулы, тогда как вторая нужна для более точной подгонки. Этот подход оправдывается тем, что вторая функция достаточно тяжела для вычислений и потребляет много ресурсов. Ниже приведены составляющие этих функций из оригинальной работы. Детальный разбор этих функций выходит за рамки работы, но можно отметить, что в простой функции, например, не задан явно потенциал для описания гидрофобных связей, а используется вероятностное описание паттернов для  $\beta$ -тяжей и  $\alpha$ -спиралей. Тогда как в сложной функции водородные связи заданы уже явно.

TABLE I  
COMPONENTS OF ROSETTA ENERGY FUNCTION<sup>a</sup>

Name	Description (putative physical origin)	Functional form	Parameters (values)
$env^b$	Residue environment (solvation)	$\sum_i -\ln [P(aa_i nb_i)]$	$i$ = residue index $aa$ = amino acid type $nb$ = number of neighboring residues <sup>c</sup> (0, 1, 2... 30, >30)
$pair^b$	Residue pair interactions (electrostatics, disulfides)	$\sum_i \sum_{j>i} -\ln \left[ \frac{P(aa_i, aa_j s_{ij}d_{ij})}{P(aa_i s_{ij}d_{ij})P(aa_j s_{ij}d_{ij})} \right]$	$i, j$ = residue indices $aa$ = amino acid type $d$ = centroid–centroid distance (10–12, 7.5–10, 5–7.5, <5 Å) $s$ = sequence separation (>8 residues)
$SS^d$	Strand pairing (hydrogen bonding)	SchemeA : $SS_{\phi, \theta} + SS_{hb} + SS_d$ SchemeB : $SS_{\phi, \theta} + SS_{hb} + SS_{d\sigma}$ where $SS_{\phi, \theta} = \sum_m \sum_{n>m} -\ln [P(\phi_{mn}, \theta_{mn} d_{mn}, s_{mn})]$ $SS_{hb} = \sum_m \sum_{n>m} -\ln [P(hb_{mn} d_{mn}, s_{mn})]$ $SS_d = \sum_m \sum_{n>m} -\ln [P(d_{mn} s_{mn})]$ $SS_{d\sigma} = \sum_m \sum_{n>m} -\ln [P(d_{mn}\sigma_{mn} \rho_m, \rho_n)]$	$m, n$ = strand dimer indices; dimer is two consecutive strand residues $V$ = vector between first N atom and last C atom of dimer $m$ = unit vector between $\hat{V}_m$ and $\hat{V}_n$ midpoints $x$ = unit vector along carbon–oxygen bond of first dimer residue $y$ = unit vector along oxygen–carbon bond of second dimer residue $\phi, \theta$ = polar angles between $\hat{V}_m$ and $\hat{V}_n$ (36° bins) $hb$ = dimer twist, $\sum_{k=m,n} 0.5( \hat{m} \cdot \hat{x}_k  +  \hat{m} \cdot \hat{y}_k )$ (< 0.33, 0.33–0.66, 0.66–1.0, 1.0–1.33, 1.33–1.6, 1.6–1.8, 1.8–2.0) $d$ = distance between $\hat{V}_m$ and $\hat{V}_n$ midpoints (< 6.5 Å) $\sigma$ = angle between $\hat{V}_m$ and $\hat{M}$ (18° bins) $sp$ = sequence separation between dimer-containing strands (< 2, 2–10, > 10 residues) $s$ = sequence separation between dimers (>5 or >10) $\rho$ = mean angle between vectors $\hat{m}, \hat{x}$ and $\hat{m}, \hat{y}$ (180° bins)

sheet <sup>e</sup>	Strand arrangement into sheets	$-\ln [P(n_{\text{sheets}}n_{\text{lonestands}} n_{\text{strands}})]$	$n_{\text{sheets}}$ = number of sheets $n_{\text{lonestands}}$ = number of unpaired strands $n_{\text{strands}}$ = total number of strands
HS	Helix-strand packing	$\sum_m \sum_n -\ln [P(\phi_{mn}, \psi_{mn} sp_{mn}d_{mn})]$	$m$ = strand dimer index; dimer is two consecutive strand residues $n$ = helix dimer index; dimer is central two residues of four consecutive helical residues $\hat{V}$ = vector between first N atom and last C atom of dimer $\phi, \theta$ = polar angles between $\hat{V}_m$ and $\hat{V}_n$ ( $36^\circ$ bins) sp = sequence separation between dimer-containing helix and strand (binned < 2, 2-10, >10 residues) $d$ = distance between $\hat{V}_m$ and $\hat{V}_n$ midpoints (< 12 Å)
rg	Radius of gyration (vdw attraction; solvation)	$\sqrt{\langle d_{ij}^2 \rangle}$	$i, j$ = residue indices $d$ = distance between residue centroids
cbeta	C $\beta$ density (solvation; correction for excluded volume effect introduced by simulation)	$\sum_i \sum_{sh} -\ln \left[ \frac{P_{\text{compact}}(\text{nb}_{i,sh})}{P_{\text{random}}(\text{nb}_{i,sh})} \right]$	$i$ = residue index sh = shell radius (6, 12 Å) nb = number of neighboring residues within shell <sup>f</sup> $P_{\text{compact}}$ = probability in compact structures assembled from fragments $P_{\text{random}}$ = probability in structures assembled randomly from fragments
vdw <sup>g</sup>	Steric repulsion	$\sum_i \sum_{j>i} \frac{(r_{ij}^2 - d_{ij}^2)^2}{r_{ij}}; d_{ij} < r_{ij}$	$i, j$ = residue (or centroid) indices $d$ = interatomic distance $r$ = summed van der Waals radii <sup>h</sup>

TABLE II  
COMPONENTS OF ROSETTA ALL-ATOM ENERGY FUNCTION<sup>a</sup>

Name	Description (physical origin)	Functional form	Parameters
rama	Ramachandran torsion preferences	$\sum_i -\ln [P(\phi_i, \psi_i aa_i, ss_i)]$	$i$ = residue index $\phi, \psi$ = backbone torsion angles ( $36^\circ$ bins) aa = amino acid type ss = secondary structure type <sup>b</sup>
LJ <sup>c</sup>	Lennard-Jones interactions	$\sum_i \sum_{j>i} \begin{cases} \left[ \left( \frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}}{d_{ij}} \right)^6 \right] e_{ij}, & \text{if } \frac{d_{ij}}{r_{ij}} > 0.6 \\ \left[ -8759.2 \left( \frac{d_{ij}}{r_{ij}} \right) + 5672.0 \right] e_{ij}, & \text{else} \end{cases}$	$i, j$ = residue indices $d$ = interatomic distance $e$ = geometric mean of atom well depths <sup>d</sup> $r$ = summed van der Waals radii <sup>e</sup>
hb <sup>f</sup>	Hydrogen bonding	$\sum_i \sum_j (-\ln [P(d_{ij} h_j ss_{ij})])$ $-\ln [P(\cos \theta_{ij} d_{ij} h_j ss_{ij})]$ $-\ln [P(\cos \psi_{ij} d_{ij} h_j ss_{ij})]$	$i$ = donor residue index $j$ = acceptor residue index $d$ = acceptor-proton interatomic distance $h$ = hybridization ( $sp^2, sp^3$ ) ss = secondary structure type <sup>g</sup> $\theta$ = proton-acceptor-acceptor base bond angle $\psi$ = donor-proton-acceptor bond angle
solv	Solvation	$\sum_i \left[ \Delta G_i^{\text{ref}} - \sum_j \left( \frac{2\Delta G_i^{\text{free}}}{4\pi^{3/2}\lambda_j r_{ij}^2} e^{-d_{ij}^2/V_j} \right) + \frac{2\Delta G_i^{\text{free}}}{4\pi^{3/2}\lambda_j r_{ij}^2} e^{-d_{ij}^2/V_i} \right]$	$i, j$ = atom indices $d$ = distance between atoms $r$ = summed van der Waal radii <sup>e</sup> $\lambda$ = correlation length <sup>h</sup> $V$ = atomic volume <sup>h</sup> $\Delta G^{\text{ref}}, \Delta G^{\text{free}}$ = energy of a fully solvated atom <sup>h</sup>

pair	Residue pair interactions (electrostatics, disulfides)	$\sum_i \sum_{j>i} -\ln \left[ \frac{P(aa_i, aa_j d_{ij})}{P(aa_i d_{ij})P(aa_j d_{ij})} \right]$	$i, j =$ residue indices aa = amino acid type $d =$ distance between residues <sup>i</sup>
dun	Rotamer self-energy	$\sum_i -\ln \left[ \frac{P(\text{rot}_i \phi_i, \psi_i)P(aa_i \phi_i, \psi_i)}{P(aa_i)} \right]$	$i, j =$ residue indices rot = Dunbrack backbone-dependent rotamer aa = amino acid type $\phi, \psi =$ backbone torsion angles
ref	Unfolded state reference energy	$\sum_{aa} n_{aa}$	aa = amino acid type $n =$ number of residues

Стоит отметить, что энергетическая функция высокой точности не является необходимой для хороших предсказаний, так же, как хорошие предсказания не гарантируют стоящей за ними хорошей энергетической функции. Это объясняется тем, что между «правильной» структурой и остальными конформациями должен существовать существенный разрыв в энергии, таким образом, энергетической функции достаточно «чувствовать» этот разрыв и укладываться в определенные пределы точности. Показано, что этот разрыв обычно составляет около 3-10 ккал/моль [5]. И это если говорить только о свободной энергии. Если же добавить еще и энтропийный фактор, то разрыв увеличивается. Экспериментальные и теоретические исследования дают оценку порядка 100 ккал/моль [6, 7]. Таким образом, энергетические функции имеют запас точности в многие ккал/моль, что, правда, не значит, неважность уточнения максимального уточнения функции получения правильной модели. Например, высокая точность нужна для оценки двух близких конформаций, разница между которыми может быть всего в несколько ккал/моль.

# De novo предсказания структур с помощью Rosetta

## Выбор фрагментов

Базовая операция для модификации конформаций в Rosetta называется “fragment insertion”. Для каждого фрагмента выбирается «окно» из трех или девяти остатков и их торсионные углы меняются на углы, полученные из белка с известной структурой. Для этих целей формируется специальная библиотека, в которой хранятся данные по торсионным углам фрагментов длиной в 3 или 9 нуклеотидов. Эта база формируется на основе рентген-структурных экспериментов с разрешением лучше 2,5 А. Все 3 и 9 остаточные окна в последовательности оцениваются против всех окон в базе данных известных белков, где сходство последовательностей с данной превышает 50%. Для выравнивания последовательностей используется несколько раундов PSIBLAST'a. Так же, предсказанная вторичная структура последовательности сравнивается с последовательностью из базы данных. Предсказания делаются с помощью трех методов: Psipred, SAM-T99 и JUF0. Полученные значения торсионных углов сравниваются с допустимыми значениями и ошибочные предсказания исключаются.

После вышеописанных итераций формируется список из лучших возможных конформаций последовательности (в него добавляются как найденные структуры, так и лучшие предсказанные). В итоге формируется 200 девяти- и 200 трех-остаточных фрагментов для каждого из перекрывающихся друг с другом окон в исходной последовательности.

## Сборка фрагментов

Сборка фрагментов в белок-подобную структуру осуществляется с помощью поиска Монте-Карло. Процесс начинается с полностью развернутого остатка. Случайным образом выбирается фрагмент из 9 остатков и случайным же образом выбирается одна из 25 лучших структур, полученных на предыдущем шаге. После подстановки структуры измеряется энергия полученной структуры. При следующем шаге применяется стандартный Metropolis критерий: если изменения уменьшило энергию, то оно принимается за новое базовое, если увеличило, то оно принимается или отвергается с определенной вероятностью. Если за 150 ходов не была принята ни одна структура, то эта вероятность ощутимо увеличивается (но после принятия следующей структуры вероятность возвращается в нормальное состояние). Каждая симуляция начинается со случайного участка и перебирает 28,000 девяти фрагментных подстановок. Стоит отметить, что не все из этих 28,000 подстановок однотипны. На разных этапах используются разные модификации оценочной функции, начиная от ее самых простых вариантов и заканчивая наиболее сложными. После этого производится так же 8000 подстановок трех-остаточных фрагментов с использованием



полной оценочной функции. Производится большое количество таких запусков с разных «зародышей» (seeds), затем этот набор кластеризуется по схожести для определения определенных регионов с минимальной энергией. Обычно правильная структура находится где-то в середине самого большого кластера.

Для последующей оптимизации структуры существует несколько базовых подходов, которые ведут не к значительным изменениям структуры, как вышеописанные действия, а к небольшим поправкам и улучшениям.

Random Angle Perturbation - самый простой подход для локальной оптимизации структуры – это случайные повороты углов. Случайно выбираются либо пары углов  $\phi$  и  $\psi$ , либо выбирается только угол  $\phi$ , а предшествующий ему угол  $\psi$  изменяется для компенсации поворота. Разрешены повороты не более чем на 3 градуса. Запрещены изменения углов в  $\alpha$ -спиралях, а изменения углов в  $\beta$ -тяжах ограничены 2 градусами. Так же, повороты, которые увеличивают Ramachandran score энергетической функции, отбираются с использованием критерия Metropolis.

Selection of Globally Nonperturbing Fragments: A Local Move. Второй подход так же основан на подстановке фрагментов, но в данном случае подставляются только фрагменты, максимально похожие на фрагменты из модели. В данном случае, существует два подхода для оценки схожести. В первом из них происходит замена всех атомов фрагмента модели на атомы из базы и подсчет схожести по методу наименьшего квадрата. Во втором, каждому фрагменту приписывается 6 степеней свободы таким образом, чтобы они зависели только от конформации фрагмента, но не от его положения в структуре. В этом случае в базе ищется фрагмент со схожими значениями этих степеней.

Torsion Angle Variation to Offset Global Perturbation and Monte Carlo Plus Minimization. По сути это две вспомогательные операции. Их суть заключается в определенной подгонке торсионных углов фрагмента, в котором производились изменения другим методом, а так же торсионных углов участка остова прилегающего к фрагменту. Для оценки изменений используется потенциал Рамачандрана (смотри таблицу с составляющими энергетической функции) и метод наименьших квадратов для первого метода и полная энергетическая функция для второго. С подробностями методов можно ознакомиться, например, тут [8].

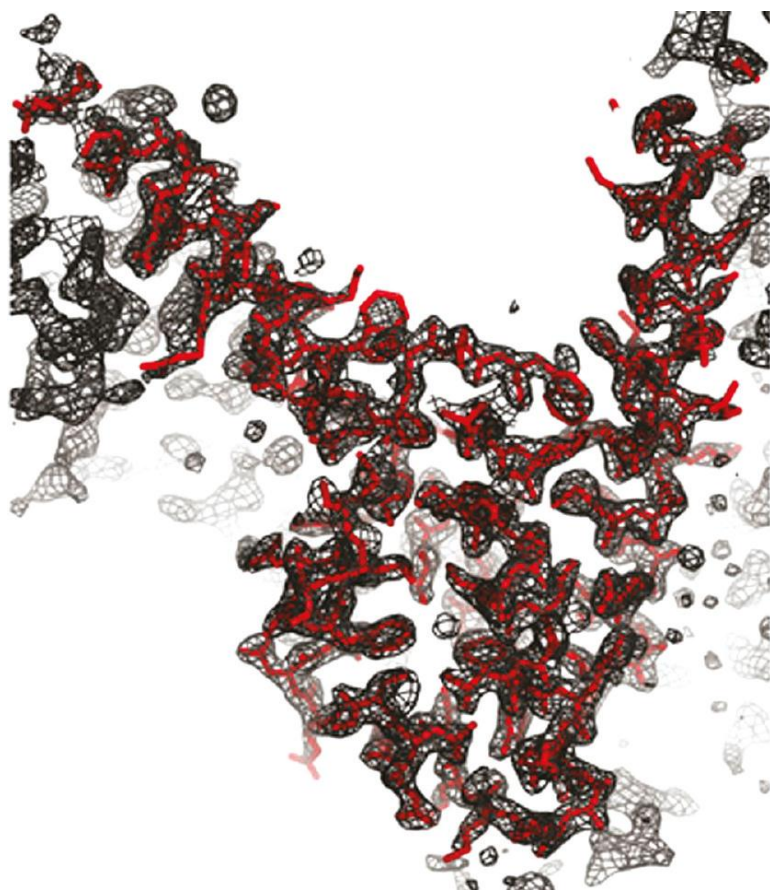
Rapid Side-Chain Optimization. Название метода говорит само за себя. Предыдущие методы оптимизировали структуру остова, этот оптимизирует боковые радикалы, меняя их конформацию и вычисляя энергетическую функцию после каждого изменения.

## Примеры успешных применений

Ниже будет представлено несколько примеров применения Розеттовских алгоритмов для решения разноплановых задач. Стоит отметить, что таких примеров очень много и ниже представлена случайная выборка из них.

### 1) Решение фазовой проблемы

С примером решения которой можно ознакомиться в работе «High-resolution structure prediction and the crystallographic phase problem» [9]



В особых комментариях работа не нуждается: с помощью Rosetta была получена модель белка, пригодная для метода молекулярного замещения. На (рис. 1) представлена структура структуры BHK3980 из *Bacillus halodurans*, полученная в данной работе. Красным отмечена модель, предсказанная с помощью Rosetta.

Рис. 1. Применения моделирования для метода молекулярного замещения.

## 2) Белковый докинг

- Многие белки образуют четвертичную структуру из нескольких цепей. Но в PDB банке для них часто представлена укладка только одной цепи. Таким образом, необходим инструмент для решения этой проблемы. Им может служить алгоритм RosettaDock. Суть алгоритма сводится к проведению сначала докинга белков с низким разрешением, а затем переход к более высокому разрешению и использованию алгоритмов коррекции боковых остатков. В одном из экспериментов было показано, что этот алгоритм может успешно восстановить 42 из 54 образцов белков, из которых была изъята одна из боковых цепей.[10]
- Так же было показано, что этот алгоритм пригоден для моделирования взаимодействия белок-антитела, что открывает большие возможности для дизайна лекарств.[18]

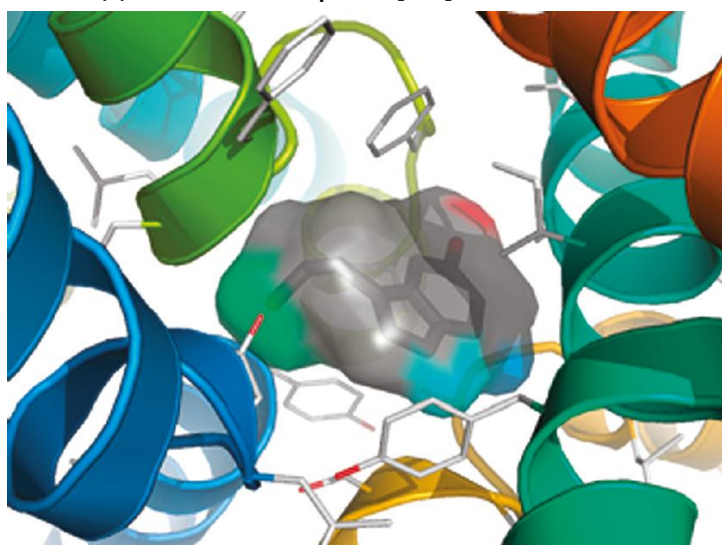


Рис. 2 Низкомолекулярный докинг с помощью Rosetta. Градиентными цветами обозначены расхождения предсказанной электронной плотности с экспериментальной, где цвет обозначает серьезность ошибки.

- Существует несколько модификаций алгоритма. Один из них RosettaLigand предназначен для докинга низкомолекулярных соединений. Его эффективность в одном из экспериментов была 80 из 100 правильно определенных белок-лигандных комплексов (см. рис. 2).[11]
- FOLD-AND-DOCK. Метод применяется для предсказания структур гомомеров белков. По сути, к обычной процедуре фолдинга добавляется rigid-body докинг. С помощью этого метода была предсказана структура R138F мутанта Rab6-GTP. [12]

### 3) Дизайн белков

- Очень интересным примером алгоритмов Rosetta может служить RosettaDesign, который решает обратную проблему: подбирает аминокислотную последовательность для того, чтобы белок складывался в заданную структуру или выполнял определенные функции. Основные направления применения данного алгоритма: дизайн новых укладок, редизайн существующих белков, создание белков с заданной поверхностью, дизайн ферментов. В качестве примера может служить дизайн белка с заданной укладкой, который затем был проверен экспериментально (см. рис. 3)[13].

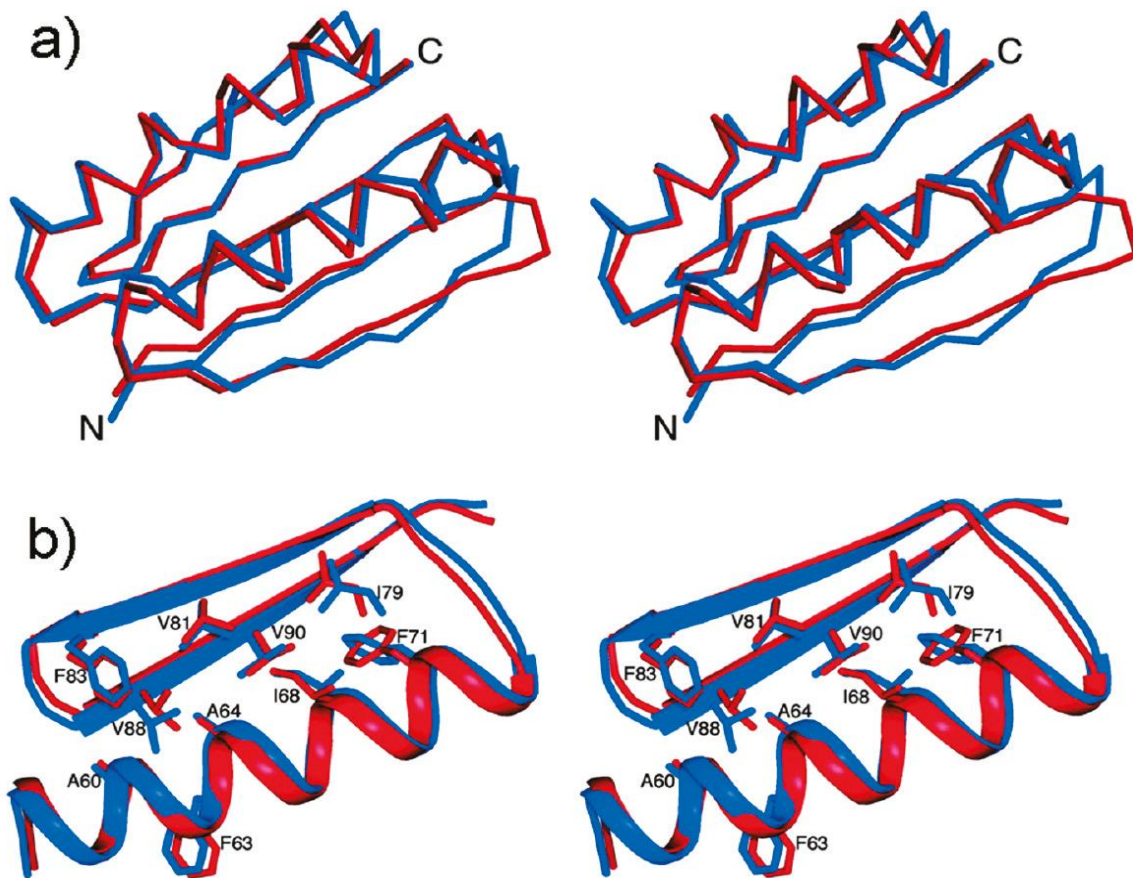


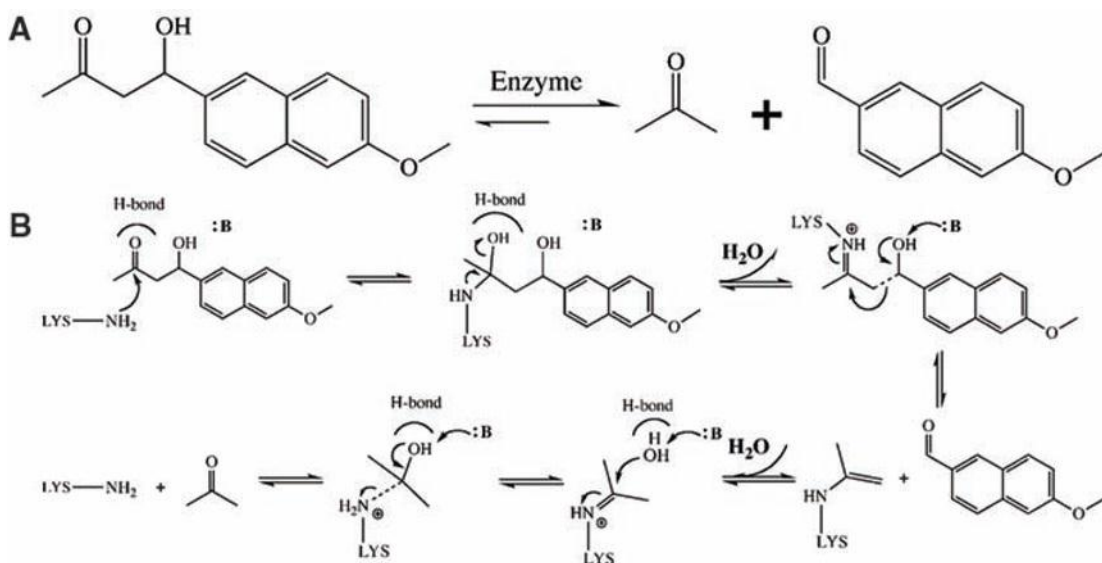
Рис.3 . Дизайн укладки белка. Красным показана экспериментально определенная структура, синим – предполагаемый дизайн. а) – общий вид белка, б) – ядро белка, где достигнута точность на атомном уровне.

- Интересным примером служит работа [14], где с помощью данного алгоритма были подобраны мутации в карбоксипептидазе, которые увеличивали стабильность белка на 5,2 ккал/моль
- Во втором интересном исследовании [15, 16] на эту тему удалось ощутимо увеличить специфичность эндонуклеазы, с помощью ее редизайна.

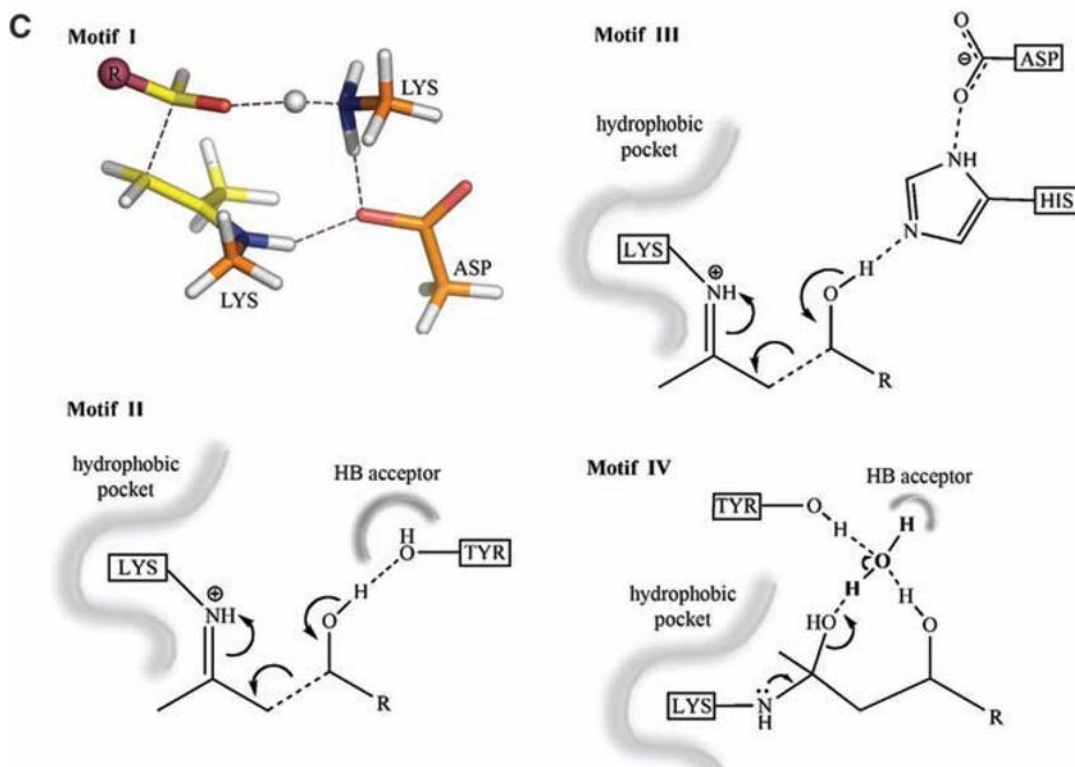
#### 4) Дизайн ферментов

Специальный алгоритм RosettaMatch предназначен для дизайна ферментов. Его суть состоит в том, что перебираются возможные положения активного сайта, а затем происходит попытка подогнать белок под выбранное положение. Крайне интересной является работа по de novo дизайну белка, катализирующего не встречающуюся в живых организмах реакцию: ретро-альдольное разрушение C-C связи в 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone. Всем интересующимся рекомендуется ознакомиться с работой непосредственно: «De Novo Computational Design of Retro-Aldol Enzymes» [17]. Здесь же будет описана лишь общая идея.

Первым шагом для подобной работы является определения одного или нескольких возможных каталитических механизмов, а так же возможных активных центров. В данном случае механизм реакции можно просуммировать следующей схемой:







Здесь А – общий ход реакции, Б – ее механизм. С – 4 возможных мотива активного центра.

Мотив 1: два лизина помещены рядом, для снижения рКа того из них, кто будет играть роль нуклеофила. Диада лизин-аспарагиновая кислота играет роль основания для депротонирования гидроксильной группы.

Мотив 2: Лизин помещен в hydrophobic pocket для снижения рКа и усиления нуклеофильных свойств. Тирозин играет роль основания.

Мотив 3: Аналогично, но роль основания играет His-Asp диада.

Мотив 4: Аналогично, но основная роль в катализе отдается молекуле воды. Используются такие остатки как Ser, Thr, Tyr для стабилизации воды в нужном положении.

Опуская подробности, из 181.555 возможных решений было отобрано 72 варианта, из которых 70 было выделено и очищено после экспрессии в бактериях. По результатам дальнейших исследований было выяснено, что 12 ферментов, имеющих первый тип активного центра, не показали ожидаемой активности, хотя 2 из них и связали субстрат. 10 белков второго типа так же не показали результатов, и из них только один смог связать субстрат. Из 14 энзимов третьего типа 10 связали субстрат и 8 обнаружили ожидаемую активность. И последний каталитический механизм, включавший в себя воду, показал наилучшие результаты. Из 36 белков 23 показали

ожидаемую активность. Увеличение скорости реакции достигало 4 порядков по сравнению с не катализируемой реакцией. Среди работающих ферментов были белки с разной структурой, например, представлены такие структуры как В-рулет и ТИМ-баррель, причем показано, что у этих ферментов отличается кинетика реакции.

## Заключение

В данном обзоре не ставилось цели всесторонне осветить такой подход к моделированию белков, как Rosetta. Его целью было показать общую идею, стоящую за этим методом, а так же продемонстрировать некоторые примеры его практического применения. За рамками осталось еще достаточно много сервисов и алгоритмов Rosetta, но и описанного достаточно для того, чтобы понять: компьютерное моделирование может дать крайне много как в плане теоретических, фундаментальных знаний, так и в практическом отношении, будь то моделирование ферментов для совершенно новых реакций, улучшение уже существующих белков или разработка новых лекарств.

# Список использованной литературы

В качестве основной работы по алгоритму Rosetta рекомендуется ознакомиться с

Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. (2004) Protein structure prediction using Rosetta. *Methods Enzymol.* 383, 66–93.

- 1 J. J. Gray, S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, and D. Baker, *J. Mol. Biol.* 331, 281 (2003).
- 2 C. A. Rohl, C. E. M. Strauss, D. Chivian, and D. Baker, *Proteins Struct. Funct. Genet.* In press (2004).
- 3 J. R. Gunn, *J. Chem. Phys.* 106, 4270 (1997). 13 R. Engh and R. Huber, *Acta Crystallogr. A* 47, 392 (1991).
- 4 R. L. Dunbrack and F. E. Cohen, *Protein Sci.* 6, 1661 (1997).
- 5 K. T. Simons, C. Kooperberg, E. Huang, and D. Baker, *J. Mol. Biol.* 268, 209 (1997).
- 6 K. T. Simons, I. Ruczinski, C. Kooperberg, B. Fox, C. Bystroff, and D. Baker, *Proteins Struct. Funct. Genet.* 35, 82 (1999).
- 7 T. Lazaridis and M. Karplus, *Proteins Struct. Funct. Genet.* 35, 133 (1999).
- 8 Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* 84, 6611 (1987).
- 9 Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J., and Baker, D. (2007) High-resolution structure prediction and the crystallographic phase problem. *Nature* 450, 259–264.
- 10. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., and Baker, D. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* 331, 281–299.
- 11. Roche, O., Kiyama, R., and Brooks, C. L., III (2001) Ligand-protein database: Linking protein-ligand complex structures to binding data. *J. Med. Chem.* 44, 3592–3598.
- 12. Burguete, A. S., Fenn, T. D., Brunger, A. T., and Pfeffer, S. R. (2008) Rab and Arl GTPase family members cooperate in the localization of the golgin GCC185. *Cell* 132, 286–298.
- 13. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302, 1364–1368.
- 14. Dantas, G., Corrent, C., Reichow, S. L., Havranek, J. J., Eletr, Z. M., Isern, N. G., Kuhlman, B., Varani, G., Merritt, E. A., and Baker, D. (2007) High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. *J. Mol. Biol.* 366, 1209–1221.
- 15. Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J., and Stoddard, B. L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell* 10, 895–905.
- 16. Joachimiak, L. A., Kortemme, T., Stoddard, B. L., and Baker, D. (2006) Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J. Mol. Biol.* 361, 195–208.
- 17. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. (2008) De novo computational design of retro-aldol enzymes. *Science* 319, 1387–1391.
- 18. Sircar, A., and Gray, J. J. (2010) SnugDock: Paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. *PLoS Comput. Biol.* 6, e1000644.