

**Отчёт о качестве расшифровки структуры мутантного комплекса
тиаминазы 2 с формил аминометил-пиримидином *Bacillus subtilis*
(PDB код 2QCX) методом рентгеноструктурного анализа**

**студентки четвёртого курса Факультета биоинженерии и
биоинформатики**

МГУ им. М. В. Ломоносова

Носиковой Екатерины

Москва, 2015

Аннотация

В отчёте рассмотрены некоторые индикаторы качества модели структуры 2QСХ, а также проведён анализ индикаторов локального качества структуры для ряда аминокислотных остатков. В ходе работы над отчётом удалось познакомиться с возможностями, которые предоставляют серверы для оценки качества структур белковых молекул.

Введение

ТенА (EC 3.5.99.2) – Тиаминаза 2, является уникальным семейством ферментов и катализирует 2 реакции, а именно, гидролиз тиамин (рис.1) и гидролиз 4-амино-5-аминометил-2-метилпиримидина (рис.2).

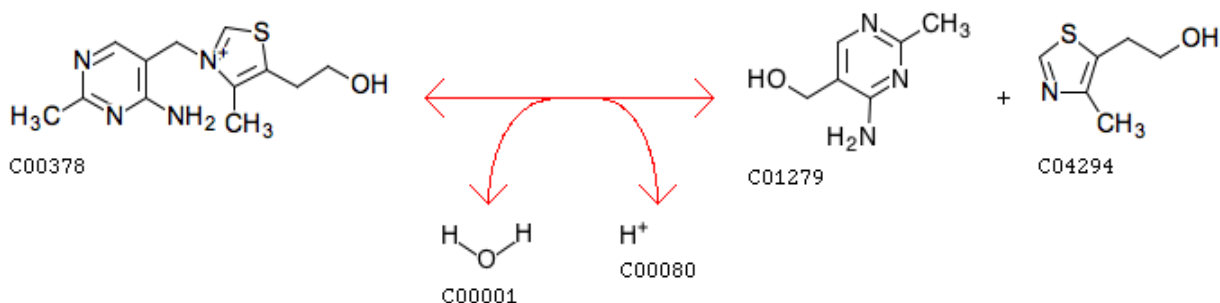


Рисунок 1. Реакция 1: Тиамин + H₂O ⇌ 4-амино-5-гидроксиметил-2-метилпиримидин+ 5-(2-гидроксиэтил)-4-метилтиазол + Н

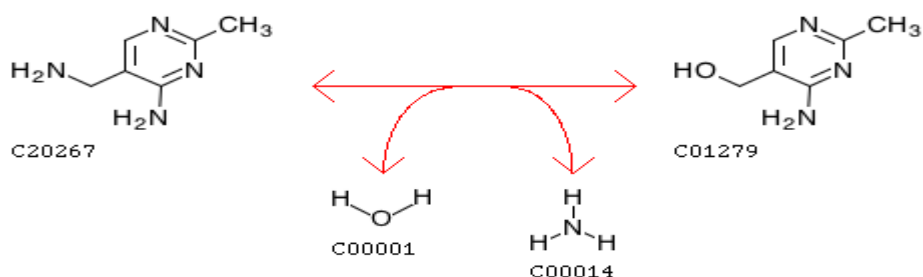


Рисунок 2: Реакция 2: 4-Амино-5-аминометил-2-метилпиримидин + H₂O ⇌ 4-амино-5-гидроксиметил-2-метилпиримидина + Аммиак

Семейство белков *TenA*, присутствующее у прокариот, растений и грибов, подразделяется на 2 подсемейства: ***TenA_C*** (в активном сайте которого содержится цистеин) и ***TenA_E*** (активный сайт не содержит цистеин).

Тиаминдифосфат является необходимым кофактором для таких ферментов как: транскетолаза, альфа-кетоглутарат дегидрогеназа и тд.

Тиамин содержит фрагменты НМР (4-амино-5-гидроксиметил-2 – метилпиримидин) и ТНЗ (4 - метил-5- (2 -гидроксиэтил) тиазол). Растения, грибы и большинство прокариотических организмов синтезируют оба из этих фрагментов *de novo*, и только затем происходит образование тиамин из НМР и ТНЗ.

Сохранение постоянного количества тиамин является важной и необходимой задачей. Известно, что при гидролизе тиамин образуется аминок- НМР. Было показано, что белки семейства *TenA* играют важную роль в поддержании баланса тиамин, путем гидролиза аминок- НМР до НМР, но, тем не менее, оставалось совершенно непонятно какое из подсемейств катализирует эту реакцию.

Авторами статьи [1] было выяснено, что именно *TenA_C* обладает активностью катализировать реакцию аминок- НМР до НМР (рис.1) *TenA_E* такой активности не проявляет. Считается, что эта реакция является ключевой для экономии затрат тиамин.

Также было выяснено, что экспрессия *TenA_C* и *TenA_E* генов значительно коррелирует.

Далее разговор пойдет только о *TenA Y112F* мутантном комплексе (pdb –код : 2QСХ) у которого присутствует остаток цистеина в активном центре.

Целью данной работы является анализ и оценка качества белка 2QСХ из организма Bacillus subtilis.

Результаты и обсуждение

Модель структуры 2QСХ была получена в 2007 году (**Zhang, Y., Jenkins, A.L., Begley, T.P., Ealick, S.E.**). Разрешение структуры – 2.2 Å.

Из данных EDS получено, что количество уникальных рефлексов 27575, количество использованных рефлексов равно 26547. Полнота данных равна 96.3 %.

Общая информация о модели 2QСХ систематизирована в таблице 1.

<i>Таблица 1. Общая информация о модели 2QСХ</i>	
Разрешение	2.2 Å
Диапазон разрешения	41.45 - 2.20 Å
Число измеренных рефлексов	27575
Число использованных рефлексов	26547
Полнота данных	96.3 %.
Пространственная группа	P 41 21 2
Параметры ячейки	a=58.62 Å, b=58.62 Å, c=296.47 Å alpha=90.00, beta=90.00, gamma=90.00

Для решения фазовой проблемы использовался метод молекулярного замещения (MR) с использованием структуры из того же семейства TenA из *Bacillus subtilis* – **1YAF**.

К показателям качества модели, приведенным в pdb-файле и статье, относятся: R-фактор, R-free, B-фактор и ранее описанные- разрешение, полнота. R-фактор – параметр, оценивающий, насколько полученная кристаллографическая модель соответствует набору экспериментально полученных рефлексов. Для данной структуры **R-фактор** равен **19.7%**, что меньше 25%, значит значение хорошее. **R-free** оказалось равным **23.3%**, это почти хороший результат, так как если говорить строго, то для хорошего Rfree должен быть меньше 20%. Для оценки качества также можно рассчитать коэффициент, равный разнице между R-free и R-фактором, если он не превышает 10%, то можно считать, что переоптимизации модели не произошло. В данном случае этот коэффициент равен **3.6%**, что подтверждает теорию об отсутствии переоптимизации. Также для оценки модели использовали карту Рамачандрана и пространственный R-фактор, Zscore для RSR.

Для построения карты Рамачандрана был использован сервер MolProbity . Результат анализа торсионных углов этим сервером представлен на рисунке 3 и содержит 6 карт (множество пар двугранных углов ϕ и ψ для всех типов аминокислотных остатков разбито для удобства на 6 групп). На карте находится 1 остаток, находящийся в запрещенной области – A 0 GLY (-64.3, -128.7) (значения в скобках соответствуют углам ϕ и ψ)

442 из 449 остатков (98.4%) остатков лежат внутри предпочитаемой области на карте Рамачандрана (Рис.3)

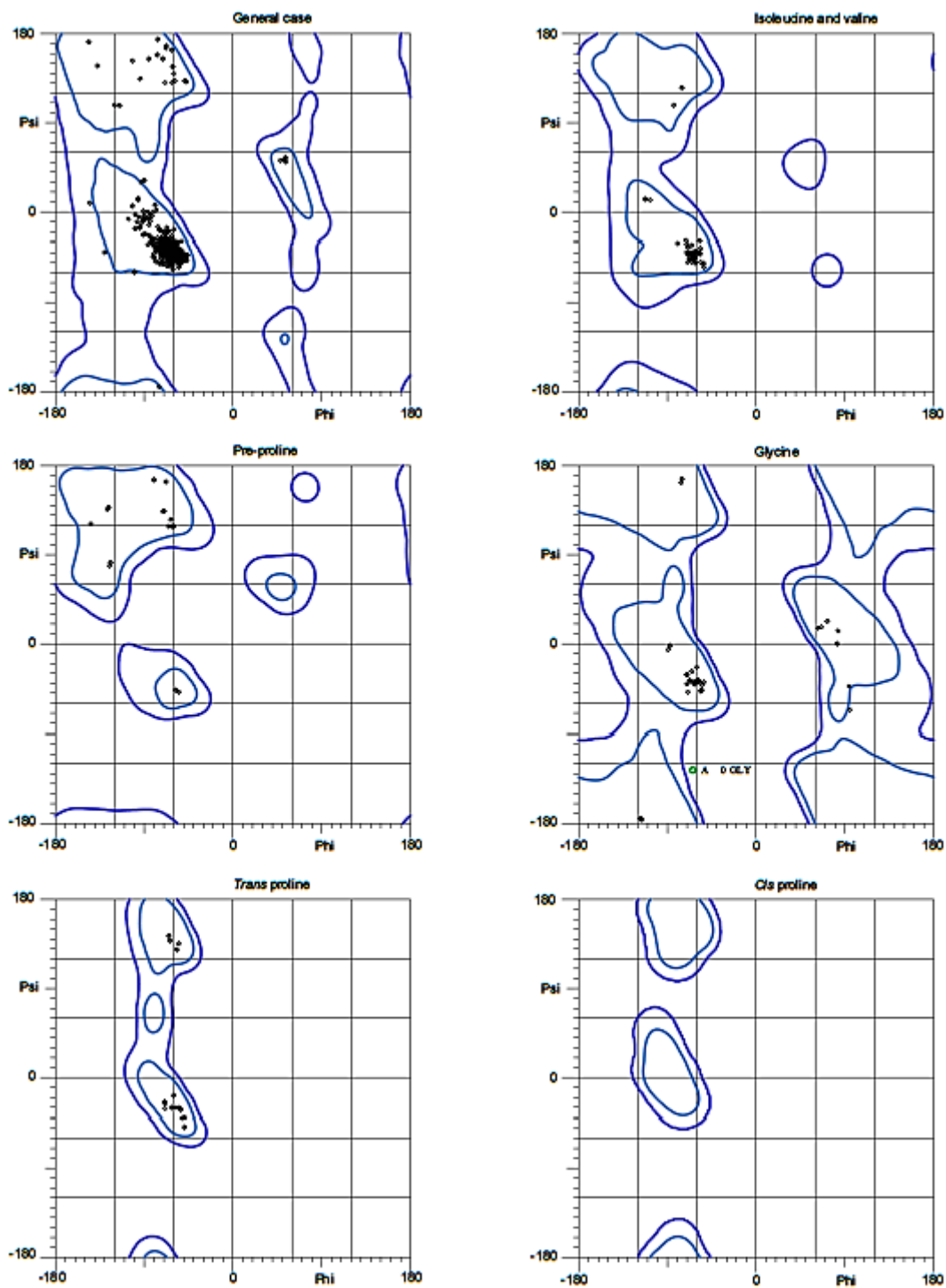


Рисунок 3: Карта Рамачандрана для структуры белка 2QCX. Карта построена сервисом MolProbity. На карте виден 1 маргинальный остаток.

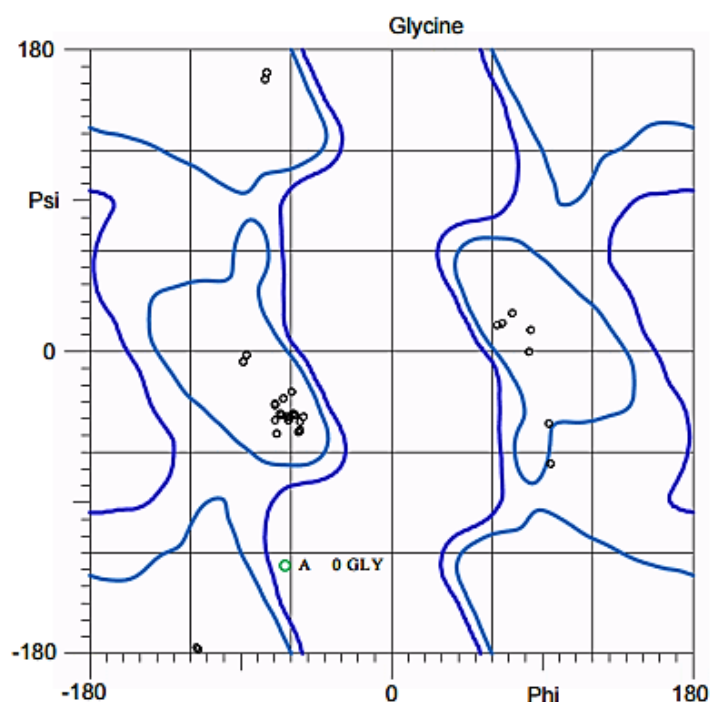


Рисунок 4: Фрагмент карты Рамачандрана для структуры белка 2QCX с маргинальным остатком A 0 GLY (-64.3, -128.7).

Общая информация о модели 2QCX систематизирована в таблице 1.

Таблица 2. Величины некоторых индикаторов качества структуры в целом для модели 2QCX

R-фактор	0.197
R_{free}	0.233
$R_{free} - R$	0.036
RSR-фактор (среднее значение и стандартное отклонение)	0.172 (0.076)
Число маргиналов по карте Рамачандрана	1
Число остатков в предпочитаемой области карты Рамачандрана	440 (98.43%)
Валентные углы, существенно отклоняющиеся от теории	0
Ковалентные связи, существенно отклоняющиеся от теории	1 (0.03%)
Отклонения $C\beta > 0.25 \text{ \AA}$	0
MolProbity score [^]	1.39

RSR – показатель того, насколько построенная модель соответствует « экспериментальной» электронной плотности. Для построенной модели равно 0.172 (0.076), что близко к 20% и не является хорошим показателем.

Если рассмотреть картину распределения RSR, для всех аминокислотных остатков всех белковых цепей (Рис.5), полученную с сервиса EDS, то, нетрудно заметить, что в цепи есть довольно много остатков, чей RSR > 0.2;

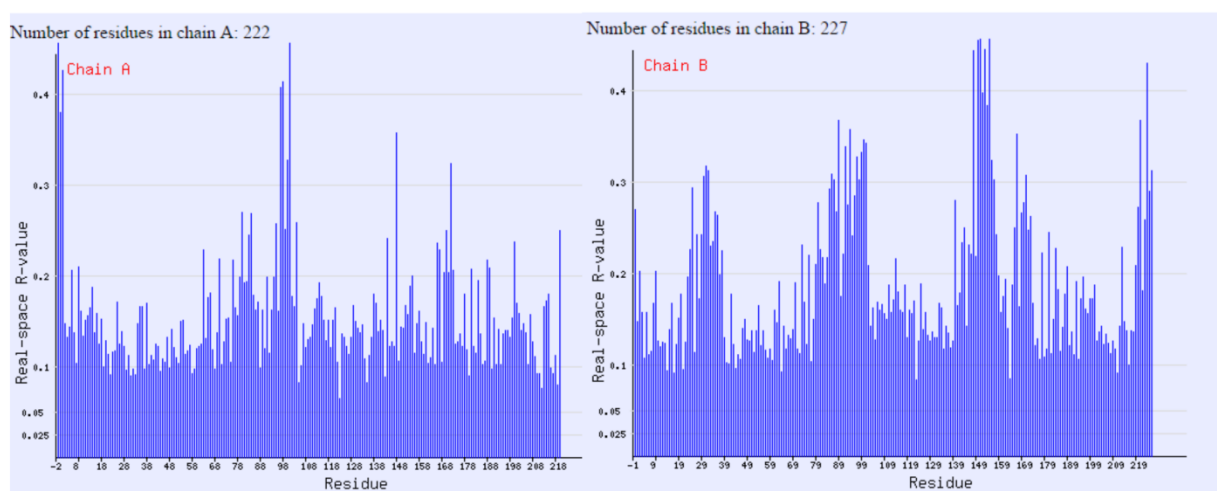


Рисунок 5. Распределение RSR для всех аминокислотных остатков всего белка, на каждой картинке графики для каждой цепи.

Z-score для RSR - параметр, позволяющий сравнивать RSR остатка модели со средним RSR в выборке моделей PDB с таким же разрешением. Проанализировав Z-score этой модели (Рис.6), можно сказать, что почти всех аминокислотных остатков белка (у 13.81% Z-score > 2) имеет RSR лучше, чем в других моделях с разрешением 2.00-2.20 Å. Среднее значение Z-score по всем остаткам равно -0.81.

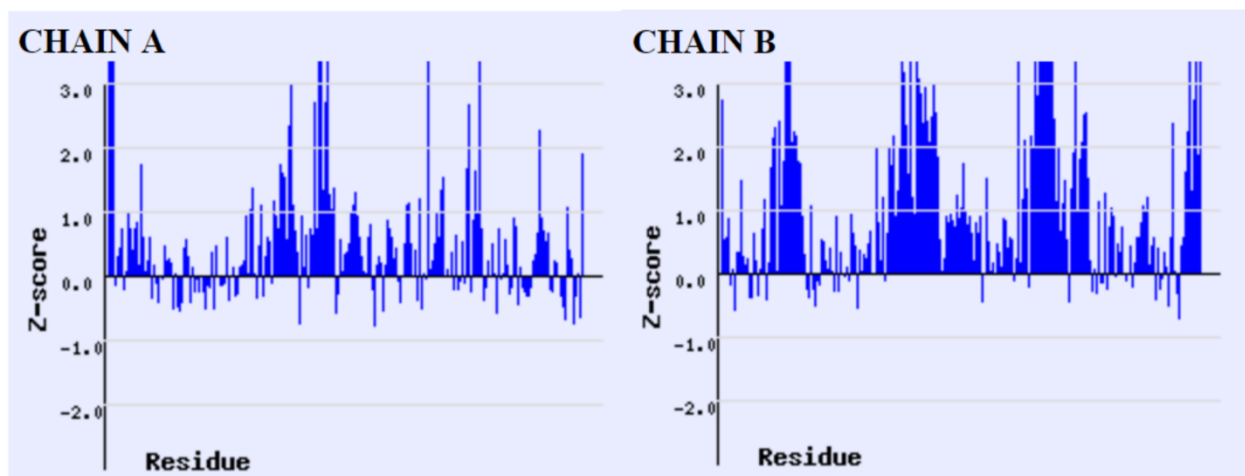


Рисунок 6. Изображение Z- score для цепи А, В.

Высокие положительные значения $Z > 2$ свидетельствуют о том, что остаток плохо вписан в электронную плотность по сравнению с другими структурами с тем же разрешением (=>маргинал)

Как видно из рисунка 6, в структуре присутствует довольно много маргинальных остатков.

Далее было выбрано 12 различных примеров маргинальных остатков.

Таблица 3. Примеры аминокислотных маргинальных остатков.

Маргинальный остаток	Критерий выбора
0 GLY A	В запрещенной области на карте Рамачандрана
47 TYR A 29 Asp B 103 Phe B	отклонение от нормы углов, образующих остов,
34 ILE A 87 ARG B	Нехарактерные торсионные углы (Zscore < -2)
136 TYR A	Угол tau (N-Calpha-C) имеет SD > 4σ
22 Pro A	Имеет "puckering" фазу с которой невозможно существовать внутри структуры белка
20 Val B	Zscore = -2.62
219 Ser A 220 ASP A 1 Met B	Нетипичная конформация боковых цепей (остатки могут являться частью странной петли)

Далее был произведен детальный анализ 5 маргинальных остатков.

I. Глицин-0 – маргинал по карте Рамачандрана.

Как показал анализ, глицин - 0 попадает в запрещенную область на карте Рамачандрана ($\phi = -64.3$, $\psi = 128.7$ 3). Эти остаток является первым в цепи А можно предположить инверсию пептидной цепи.

На Рисунке 7 изображен остаток и срез электронной плотности на уровне 1.5σ. Видно, что электронная плотность плохо описывает остаток глицина, поэтому, вероятно, действительно нужно провести инверсию вокруг пептидной связи на 180°.

Помимо нахождения в запрещенной области карты Рамачандрана, остаток глицина-0 также попадает в маргиналы по Zscore=5.51 (больше установленного порога 2).

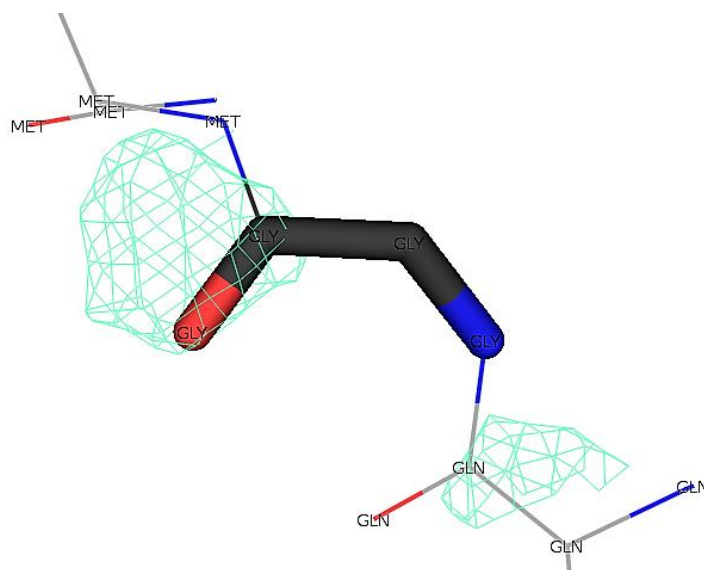


Рисунок 7. Изображение глицина-0 . Светло-голубым показан срез электронной плотности на уровне 1,5 σ

Таким образом, очевидно, что Gly 0 является маргинальным остатком.

11. Пролин-22 – согласно pdbreport имеет “ruckerling” фазу с которой невозможно существовать внутри структуры белка.

Было очень интересно, как выглядит такой остаток и как коррелирует со своей электронной плотностью. Было решено построить изображение на уровне подрезки 2 σ .

Изображение представлено на рисунке 8.

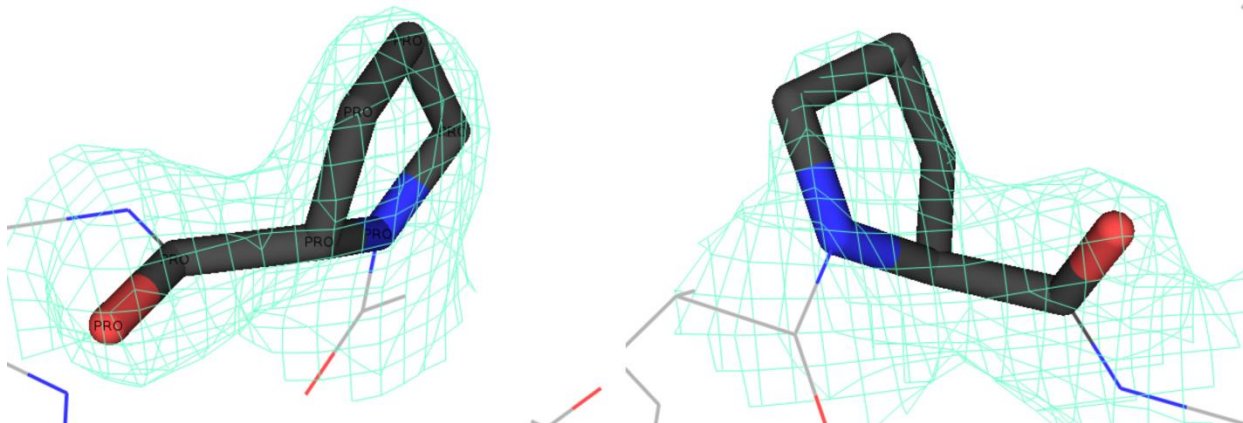


Рисунок 8. Слева маргинальный остаток Pro-22, справа типичный остаток Pro. Как видно из изображения Pro-22 действительно имеет нетипичную конформацию.

Далее был рассмотрен остаток 219 Ser A (нетипичная конформация боковых цепей). Была построена функция электронной плотности с уровнем подрезки 1.5 σ .

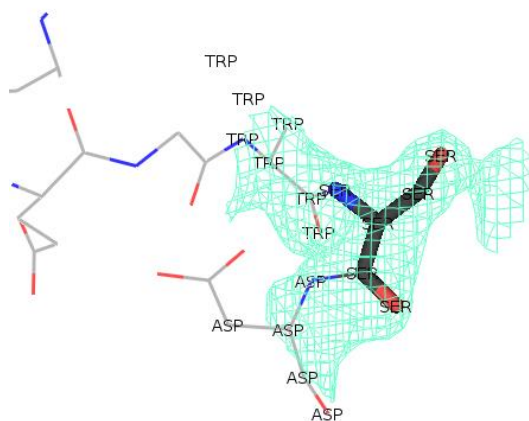


Рисунок 9. Остаток 219 Ser A в модели 2QCX и изображение электронной плотности (уровень подрезки 1.5 σ).

Рисунок 9 позволяет сделать вывод, что маргинальность Ser219 связана с ошибкой расшифровки и не является правдой.

Интересно также было посмотреть на 20 Val B, который является маргиналом из-за того, что обладает Zscore = -2.62. Была построена структура этого остатка и изображение электронной плотности с уровнем подрезки 2σ .

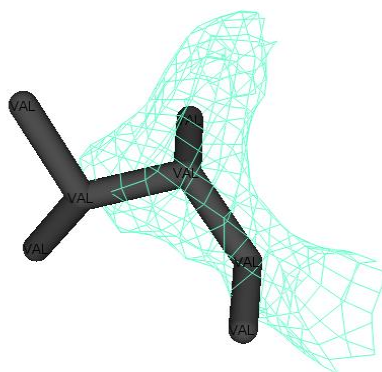


Рисунок 10. Остаток 20 Val B в модели 2QCX и изображение электронной плотности (уровень подрезки 2σ).

Видно, что ЭП плохо описывает остаток, что говорит о несоответствии модели ЭП.

Последним было решено проанализировать 136 TYR A, который является маргинальным, так как угол tau (N-Calpha-C) имеет $SD > 4\sigma$. На рисунке 11 можно увидеть изображение остатка и электронной плотности (уровень подрезки 2σ).

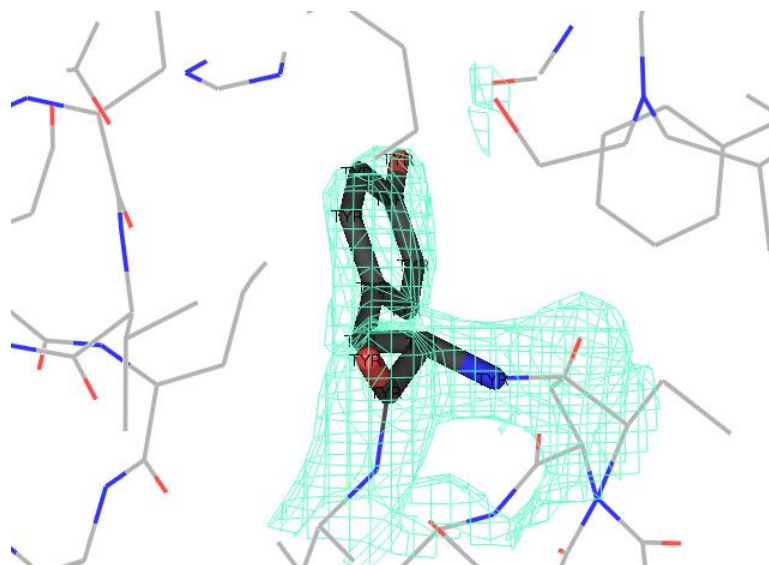


Рисунок 11. Остаток 136 TYR A в модели 2QCX и изображение электронной плотности (уровень подрезки 2σ).

Действительно, на рисунке видно, что угол tau (N-Calpha-C) имеет совсем не стандартное значение, но при этом ЭП довольно хорошо описывает остаток, что является удивительным.

Сравнение модели из PDB с моделью из PDB_redo

На сервере PDB REDO была построена структура 2qcx_final по имеющимся экспериментальным данным о 2qcx. В таблице 4 перечислены некоторые характеристики для 2qcx и 2qcx_final.

Значения R и R_{free} улучшились, и разница их также уменьшилась. Значит, новая модель лучше подогнана под экспериментальные данные. Число маргинальных остатков по карте Рамачандрана стало равно 0. Можно сделать вывод, что оптимизация структуры средствами PDB_REDO была проделана успешно.

Таблица 4. Величины некоторых индикаторов качества структуры в целом для модели 2QCX и 2QCX_REDO

Характеристика	2qcx	2qcx_final(pdb_redo)
R-фактор	0.197	0.1779
R_{free}	0.233	0.2072
$R_{free} - R$	0.036	0.0293
Число маргиналов по карте Рамачандрана	1	0
Число остатков в предпочитаемой области карты Рамачандрана	440 (98.43%)	445 (99.55%)
Валентные углы, существенно отклоняющиеся от теории	0	0
Ковалентные связи, существенно отклоняющиеся от теории	1 (0.03%)	0
Отклонения $C\beta > 0.25 \text{ \AA}$	0	0
MolProbity score [^]	1.39	1.26

Заключение

2QCX представляет собой модель хорошего разрешения. Качество её также можно оценить как «среднее», о чём свидетельствуют индикаторы оценки качества этой структуры. Из наблюдаемых в 2QCX аномалий можно отметить высокие показатели RSR-фактора и Zscore. В модели структуры удалось также обнаружить остатки, для которых предполагается инверсия боковых цепей (рисунок 7). Стоит отметить, что сервер PDB_REDO значительно улучшил показатели модели.

Список литературы

1. <http://www.ncbi.nlm.nih.gov/pubmed/25014715>
2. <http://www.ncbi.nlm.nih.gov/pubmed/18054064>
3. http://www.genome.jp/dbget-bin/www_bget?ec:3.5.99.2
4. <http://www.rcsb.org/pdb/explore.do?structureId=2qcx>
6. <http://molprobity.biochem.duke.edu/>
7. <http://www.cmbi.ru.nl/pdbreport/cgi-bin/nonotes?2QCX>
8. <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=2QCX>
9. http://www.cmbi.ru.nl/pdb_redo/uw/2qcx/