

**Отчёт о качестве расшифровки структуры
металл-зависимой карбоксипептидазы
Bacillus subtilis (PDB код 3HQ2) методом рентгеноструктурного анализа
студентки четвёртого курса
факультета биоинженерии и биоинформатики МГУ им. М.В. Ломоносова**

Бесединой Елизаветы

Москва, 2015

Аннотация

В отчёте рассмотрены некоторые индикаторы качества модели структуры 3HQ2, 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

Изучаемый белок – металлозависимая пептидаза урwA *Bacillus subtilis* – является предположительной карбоксипептидазой. Это значит, что она осуществляет гидролиз пептидной связи, действуя с С-конца пептида. Представляет собой димер, масса белка 58.175 кДа, длина обеих цепей 501 а.о. Белок связывает (рис. 1) множество лигандов: различные производные цинка (1 частица Zn_2Cl_2 , 2 молекулы $ZnCl_2$, 1 ион $ZnCl^-$, 2 иона $ZnPO_4^{2-}$) и 1 ион F^- .

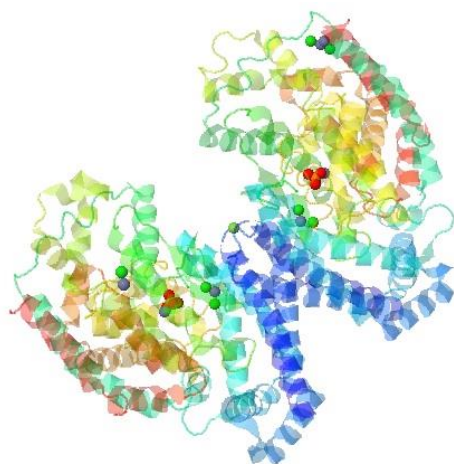


Рис.1. Лиганды в составе белка 3HQ2.
Zn - синие-голубые, F^- - светло-зеленый, Cl - зеленые, P^{+5} - оранжевые, O - красный.

В работе [1] была разрешена структура комплекса металлопротеазы с вышеописанными лигандами.

Белок относится к семейству M32-пептидаз. Для них характерен высоко консервативный участок HEXXH, играющий первостепенную роль в катализе гидролиза пептидной связи субстрата [2].

Интерес к семейству M32 появился еще довольно давно как к семейству, где предполагалось найти протеиназы, устойчивые к высоким температурам. Эти протеазы хотели использовать в масс-спектрометрии (когда не была известна технология MALDI).

Семейство M32 обширное и включает в себя много белков, гомологичных друг другу. Было выдвинуто предположение, что семейство естественно разделить на несколько подсемейств сообразно особенностям их структуры, каталитическим и субстратным особенностям. Именно с такой целью авторы статьи [1] решили произвести анализ структур двух разных белков из семейства M32. Для подтверждения существования внутри семейства двух подсемейств было решено произвести кристаллизацию и PCA для двух белков: BsuCP (ген *urwA*) и TthCP (ген *tthcp*).

Результаты и обсуждение

Общая информация о модели

Комплекс 3HQ2 состоит из двух одинаковых полипептидных цепей А и В в составе белка-димера. Каждый мономер связывает $ZnCl_2$, $ZnPO_4$; между субъединицами располагается ион F. Кроме того, в субъединице А дополнительно присутствует частица Zn_2Cl_2 , а в субъединице В – $ZnCl$.

Комплекс был расшифрован в 2009 году учеными M.M.Lee, C.E.Isaza, J.D.White, R.P.-Y.Chen, G.F.-C.Liang, H.T.F.He, S.I.Chan, M.K.Chan. Авторами было получено 32 740 рефлексов, из которых использовалось для оптимизации модели 31 847 рефлексов. Тестирование модели проводилось на 10% рефлексов. Для решения фазовой проблемы использовался метод молекулярного замещения с использованием карбоксипептидазы из *Leishmania major* (LmaCP – неопубликованные данные о ее структуре) [1].

- Разрешение данной структуры – 2.9 Å.
- Диапазон разрешений структурных факторов: 43.37 - 2.87 Å.
- Полнота набора структурных факторов - 94.6 %
- Параметры кристаллографической ячейки: a=88.96 Å, b=149.00 Å, c=217.84 Å
alpha=90.00, beta=90.00, gamma=90.00

Данные из источников [1], [3]

Значения индикаторов качества модели в целом

Значение R-фактора модели равно 0.188, а R-free – 0.251. R-фактор меньше 0.25 – это является показателем хорошего качества модели. Тем не менее, R-free больше 0.2.

Разница между значениями равна 6.3% - меньше 10%, то есть не доводов в пользу переоптимизации модели. Значение RSR равно 0.161 (со стандартным отклонением 0.052), что является показателем среднего качества модели (хороший показатель – <10%)

Для построения карты Рамачандрана был использован сервер MolProbity [4]. В результате обработки pdb файла сервис выдал информацию о множестве пар двугранных углов ϕ и ψ для всех типов аминокислотных остатков (6 групп) в виде карт (рис. 2), а также критерии качества для некоторых характеристик (табл. 1).

На карте есть 7 остатков, находящихся в запрещенной области: в субъединице А – Asp75, Asp255, Lys309, His462, Ile463; в субъединице В – Asp255 и His462.

MolProbity Ramachandran analysis

3hq2.pdb, model 1

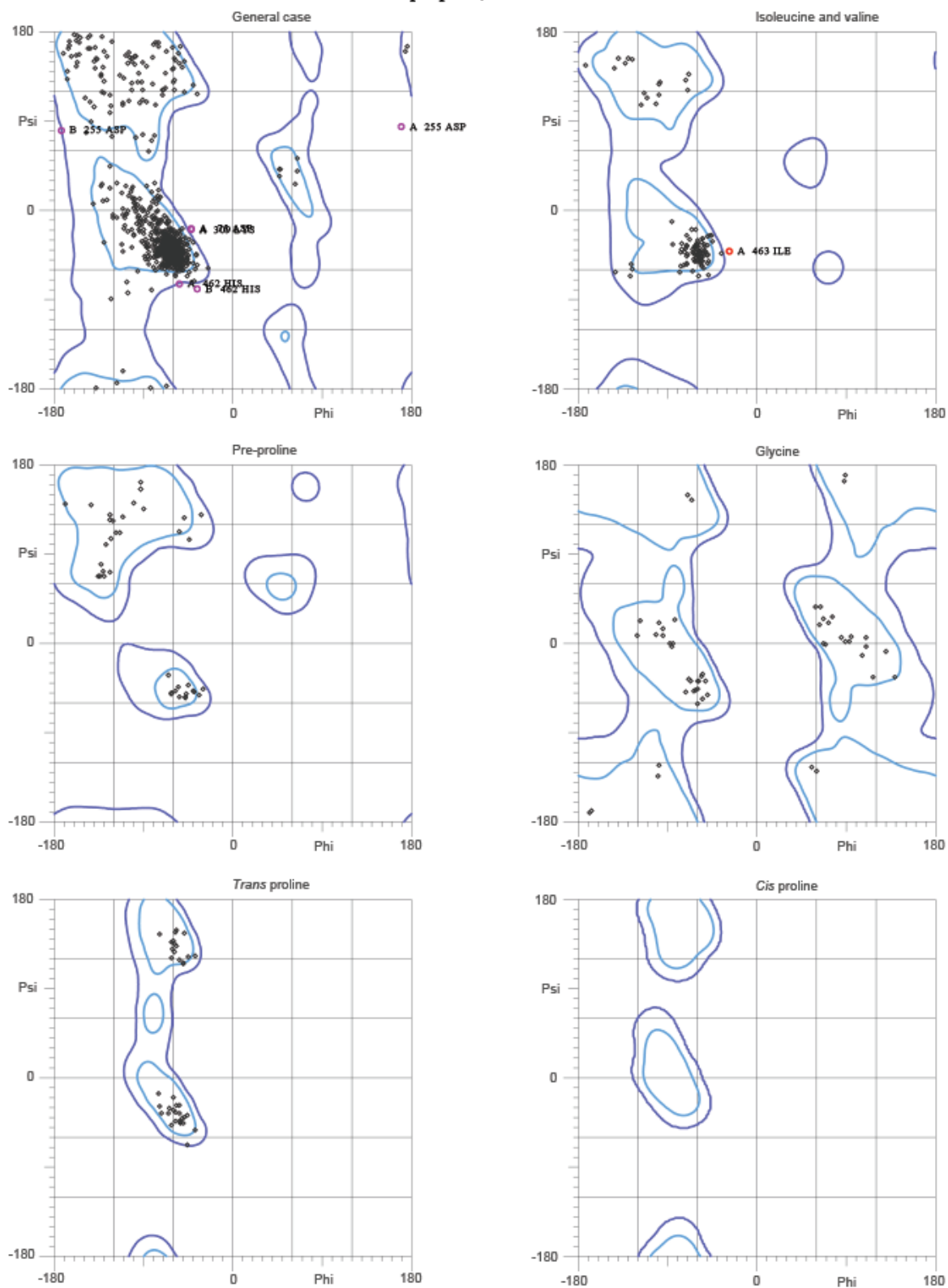


Рис. 2. Карта Рамачандрана для структуры белка 3hq2. Карта построена сервисом MolProbity. На карте видны 7 маргинальных остатка, находящиеся в запрещенной зоне углов – А: Asp75, Asp255, Lys309, His462, Ile463; В: Asp255 и His462.

Табл. 1. Значения критериев качества модели, предоставленных сервисом MolProbity.

Красным цветом выделены значения параметров, не удовлетворяющие условиям хорошей модели.

Protein Geometry	Poor rotamers	15	1.71%	Goal: <0.3%
	Favored rotamers	818	93.06%	Goal: >98%
	Ramachandran outliers	7	0.71%	Goal: <0.05%
	Ramachandran favored	923	93.33%	Goal: >98%
	C β deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 8371	0.00%	Goal: 0%
	Bad angles:	2 / 11323	0.02%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 38	0.00%	Expected: ≤ 1 per chain, or $\leq 5\%$
Low-resolution Criteria	CaBLAM outliers	5	0.51%	Goal: <1.0%
	CA Geometry outliers	0	0.00%	Goal: <0.5%

Из данных таблицы 1 нетрудно заметить, что в белке довольно много (0.71%) маргинальных остатков – больше допустимого в процентах значения (<0.05%). Это не очень хороший показатель качества модели.

Общая информация о модели 3HQ2 систематизирована в таблице 2.

Таблица 2. Величины некоторых индикаторов качества структуры в целом для модели 3HQ2. Зеленым помечены значения, удовлетворяющие нормам, красным – не удовлетворяющие, желтым – показатели со средними значениями.

R-фактор	0.188
R-free	0.251
R-free - R-фактор	6.3%
RSR (среднее значение и стандартное отклонение)	0.161 (0.052)
Число маргинальных остатков по карте Рамачандрана	7 (0.71%)
Число остатков в предпочитаемой области карты Рамачандрана	923 (93.33%)
«Слабые» ротамеры (число остатков, в нетипичных для данного типа остатков конформациях)	15 (1.71%)
Правильные ротамеры	818 (93.06%)
Отклонения угла C β	0
Ковалентные связи, существенно отклоняющиеся от теории	0
Валентные углы, существенно отклоняющиеся от теории	3 из 11323

Z-score для RSR - параметр, позволяющий сравнивать RSR остатка модели со средним RSR в выборке моделей PDB с таким же разрешением. Если Z-score остатка выше 2, значит, остаток вписывается в электронную плотность значительно хуже, чем остатки в структурах с таким же разрешением. Исходя из таких показателей, можно утверждать, что остаток является маргиналом (рис. 3).

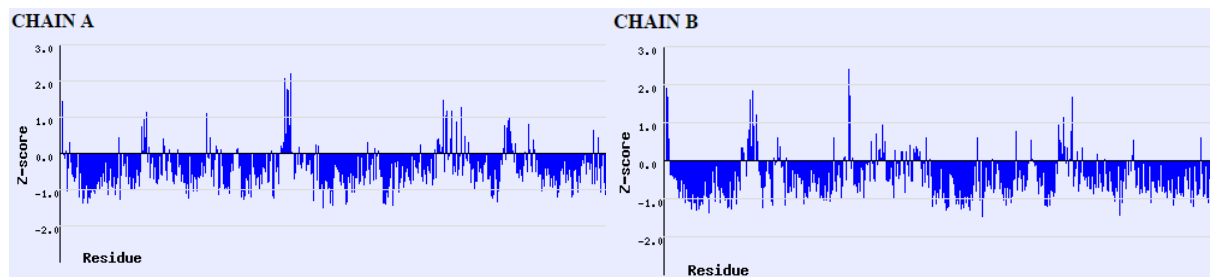


Рисунок 3. Значение Z- score для цепей А и В [3].

Для анализа маргинальных остатков использовались данные из карты Рамачандрана, сервиса EDS [3] и PDBREPORT [5]. Было выбрано 5 остатков для дальнейшего анализа.

Табл. 3. Примеры аминокислотных маргинальных остатков

Критерий	Маргинальный остаток
Находится вне разрешенной области на карте Рамачандрана	Asp75, цепь А
	Lys309, цепь А
RSR Z-score > 2	Gly150, цепь В
	Ala190, цепь А
	Pro195, цепь А
Максимальный температурный фактор (80.96)	Asn193, цепь А
Нехарактерные углы для ротамеров	Ser382, цепь В
	Ser213, цепь А

При подробном рассмотрении кандидатов в маргинальные остатки использовались все доступные источники информации.

Asp75, цепь A

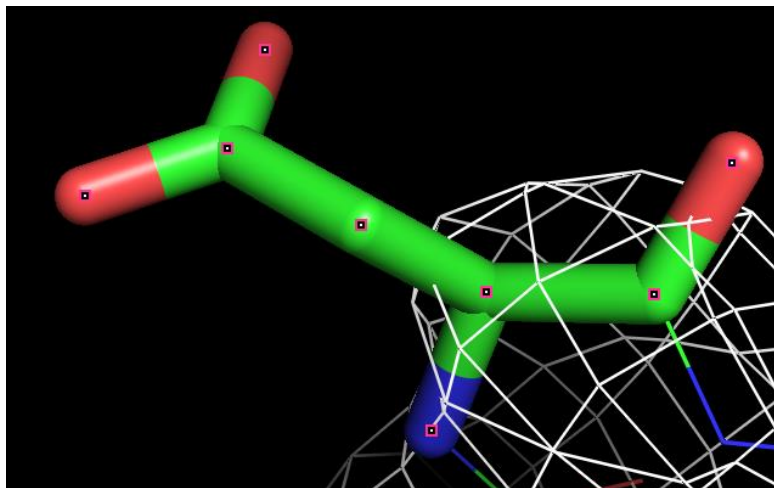


Рис. 4. Asp75, срезка электронной плотности на уровне 1.5σ

Для данного остатка торсионные углы принадлежат запрещенной области: ($\phi = -42.1$, $\psi = -18.5$). RSR для данного остатка равен 0.284 – и, как видно на рис. 3, остаток плохо вписывается в электронную плотность. Температурный фактор равен 69.27. Z-score = 1.14. На основании данных показателей видно, что маргинальность остатка – следствие недостаточной точности экспериментальных данных.

Lys309, цепь A

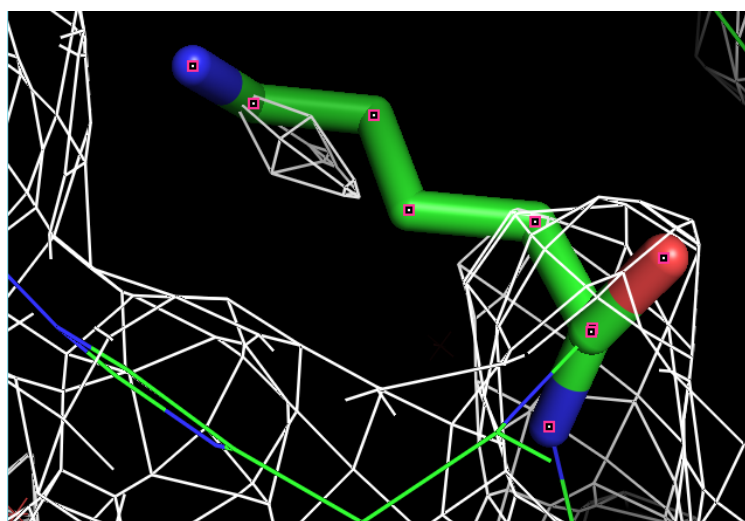


Рис. 5. Lys309, срезка электронной плотности на уровне 1.5σ

Для данного остатка торсионные углы принадлежат запрещенной области: ($\phi = -43.0$, $\psi = -19.6$). RSR для данного остатка равен 0.209. Температурный фактор равен 36.44. Z-score = -0.37. В отличие от предыдущей ситуации, основные показатели не очень велики, поэтому сделать вывод, что маргинальность остатка – следствие недостаточной точности

экспериментальных данных мы не можем. Возможно, маргинальность остатка – его особенность.

Ser382, цепь B

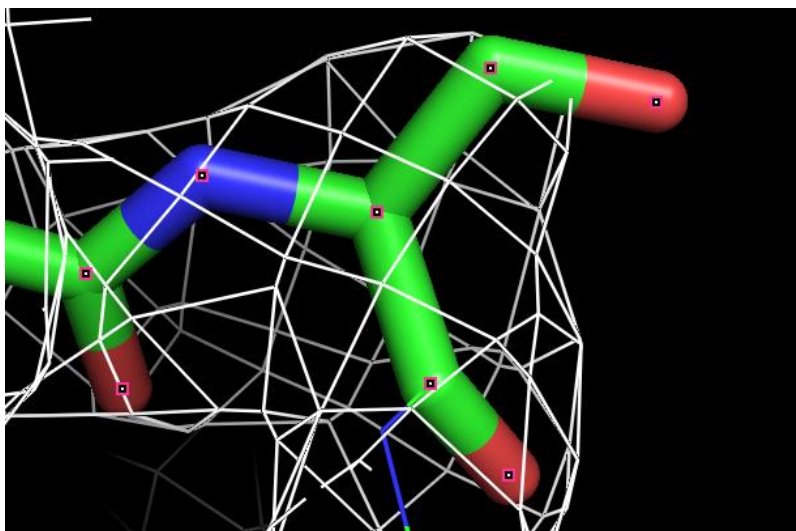


Рис. 6. Ser382, срезка электронной плотности на уровне 1.5 σ

Температурный фактор равен 27.74. Z-score = -0.2. Рисунок позволяет судить о том, что маргинальность данного остатка не является ошибкой расшифровки – остаток довольно хорошо вписывается в электронную плотность. Следовательно, как и в предыдущем случае, маргинальность остатка – не следствие ошибки расшифровки.

Pro195, цепь A

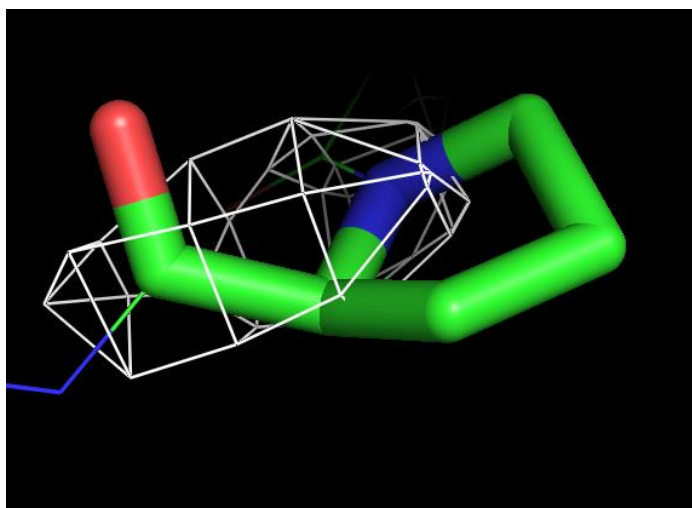


Рис. 7. Pro195, срезка электронной плотности на уровне 1.5 σ

Пролин плохо вписывается в электронную плотность - Z-score = 2.18. Температурный фактор равен 69.97. Поэтому, скорее всего, маргинальность остатка является ошибкой расшифровки.

Asn193, цепь А

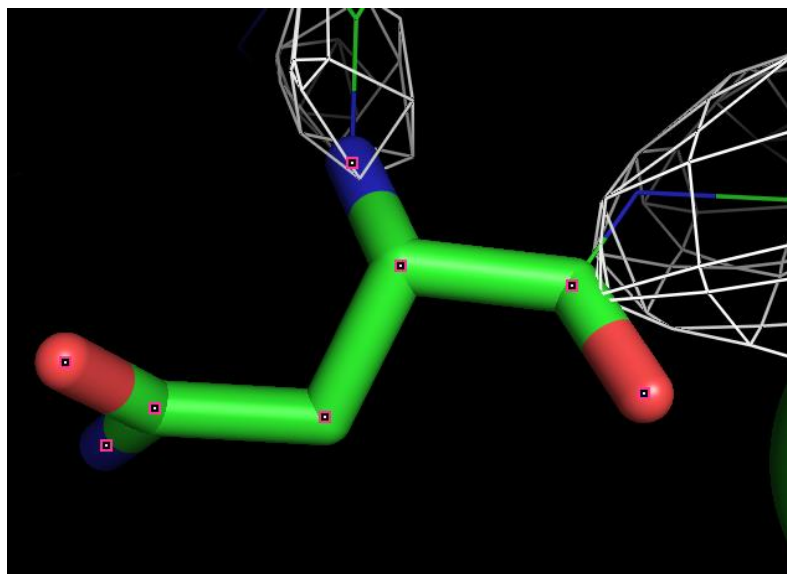


Рис. 8. Asn193, срезка электронной плотности на уровне 1.5σ

Как видно на рис. 8, для аспарагина практически отсутствует электронная плотность. Z -score = 1.71. Маргинальность остатка, основанная на высоком температурном факторе (80.96), является недостаточным показателем и скорее свидетельствует о низком качестве расшифровки модели.

Сравнение модели из PDB с моделью из PDB_redo

На сервере PDB_REDO была построена структура 3HQ2_final по имеющимся экспериментальным данным о 3HQ2. Данные в этой базе более новые и оптимизированные, что вызывает больше доверия. В таблице 4 перечислены некоторые характеристики для 3HQ2 и 3HQ2_final.

Таблица 4. Величины некоторых индикаторов качества структуры в целом для модели 3HQ2. Зеленым помечены значения, удовлетворяющие нормам, красным – не удовлетворяющие, желтым – показатели со средними значениями.

Параметр	3HQ2	3HQ2_final
R-фактор	0.188	0.1671
R-free	0.251	0.2079
R-free - R-фактор	6.3%	4.08%
Z-score	-2.89	-2.00

RSR (среднее значение и стандартное отклонение)	0.161 (0.052)	
Число маргинальных остатков по карте Рамачандрана	7 (0.71%)	2 (0.20%)
Число остатков в предпочитаемой области карты Рамачандрана	923 (93.33%)	963 (97.37%)
«Слабые» ротамеры (число остатков, в нетипичных для данного типа остатков конформациях)	15 (1.71%)	18 (2.05%)
Правильные ротамеры	818 (93.06%)	830 (94.43%)
Отклонения угла C β	0	1 (0.11%)
Ковалентные связи, существенно отклоняющиеся от теории	0	0
Валентные углы, существенно отклоняющиеся от теории	3 из 11323	1 из 11323
Unsatisfied H-bond donors/acceptors	66	52

Как видно на примерах из таблицы, почти все характеристики значительно улучшились (хотя и не всегда на столько, чтобы считаться идеальными). Таким образом, можно считать, что новая модель структуры 3NQ2 на основе старых данных имеет большее отношение к реальности, чем старая.

Заключение

Модель 3NQ2 является моделью среднего качества: при рассмотрении основных индикаторов качества модели было выявлено довольно много недостатков модели. Помимо большого количества остатков, находящихся вне допустимой части карты Рамачандрана, модель имеет большое количество неправильных ротамеров, показатель RSR слегка великоват.

Тем не менее, это лишь отчасти связано с недостаточными экспериментальными данными: построенная по тем же самым данным более оптимальным способом модель 3NQ2_final имеет более приемлемые характеристики, хотя и не всегда вписывается в общепринятые стандарты.

Список литературы

- [1] M. M. Lee, C. E. Isaza, J. D. White, R. P.-Y. Chen, G. F.-C. Liang, H. T.-F. He, S. I. Chan, and M. K. Chan, "Insight into the substrate length restriction of M32 carboxypeptidases: characterization of two distinct subfamilies.," *Proteins*, vol. 77, no. 3, pp. 647–57, 2009.
- [2] Electronic Source
<http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=188998>
- [3] Electronic Source
<http://eds.bmc.uu.se/eds/>
- [4] Electronic Source
<http://molprobitry.biochem.duke.edu/>
- [5] Electronic Source
<http://swift.cmbi.ru.nl/gv/pdbreport/>