

**Московский государственный университет им. М.В.Ломоносова
Факультет биоинженерии и биоинформатики**

**Отчёт по качеству расшифровки структуры белка 5A77, полученной с помощью
рентгеноструктурного анализа**

Выполнил:
студент 4-го курса
Котюргин Александр

Москва, 2018

Аннотация

В данной работе проводилась оценка расшифровки структуры самонаводящейся эндонуклеазы I-CvuI в комплексе с расщепленной ДНК, PDB ID 5A77. Рассмотрен ряд индикаторов качества модели, некоторые маргинальные остатки и оптимизированная структура из PDB_redo.

Введение

Комплекс 5A77 образован самонаводящейся эндонуклеазой I-CvuI (в форме гомодимера) с фрагментом двухцепочечной ДНК длиной 24 пары оснований, являющимся целью гомологичной самонаводящейся эндонуклеазы I-CreI [1]. В ДНК внесен одноцепочечный разрыв. В структуре присутствуют 2 иона Mg^{2+} .

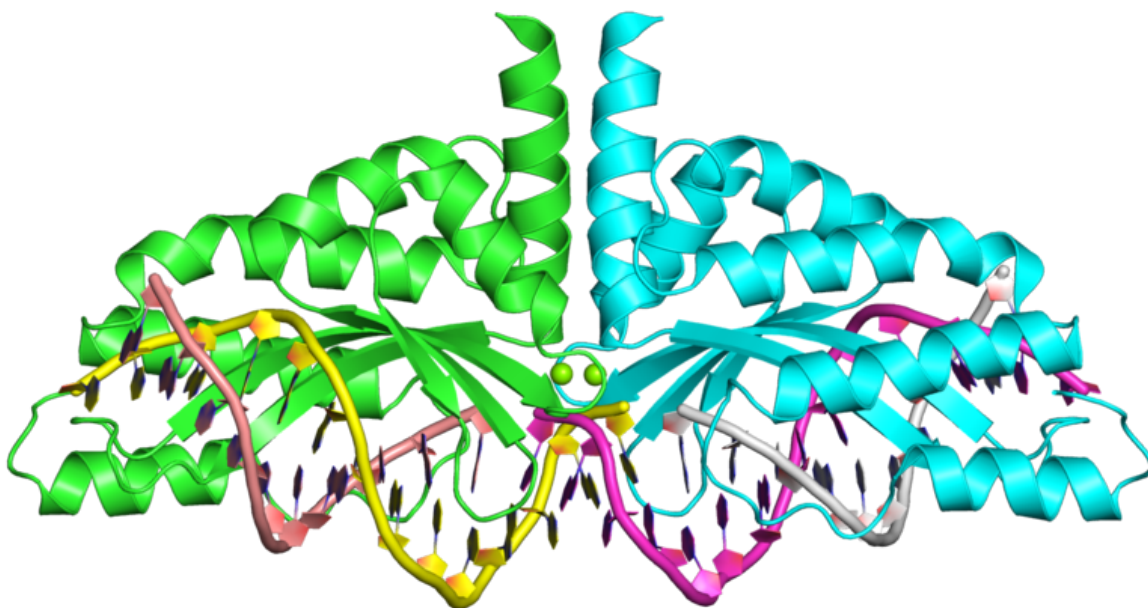


Рис. 1. Изображение структуры 5A77

Самонаводящиеся эндонуклеазы (homing endonucleases - не нашел устоявшегося перевода на русский, также известны как мегануклеазы) кодируются генами внутри интронов и распространяются путем внесения разрывов в специфические сайты, что провоцирует гомологическую рекомбинацию. Тем самым, эти эндонуклеазы могут быть интересны как фермент для генной инженерии.

I-CvuI относится к наиболее крупному семейству самонаводящихся эндонуклеаз LAGLIDADG, механизм действия которых описан в статье [2].

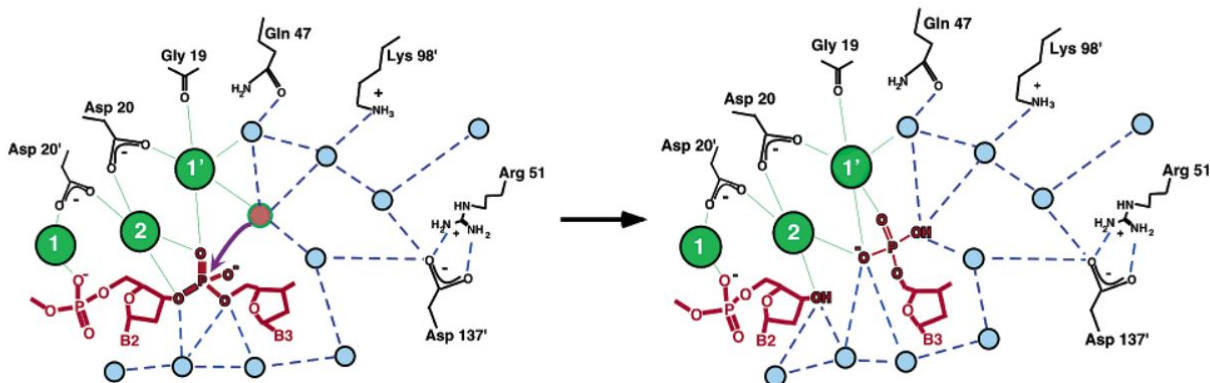


Рис. 2. Реакция протекает следующим образом: пара из ионов металла (обозначена как 1/1'), удерживаемая остатками аспарагиновой кислоты (Asp20/Asp20') координируют воду, которая осуществляет нуклеофильную атаку, в то время как ион металла 2 стабилизирует переходное состояние [2].

Ген I-CvuI находится в интроне гена 23S рПНК в хлоропластах *Chlorella vulgaris*.] В работе [1] авторы определяли структуры комплекса I-CvuI с его нативной целью Sro1.3 и с целью I-Crel эндонуклеазы (C1221), а также исследовали зависимость активности фермента от концентрации Mg^{2+} .

Результаты

Общая информация о структуре

Модель опубликована в 2015 году в статье “Crystal Structure of the Homing Endonuclease I-CvuI Provides a New Template for Genome Modification”, Rafael Molina et al. Разрешение - 2.50 Å, среднее. Диапазон разрешений - 47.54–2.50 Å. Число измеренных рефлексов - 15798, полнота данных 98.7%. Параметры кристаллической решетки: a=48.57 Å, b=95.08 Å, c=52.15 Å

$\alpha=90.00$, $\beta=103.71$, $\gamma=90.00$. Кристаллическая решетка относится к группе P 1 2₁ 1, некристаллографическая симметрия присутствует - C2.

R_{work} - 0.227 (согласно EDS[3]), 0.196 (согласно PDB[4]), R_{free} - 0.265 (согласно обоим источникам).

Фазовая проблема решена методом молекулярного замещения, замещающая структура - 1G9Z, комплекс мегануклеазы I-Crel с продуктом. Асимметрическая единица совпадает с биологической.

Также нужно отметить, что, согласно PDB Validation Report [5] 9% цепи А и 8% цепи В не моделировалось.

Качество модели

Начнем оценку модели с общих параметров R и R_{free} . При подсчете R_{free} использовалось 3094 рефлекса или около 20% от всех рефлексов. Значения, приведенные в EDS и PDB, различаются, но в обоих случаях можно сказать, что $(R_{free} - R) < 0.1$, а $R < 0.25$. R_{free} при этом больше 0.2, но меньше 0.4. Глядя на эти данные, можно сказать, что эта модель относится к хорошим не подвергнутым сверхоптимизации моделям.

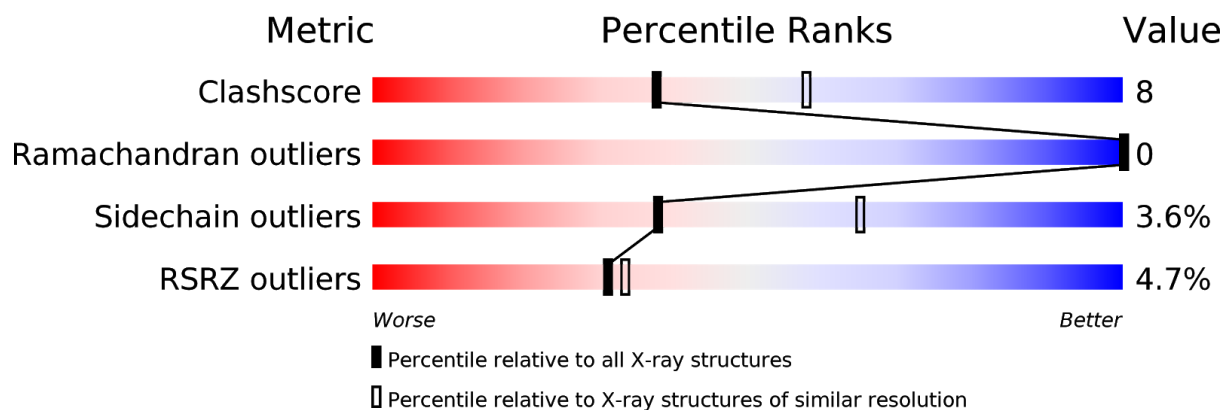


Рис. 3. Percentile plot для структуры 5A77 [5]

Рассмотрим percentile plot, взятый из полного отчета о валидации структуры. Можно отметить, что отсутствуют остатки, маргинальные по карте Рамачандрана (см. также рис. 4).. По остальным параметрам структура несколько хуже, чем в среднем по всем полученным методом PCA структурам, и хуже, чем большинство структур аналогичного разрешения по числу не вписывающихся в электронную плотность остатков (параметр RSRZ). Тем не менее, процент маргинальных остатков меньше 10, поэтому модель нельзя отнести к некачественным.

Также надо отметить, что, согласно EDS, в структуре присутствует 7 маргиналов по карте Рамачандрана, что находится в рамках нормы. Однако, данные EDS будут здесь и далее восприниматься как менее приоритетные, т.к. EDS не обновляется с 2017 года.

Исследуем качество модели с помощью сервиса MolProbity [6]. В ходе обработки боковые цепи остатков A/29/GLN и B/111/GLN были инвертированы согласно рекомендациям.

Таблица 1. Сводная информация о качестве модели 5A77 из MolProbity.

Clashscore, all atoms:	8.61		98 th percentile* (N=271, 2.50Å ± 0.25Å)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Poor rotamers	17	6.18%	Goal: <0.3%
Favored rotamers	228	82.91%	Goal: >98%
Ramachandran outliers	0	0.00%	Goal: <0.05%
Ramachandran favored	298	95.82%	Goal: >98%
MolProbity score [^]	2.35		84 th percentile* (N=6960, 2.50Å ± 0.25Å)
Cβ deviations >0.25Å	1	0.34%	Goal: 0
Bad bonds:	0 / 2504	0.00%	Goal: 0%
Bad angles:	1 / 3374	0.03%	Goal: <0.1%
Cis Prolines:	0 / 6	0.00%	Expected: ≤1 per chain, or ≤5%

Согласно данным MolProbity, 5A77 относится к хорошим моделям (84 перцентиль по показателю общего качества, 98 перцентиль по количеству атомов с недопустимым перекрытием ван-дер-ваальсовых радиусов. Главными проблемными пунктами является обилие маргинальных и недостаточно хорошо соответствующих табличным данным ротамеров. Также MolProbity выделяет, что предпочтительными по Рамачандрану является меньше 98% остатков, однако, возможно, порог в 98% является слишком строгим. Также к маргинальным можно отнести единственный остаток с высоким отклонением положения Cβ-атома боковой цепи и остаток с маргинальным углом. Так как для рассмотрения в рамках данного отчета маргинальных остатков достаточно много, я выделил следующие:

Таблица 2. Список выбранных для рассмотрения маргинальных остатков

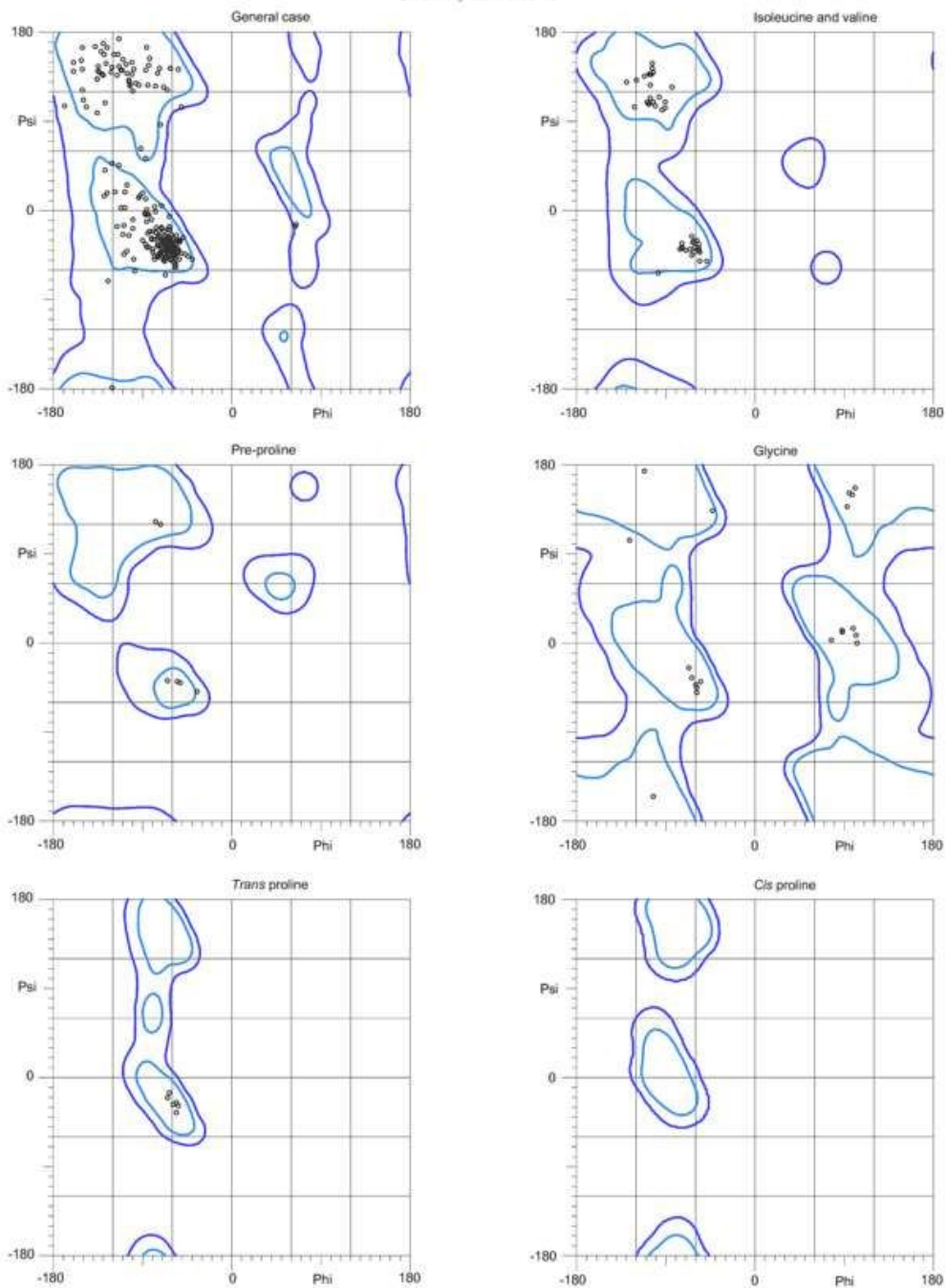
Остаток	Описание
A/7/PHE	Перекрытие 0.92Å A/10/GLN/HG2 с A/7/PHE/HA
A/10/GLN	
B/136/GLY	Перекрытие 0.83Å B/142/LYS/NZ с B/136/GLY/O

B/142/LYS	
A/152/ARG	Плохой ротамер, углы 132.9, 176.3, 253.5, 142.3
B/93/GLU	Плохой ротамер, углы 349.3, 204.7, 87.2
B/156/LYS	Плохой ротамер, углы 228.5, 117.2, 155, 137.6
B/144/ARG	Отклонение положения C β -атома 0.38Å
A/29/GLN	Отклонение угла от среднего N-CA-CB в 4.7 σ

Данные остатки являются либо наиболее подозрительными в своем классе (максимальное перекрытие радиусов, худшие ротамеры) либо единственными.

MolProbity Ramachandran analysis

5A77, model 1



95.8% (298/311) of all residues were in favored (98%) regions.

100.0% (311/311) of all residues were in allowed (>99.8%) regions.

There were no outliers.

Рис. 4. Карта Рамачандрана для структуры 5A77.

Маргинальные остатки

Перекрытие A/7/PHE и A/10/GLN

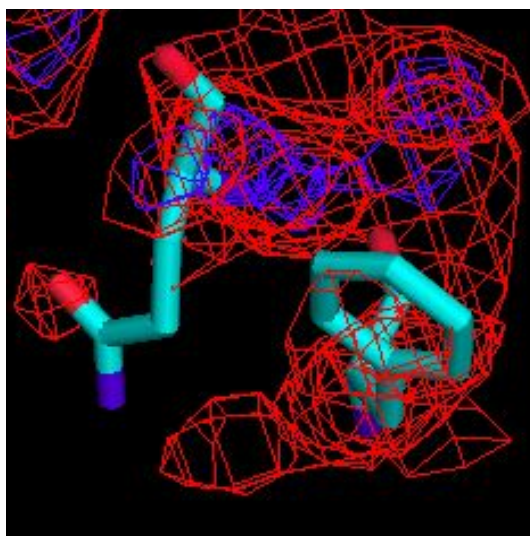


Рис. 5. Остатки A/7/PHE и A/10/GLN с различными уровнями электронной плотности (красный - 1 σ , синий - 2 σ , белый - 3 σ , здесь и далее обозначения аналогичные, если не указано другое)

Часть боковой цепи глутамина находится за пределами обозначенных поверхностей электронной плотности. Можно предположить, что это происходит из-за стягивания электронной плотности к кислороду глутамина. Если визуализировать поверхность уровня электронной плотности в 0.5 σ , можно увидеть подтверждение этому (см. рис. 6) - большая часть боковой цепи глутамина находится в пределах этой электронной плотности, за исключением азота. В то же время дальняя часть теоретически электрон-плотного ароматического кольца фенилаланина находится за пределами всех 4 поверхностей электронной плотности, что дает основание заподозрить ошибку расшифровки структуры/оптимизации, которая привела к большому перекрытию ван-дер-ваальсовых радиусов атомов водорода.

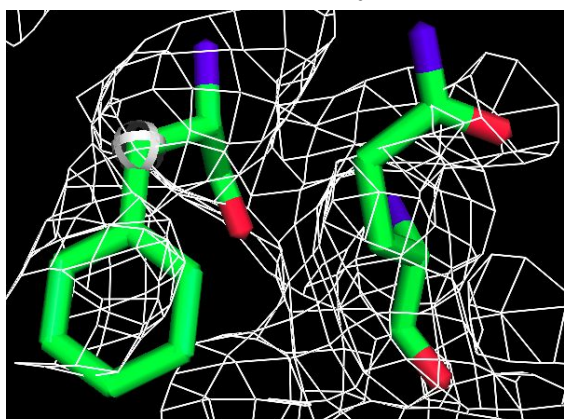


Рис. 6. Остатки A/7/PHE и A/10/GLN с уровнем электронной плотности 0.5 σ

Плохой ротамер A/152/ARG

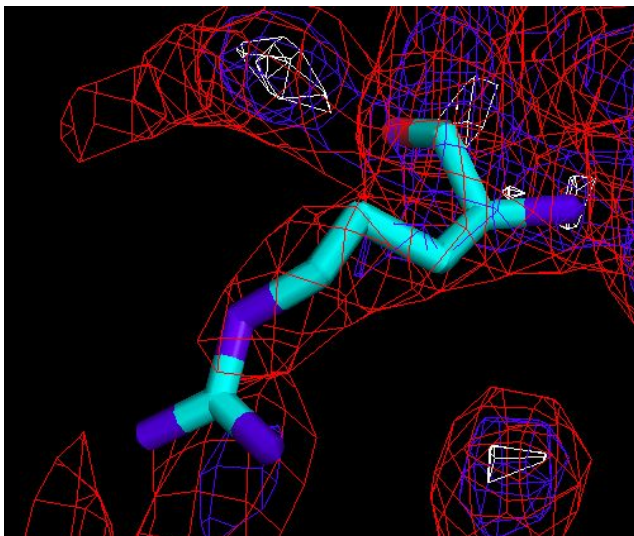


Рис. 7. Остаток A/152/ARG с различными уровнями электронной плотности
Боковая цепь аргинина практически целиком находится внутри поверхности электронной плотности, за исключением концевой части, где можно предположить наличие электрон-дефицитной области, аналогичной атому азота глутамина из предыдущего пункта. Скорее всего, здесь нет ошибки расшифровки.

Плохой ротамер B/93/GLU

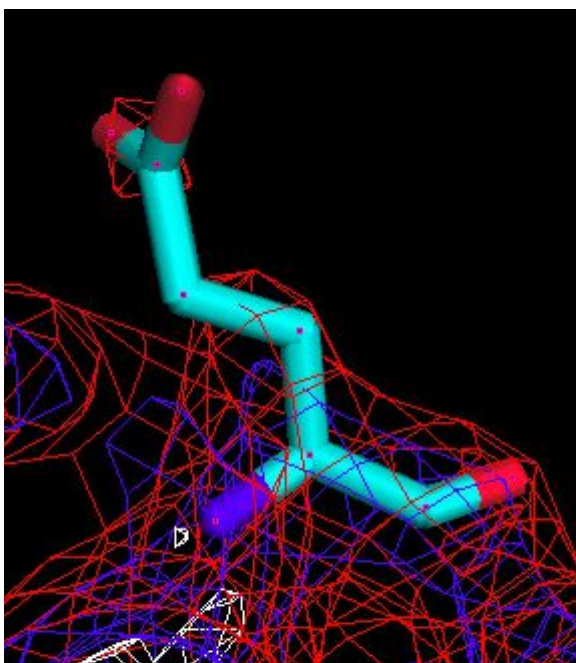


Рис. 8. Остаток B/93/GLU с различными уровнями электронной плотности
Здесь виден тот же самый эффект - максимум электронной смещен на кислороды, поэтому часть боковой цепи находится за пределами отмеченных поверхностей. Если рассмотреть поверхность электронной плотности 0.5σ , будет видно, что боковая цепь целиком лежит внутри.

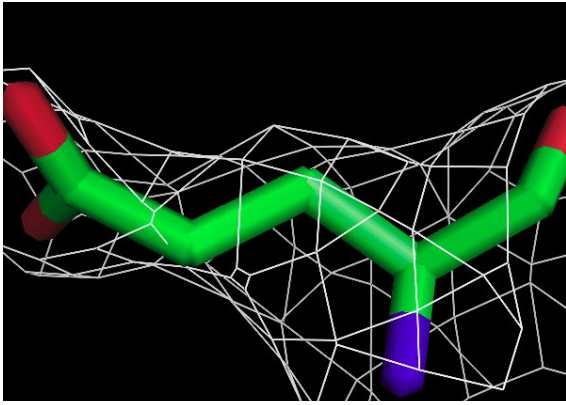


Рис. 9. Остаток B/93/GLU внутри уровня электронной плотности 0.5σ

Аномальное отклонение угла от среднего N-CA-CB в A/29/GLN

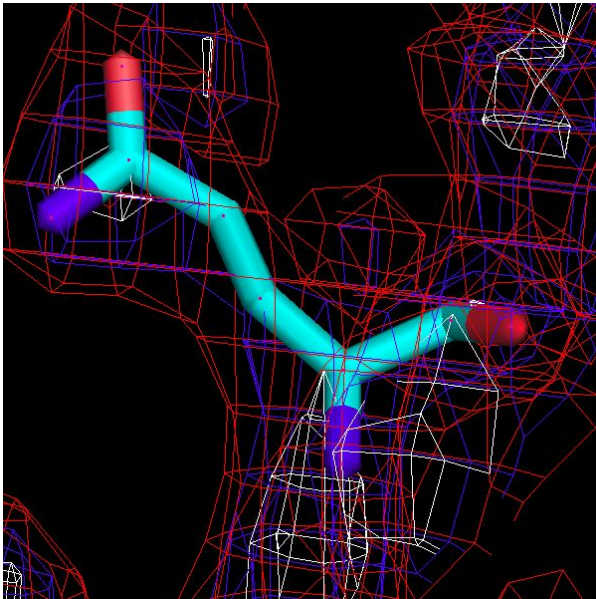


Рис. 10. Остаток A/29/GLN с различными уровнями электронной плотности
Данный остаток практически идеально вписан в электронную плотность по сравнению со всеми прочими рассмотренными остатками, что позволяет предположить, что здесь нет ошибки расшифровки. Связано ли это с функциональной ролью остатка, неясно, но этот остаток располагается на части белка, контактирующей с ДНК.

Структура из PDB_redo

Оптимизированная структура из базы PDB_redo [7] была загружена для анализа в сервис MolProbity и были получены следующие результаты:

Таблица 3. Сводная информация о качестве оптимизированной модели 5A77_redo из MolProbity.

All-Atom Contacts	Clashscore, all atoms:	2.8		100 th percentile* (N=271, 2.50Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	21	7.64%	Goal: <0.3%
	Favored rotamers	233	84.73%	Goal: >98%
	Ramachandran outliers	1	0.32%	Goal: <0.05%
	Ramachandran favored	305	98.07%	Goal: >98%
	MolProbity score [^]	1.74		99 th percentile* (N=6960, 2.50Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.34%	Goal: 0
	Bad bonds:	0 / 2504	0.00%	Goal: 0%
	Bad angles:	1 / 3374	0.03%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 6	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations*:	33	68.75%	Goal: ≤= 5%
	Bad bonds:	0 / 1096	0.00%	Goal: 0%
	Bad angles:	7 / 1686	0.42%	Goal: <0.1%

Оценка качества MolProbity относительно улучшилась (99 перцентиль), но в абсолютных единицах улучшение качества незначительно (0.61). Также увеличилось число как плохих, так и предпочтительных ротамеров, появились маргиналы по карте Рамачандрана. При этом серьезно уменьшилось число перекрываний ван-дер-ваальсовых радиусов атомов (с 8.61 до 2.8), увеличилось число предпочитаемых по Рамачандрану остатков, но то, что процент остатков всего на 0.07% выше порогового, заставляет подозревать, что произошла подгонка по пороговому параметру. Заодно резко ухудшилось качество структуры ДНК, при том, что в изначальной структуре все параметры качества были равны нулю. Поэтому, несмотря на некоторое улучшение показателей качества, нет причин предпочесть оптимизированную структуру изначальной.

Выводы

Структура 5a77 является хорошей структурой для своего разрешения. Реальная конформация большинства рассмотренных маргинальных остатков, скорее всего, совпадает с указанной в модели. Структура из базы PDB_redo не обладает особыми преимуществами в качестве перед изначальной структурой.

Список литературы

1. <http://www.doi.org/10.1074/jbc.M115.678342>
2. <http://www.doi.org/10.1021/bi048970c>
3. <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=5A77>
4. <https://www.rcsb.org/structure/5a77>
5. https://files.rcsb.org/pub/pdb/validation_reports/a7/5a77/5a77_full_validation.pdf
6. <http://molprobit.biochem.duke.edu>
7. <https://pdb-redo.eu/db/5a77>