

Отчёт о качестве расшифровки структуры белка
YABJ_BACSU
(PDB код 1QD9)
методом рентгеноструктурного анализа

Студентки четвёртого курса Факультета биоинженерии и биоинформатики
МГУ им. М. В. Ломоносова
Босхомджиевой Баины

Москва, 2015

Аннотация

В отчёте рассмотрены некоторые индикаторы качества модели структуры 1QD9, 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

Изучение белка *YabJ* началось с обнаружения открытой рамки считывания при расшифровке последовательности ДНК грамположительной бактерии *Bacillus subtilis*. Этот ген был впервые секвенирован в 1994 году, когда был проанализирован участок ДНК длиной 180 тысяч пар оснований. Авторами этой работы были проанализированы множество открытых рамок считывания, также была получена последовательность гена, кодирующего искомым белок.

В 1999 году было впервые экспериментально показано участие белка *Yabj* в регуляции биосинтеза пуринов. Одной из первых реакций этого метаболического пути является превращение инозин монофосфата в аденилосукцинат с образованием ГТФ (гуанозинтрифосфата). Эту реакцию катализирует фермент аденилосукцинат синтаза (*PurA*, EC=6.3.4.4). В дальнейшем из аденилосукцината синтезируется АМФ (аденозинмонофосфат), который является важнейшей сигнальной молекулой и, кроме того, является основой для производства пуриновых нуклеотидов. Активность *PurA* регулируется транскрипционным фактором *PurR*, кодирующий который ген располагается непосредственно перед *yabj* в геноме *Bacillus subtilis*. Было показано, что белок *YabJ* играет роль в регуляции активности транскрипционного фактора *PurR* в присутствии аденина и, как следствие, регуляции всего процесса биосинтеза пуринов.

Результаты и обсуждения

Общая информация о модели и индикаторы её качества в целом

Было установлено, что белок YABJ_BACSU представляет собой гомотример. Мономер имеет длину 124 аминокислотных остатка. Вторичная структура мономера представляет собой β -слой из шести тяжей, по большей части антипараллельных, и две α -спирали. В тримерной структуре β -листы складываются в единый замкнутый β -бочонок, окруженный со всех сторон α -спиралями. Также с белком связаны две молекулы этилартути и один ион ртути — с помощью аномально рассеивающих атомов решалась фазовая проблема. Кроме того, белок кристаллизовался в кислой среде в присутствии ацетата аммония

Модель была получена в 1999 году (Sangita Sinha, Pekka Rappu, S. C. Lange, Pekka Mäntsälä, Howard Zalkin and Janet L. Smith) Разрешение структуры 1.7 Å.

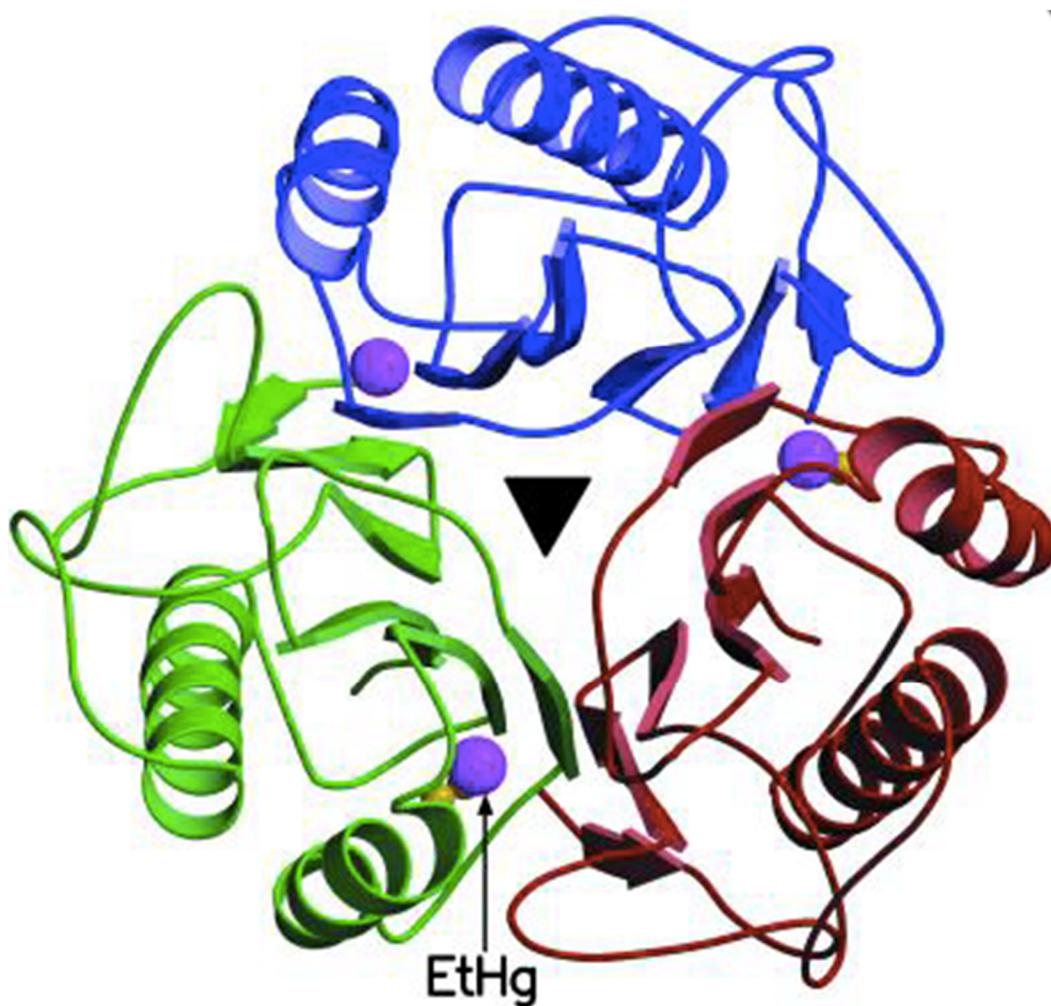


Рис 1. Структура 1QD9. Треугольником отмечена ось симметрии третьего порядка.

Фазовая проблема была решена методом многоволнового аномального рассеяния. Для этого кристаллы белка обрабатывались раствором этилртути, который ковалентно связывается с цистеинами белка. Так как атом ртути обладает свойством аномального рассеяния при разных длинах волн рентгеновского излучения, то это свойство использовалось для определения фаз. Образец помещали в рентгеновские лучи с длиной волны 1.0095, 1.0087, 1.0087 и 1.037 Å и определяли интенсивность рассеяния. Для определения координат атомов ртути и расчета фаз по измеренным модулям структурных факторов использовался Паттерсоновский метод с использованием разностных пар. При таком подходе координаты аномально рассеивающих атомов определяются по разности модулей структурных факторов с противоположными индексами

Эксперимент проводился при четырех разных длинах волн рентгеновского излучения. При этом три эксперимента проводились на одном приборе, который позволял получать данные с разрешением до 2.0 – 2.2 Å, а четвертый — на другом, позволившем достигнуть разрешения

1.7 Å. Все эксперименты проводились с одним и тем же кристаллом белка. Окончательная модель строилась по сумме данных с двух экспериментов, один из которых был с высоким разрешением

Для создания модели использовались объединенные данные двух экспериментов при разной длине волны рентгеновского излучения. В первом эксперименте было получено 19335 уникальных рефлексов, во втором – 19335. После объединения результатов осталось 35377 уникальных рефлексов, по словам авторов статьи. При этом в файле структурных факторов, полученных из банка PDB присутствует всего 35317 значений. Кроме того, там указана лишь одна длина волны возбуждающего излучения, и она не соответствует ни первой, ни второй из соответствующих экспериментов. Опыт с указанной длиной волны (1.0095 Å) также был проведен, однако, согласно статье, полученные данные использовались только для решения фазовой проблемы, а не для создания модели.

Из этих 35317 измеренных структурных факторов 34314 (97.2%) имеют силу сигнала, превышающую стандартное отклонение более чем в три раза. Среднее отношение измеренной интенсивности к стандартному отклонению составило 22.9. Это большое значение, говорящее о том, что уровень шума в измерениях мал и что значения интенсивностей измерены достаточно точно.

Максимальное и минимальное разрешение для использованных рефлексов составило 27.47 и 1.70 Å, соответственно. Такое разрешение для белковой структуры считается хорошим. Согласно EDS, количество использованных рефлексов составляет 34805. Полнота этого набора (процент использованных рефлексов от всех возможных с разрешением больше 1.7 Å) составила 96.6%. Это хорошее значение, говорящее о том, что фактическое разрешение действительно составляет около 1.7 Å, а не ухудшилось из-за неполноты набора данных.

Структура белка обладает некристаллографической осью симметрии третьего порядка : белок состоит из трех одинаковых цепей (мономеров), объединенных в тример, который, по-видимому, является биологической единицей. Элементарная ячейка кристалла примитивная. Длины ее сторон составляют 55.35, 55.35 и 204.96 Å, а углы между ними — 90, 90 и 120 градусов, соответственно. Ячейка принадлежит к группе симметрии C_3 , то есть симметрия кристалла гексагональная (образуется призма с шестиугольником в основании).

Из всех использовавшихся при построении модели рефлексов (34805 рефлексов) 5.1% (1759 рефлексов) были выделены как тестовый набор для подсчета R_{free} , а остальные — в рабочий набор. Получившаяся начальная модель имела $R_{work} = 35.5\%$ и $R_{free} = 35.1\%$. что составляет большие значения и говорит о плохом соответствии модели результатам эксперимента. Далее модель оптимизировалась и уточнялась с помощью программы CNS, причем для первых итераций уточнения использовалось ограничение на ротационную

симметрию третьего порядка. После уточнения R_{work} равнялось 16.6%, а R_{free} — 19.7%.

Эти значения говорят об успешном уточнении модели. Помимо простого R-фактора ($R_{\text{X-ray}}$), авторы привели значения RMSD для длин связей и значений углов. При этом значения связей и углов из модели сравнивались с соответствующими известными значениями. $\text{RMSD}_{\text{bonds}}$ составило 0.011 Å, $\text{RMSD}_{\text{angles}}$ — 1.60°, а $\text{RMSD}_{\text{dihedrals}}$ — 24.6°

Карты Рамачандрана для всех остатков структуры, построенная при помощи сервисов MolProbability и EDS представлена на рис. 2 и 3.

По результатам анализа с помощью обоих сервисов различалось количество найденных маргинальных остатков. Согласно EDS, таких остатков 16 (рис. 2). Видно, что большая часть из них лежит непосредственно около разрешенной области. По результатам MolProbability (рис. 3), у которого разрешенная область больше, лишь один из этих остатков действительно является маргинальным — это Asn43 цепи В. Однако, как и стоило ожидать, в предпочитаемую область попало всего 95.9% остатков (ожидаемое значение — 98%). Те остатки, которые были помечены EDS как маргинальные, в результате MolProbability попали в допустимую область, но не в предпочитаемую. Всего в допустимой области MolProbability расположено 99.7% остатков, что является нормальным значением.

По результатам анализа MolProbability в структуре есть только один аминокислотный остаток, попадающий в запрещенную область. Это Asn43 цепи В, углы ϕ и ψ которого равны 44° и 115.7°, соответственно. Кроме того, у этого остатка и предшествующего ему (Val42) угол τ (C-CA-N) больше чем на 4σ отклоняется от среднего значения. Все это говорит о том, что на этом участке произошло какое-то нарушение конформации основной цепи. При этом на таких же цепях А и С такого не наблюдается.

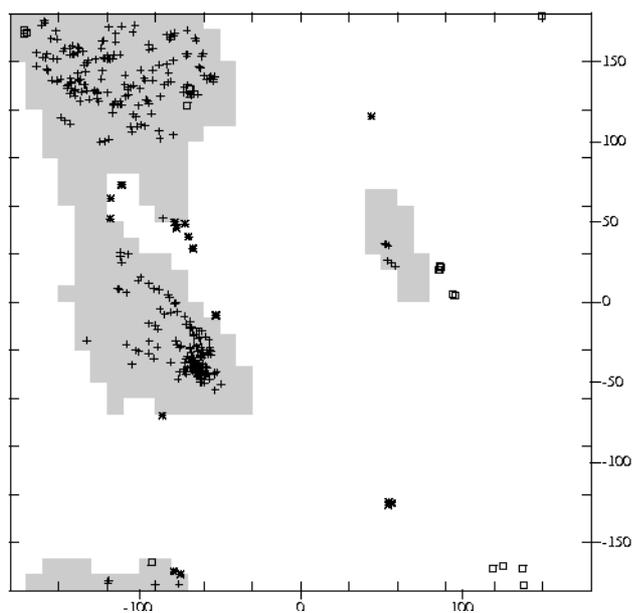


Рис 2. Карта Рамачандрана для структуры 1QD9 по результатам анализа EDS

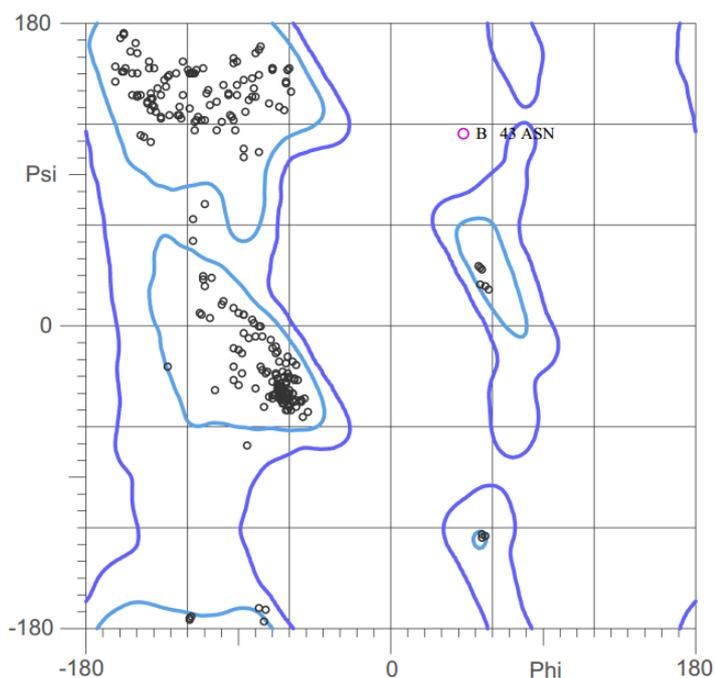


Рис 3. Карта Рамачандрана для структуры 1QD9 по результатам анализа MolProbity

По результатам анализа конформаций основной и боковых цепей аминокислотных остатков, их пространственного R-фактора и окружения был выявлен ряд маргинальных остатков. В итоговую сводную таблицу попали 15 из них. Предпочтение отдавалось остаткам, маргинальным сразу по нескольким критериям. Для каждого остатка в таблице обозначены критерии выбора и соответствующие значения оценочных функций

Остаток	Карта Рамачандрана и угол τ	Конформация боковых цепей	RSR фактор	Неблагоприятное окружение
Asn43 B	запрещенная область	нетипичные углы χ	Z-score ≈ 6	отклонение C β > 2Å
Asn43 AC		нетипичные углы χ	Z-score ≈ 6	нет водородных связей
Val42 A		нетипичные углы χ		
Val42 B	отклонение угла $\tau > 4\sigma$			
Phe84 A		нетипичные углы χ		
Lys125 A		нетипичные углы χ		
Glu48 BC		нетипичные углы χ		
Val116 ABC		Z-score двугр. углов < 2		
Thr7 B		Z-score двугр. углов < 2		

Остаток	Карта Рамачандрана и угол τ	Конформация боковых цепей	RSR фактор	Неблагоприятно е окружение
Ile14 B			Z-score = 3.9	
Thr97 A			Z-score = 2.3	
Ser67 C				нет водородных связей
Arg102 ABC				нет водородных связей
Gln52 A				инверсия
His9 C				инверсия

Анализ маргинальных остатков

Gln52 цепи A

В структуре было найдено всего шесть остатков с инверсией боковой цепи. Такая инверсия может происходить у остатков гистидина, глутамина и аспарагина. Детектируется она с помощью поиска водородных связей этих остатков. Если при повороте боковой группы остатка на 180° образование водородных связей энергетически более выгодно, значит, вероятно, электронная плотность была интерпретирована неправильно и имеет место инверсия. Так, например, если повернуть остаток глутамина 52 цепи A, то может образоваться водородная связь с остатком аспарагина 56 той же цепи, который находится в той же α -спирали (рис. 5). При этом по карте электронной плотности сложно различить кислород и азот, так что такая инверсия при определении конформации остатка вполне могла произойти. Других водородных связей, в том числе с водой, этот остаток глутамина ни в одной конформации не образует.

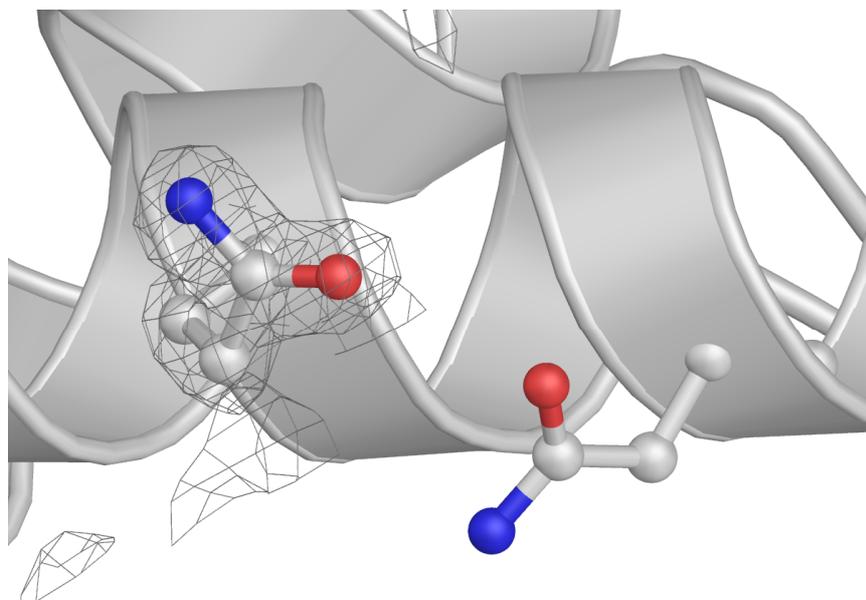


Рис 4. Остаток Gln52 цепи A с инверсией боковой цепи. Он представлен в шаростержневой модели, как и остаток аспарагина 56 той же цепи, с которым он может образовывать водородную связь при повороте боковой группы. На рисунке показан срез электронной плотности вокруг Gln52 (уровень подрезки — 1.7σ) Конформация в структуре 1QD9 (водородной связи не образуется).

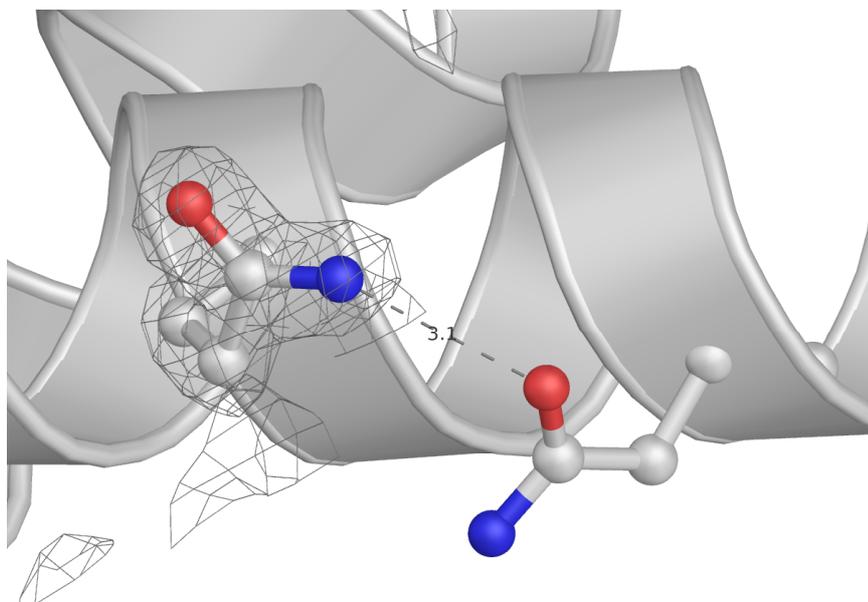


Рис 5. Конформация после поворота бокового радикала на 180° с помощью MolProbility, при которой образуется водородная связь.

Arg102

Аргинины 102 всех трех цепей белка были отмечены как маргинальные из-за отсутствия водородных связей (по протоколу WHAT CHECK водородные связи не образует атом NH1; связи с остатками воды не учитываются).

Анализ этого остатка вручную показал, что на самом деле, по-видимому, этот аргинин образует водородные связи. На соответствующем расстоянии от атома NH1 находится акцептор водородной связи OD1 остатка Asn88 той же цепи, находящегося в соседней α -спирали. Кроме того, оба NH-атома аргинина образуют водородные связи с водой. Водородные взаимодействия представлены на рисунке 6.

Стоит отметить, что Asn88 также входит в состав предполагаемого активного центра белка. Таким образом, была найдена водородная связь между двумя остатками этой группы, находящимися достаточно далеко друг от друга в последовательности и располагающимися в различных элементах вторичной структуры. Может быть, эта связь стабилизирует центр связывания субстрата.

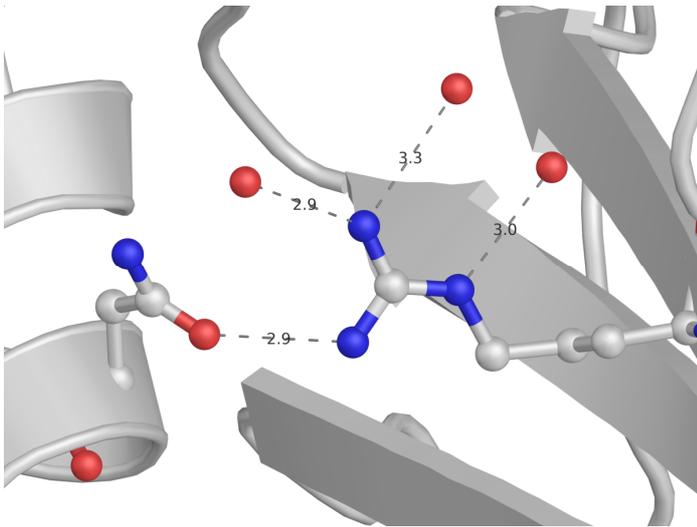


Рис. 6 Водородные связи остатка Arg102 цепи В

Asn43 цепи В

Остатки аспарагина 43 всех трех цепей были отмечены как маргинальные из-за нестандартной конформации боковых цепей и плохого пространственного R-фактора. Однако один из принадлежащий цепи В, является к тому же единственным остатком в запрещенной области карты Рамачандрана и имеет сильное отклонение положения C β -атома. Поэтому он и стал первым претендентом на подробный анализ. Остаток Asn43 цепи В, так же, как и аналогичные в других цепях, располагается на поверхности белка. Кроме того, он находится в составе достаточно подвижной петли рядом с α -спиралью (рис 7.)

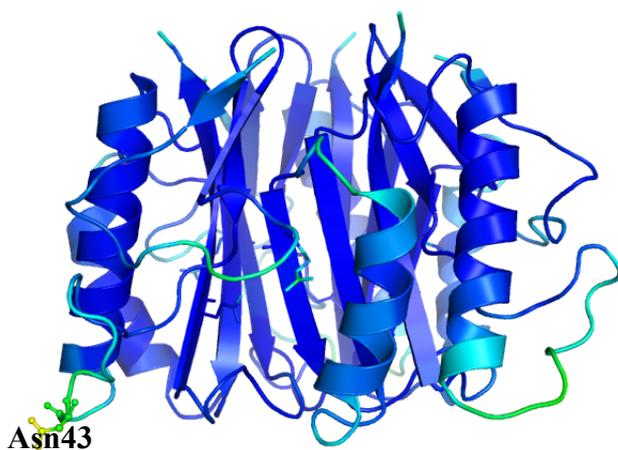


Рис 7. Расположение остатка Asn43 цепи В в структуре белка

Asn24

Остаток аспарагина 24 всех цепей также был отмечен в протоколе WHAT CHECK из-за отсутствия водородных связей у атома N. Однако здесь я хотела бы проанализировать его по другой причине: он упоминается в статье как один из образующих водородные связи между мономерами и, таким образом, отвечающих за четвертичную структуру белка. Протокол WHAT CHECK обращает внимание на то, что именно один атом N, входящий в основную цепь белка, не образует водородных связей с другими остатками. Это не является однозначным индикатором того, что с этим остатком что-то не в порядке или что присутствует ошибка расшифровки. Было обнаружено, что боковые радикалы Asn24 всех цепей взаимодействуют с Asn25 из соседней цепи (рис. 8). Таким образом, формируется кольцо из связанных тремя парами водородных связей дипептидов из трех мономеров. Оно образует “вход” в пространство между мономерами, заполненное водой. Авторы статьи совершенно верно описывают структурную роль этих остатков. Аспарагины также образуют водородные связи с водой, так как находятся на поверхности белка.

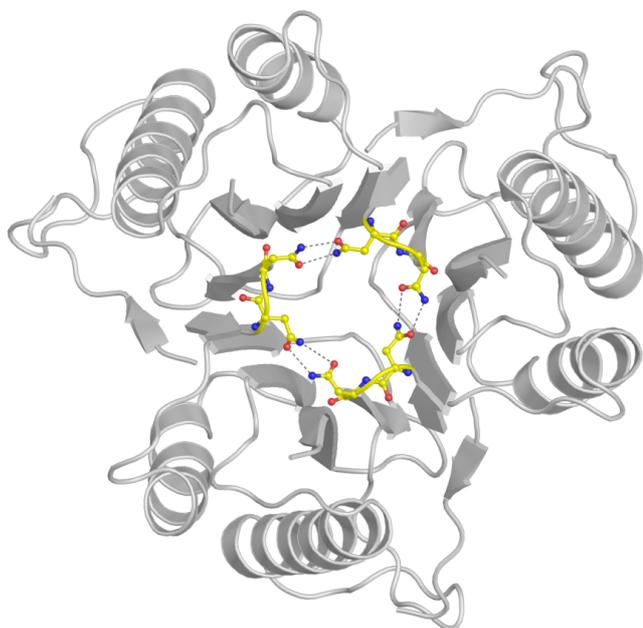


Рис. 8 “Кольцо”, образованное связанными водородными связями остатками Asn24 и Asn25 (показаны желтым цветом) мономеров белка.

PDB_redo

PDB redo оптимизирует структуру полностью автоматизированными методами и предоставляет сравнение полученного результата с исходной структурой. Для подсчета R_{free} в качестве тестового набора бралось 5% от всех используемых рефлексов, так же как для решения структуры 1QD9. Изменение индикаторов качества модели после оптимизации показано в таблице

Таблица 2: Индикаторы качества исходной модели и модели, оптимизированной PDB_redo

	Из заголовка pdb файла	Посчитанные по pdb файлу	Оптимизация PDB_redo
R_{work}	0.166	0.158	0.143
R_{free}	0.197	0.186	0.177
$R_{free}Z - score$		1.25	-0.93
Карта Рамачандрана (Z-score)		-1.23 -0.86	-0.83 -0.70
Ротамеры (Z-score)		0.46	0.51
Конформация основной цепи (Z- score)		0.44	0.74
$RMSD_{bonds} Z-score$		0.74	0.86
$RMSD_{angles} Z-score$			

	Из заголовка pdb файла	Посчитанные по pdb файлу	Оптимизация PDB_redo
Незадействованные акцепторы/доноры водородной связи		23	22

При наложении исходной и оптимизированной структур видно, что расположение основной цепи практически не изменилось. Все изменения затрагивают боковые группы остатков. В большинстве случаев отклонения незначительны, но для остатков, переставших быть маргинальными после оптимизации, можно увидеть серьезные изменения. В качестве примера был рассмотрен Asn43 цепи В, который был маргинальным по нескольким критериям. После оптимизации поменялась конформация боковой цепи этого остатка, однако нельзя сказать, что он стал вписан в электронную плотность лучше

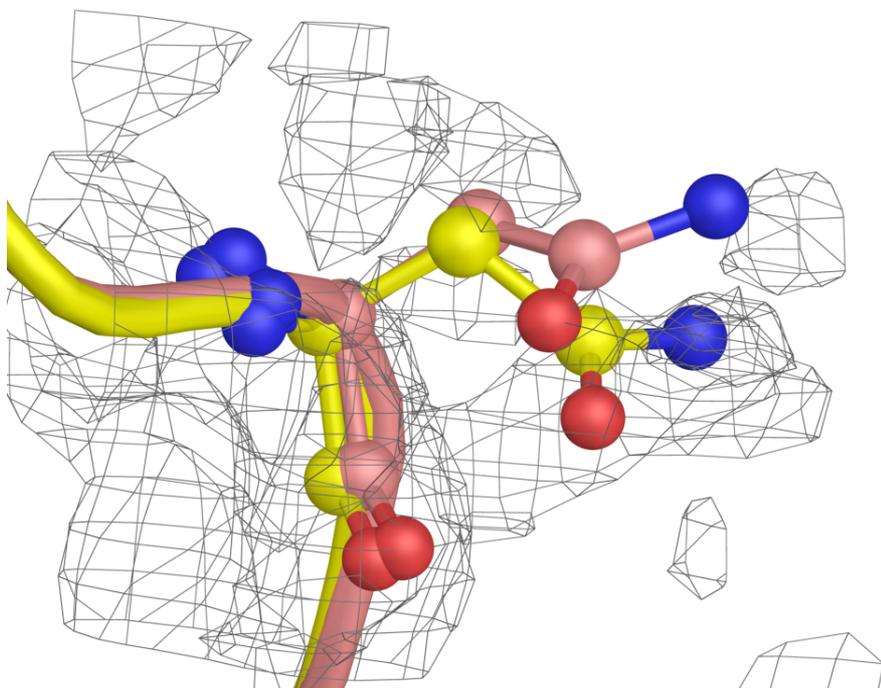


Рис 9. Различия в конформации боковой группы Asn43 цепи. Исходная структура показана желтым, а после оптимизации PDB_redo розовым. Уровень подрезки электронной плотности 0.5 σ .

Итак, оптимизацию модели с помощью автоматизированного алгоритма PDB_redo нельзя назвать намного лучше опубликованной структуры. Видно, что при подгонке моделей разными способами были по-разному расставлены приоритеты в минимизируемых критериях и ограничениях. Так, например, авторы структуры поставили сильные ограничения на длины связей и углы, из-за чего соответствующие RMSD снизились, но R-фактор стал больше, чем мог бы (и стал после PDB_redo). При этом качество вписывания модели в электронную плотность возросло: RSR значительно улучшилось для 72 остатков и ухудшилось только для трех. Таким образом, важно соблюдать баланс важности различных критериев при построении и подгонке модели

Заключение

1QD9 представляет собой модель высокого разрешения. Это подтверждают общие индикаторы качества, такие как R-фактор, RMSD связей и углов — все они имеют хорошие значения. В структуре есть маргинальные остатки по различным признакам, но серьезных нарушений, как, например, остатки в запрещенной области карты Рамачандрана, немного. По моему мнению, структура 1QD9 определена хоть и не идеально, но достаточно хорошо, чтобы ее можно было анализировать и делать правдоподобные предсказания свойств и предназначения тех или иных элементов структуры.

Литература

1. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis* / F. Kunst, N. Ogasawara, I. Moszer et al. // *Nature*. 1997. Nov. Vol. 390, No 6657. P. 249– 256.
2. Uppsala Electron-Density Server // <http://eds.bmc.uu.se/eds/>.
3. Crystal structure of *Bacillus subtilis* YabJ, a purine regulatory protein and member of the highly conserved YjgF family / S. Sinha, P. Rappu, S. C. Lange et al. // *Proc. Natl. Acad. Sci. U.S.A.* 1999. Nov. Vol. 96, No 23. P. 13074–13079.
4. RCSB Protein Data Bank // <http://pdb.org/>.
5. http://www.cmbi.ru.nl/pdb_redo/uw/1QD9/