

Московский государственный университет
имени М.В. Ломоносова

Факультет биоинженерии
и биоинформатики

Отчет по структурной биоинформатике студентки 4-го курса
Столяровой Анастасии Валерьевны

**Отчет о качестве расшифровки структуры белка RIDA_BACSU
(PDB-код 1QD9) методом рентгеноструктурного анализа**

Преподаватели:
А. В. Алексеевский
С. А. Спирин
В. Ю. Лунин
Е. А. Аксиянов

Москва
2014

Оглавление

Введение	3
Открытие белка	3
Результаты	6
Общая информация о структуре	6
Публикация структуры 1QD9	6
Решение фазовой проблемы	6
Состав комплекса	9
Измеренные рефлексы	9
Некристаллографическая симметрия и элементарная ячейка	10
Индикаторы качества модели в целом	12
R-factor и R-free	12
RMSD	12
Карта Рамачандрана	13
B-фактор	14
Поиск маргинальных остатков	15
Карта Рамачандрана	15
Конформация боковых цепей	15
RSR фактор	16
Неблагоприятное окружение	17
Сводная таблица маргинальных остатков	20
Подробный анализ нескольких маргинальных остатков	22
Asn43 цепи B	22
Phe84 цепи A	24
Arg102	25
Asn24	27
Gln52 цепи A	28
PDB_redo	30
Заключение	33
Литература	33

Введение

Открытие белка

Изучение белка *YabJ* (впоследствии *RidA*) началось с обнаружения открытой рамки считывания при расшифровке последовательности ДНК грам-положительной бактерии *Bacillus subtilis*. Первый раз этот ген был секвенирован в 1994 году,¹ когда был проанализирован участок ДНК длиной 180 тысяч пар оснований. Авторы этой работы проанализировали множество открытых рамок считывания и нашли возможные гомологичных закодированных в этой ДНК белков. Была получена последовательность гена, кодирующего искомый белок, однако авторами не было найдено его гомологов в других организмах и ему даже не было присвоено название. Результаты секвенирования были выложены в базу данных EMBL² (запись D26185).

Название *yabJ* было дано гену только в 1997 году, когда была опубликована большая статья секвенированию полного генома *Bacillus subtilis*.³ Авторы определили более четырех тысяч белок-кодирующих последовательностей (CDS) и проанализировали их (запись EMBL² AL009126). Тогда было определено точное местонахождение гена *yabJ* в геноме бактерии — он располагается по ходу цепи ДНК сразу за геном *purR*, кодирующим репрессор биосинтеза пуринов. Была предпринята попытка определить функцию белка по найденным гомологам из других организмов (*Escherichia coli*, *Salmonella enterica* и другие), однако функция этих гомологов тоже не была определена достоверно. Исходя из последовательности этих белков, было высказано предположение об их гидролазной либо деаминазной активности.

По мере анализа последовательностей ДНК других организмов было выделено семейство белков *YjgF/Yer057p/UK114*. Это семейство оказалось очень большим и крайне консервативным — белки из него были найдены в уже упомянутых *Bacillus subtilis*,³ *Escherichia coli*⁴ и *Salmonella enterica*,⁵ а также у дрожжей (*Saccharomyces cerevisiae*⁶), цианобактерий⁷ и млекопитающих (*Rattus rattus*⁸). Однако до 1999 года, в котором в банк PDB⁹ была выложена структура белка *YabJ* (1QD9) и написана соответствующая ста-

тя,¹⁰ не было опубликовано трехмерных структур представителей этого семейства. Годом раньше вышла статья, в которой сообщалось о кристаллизации гомолога из организма крысы,¹¹ однако в банке структур PDB нет соответствующей записи. Также не была найдена катализируемая белками этого семейства химическая реакция. Высказывались предположения, что столь консервативное семейство белков должно играть роль в общих для многих организмов метаболических путях. Также авторы структуры подробно описали пространственное устройство белка и даже предложили группе остатков как потенциально играющую важную функциональную роль в белке. Эта группа остатков будет подробно рассмотрена при анализе структуры.

Наконец, потом этот ген начали изучать детально. В 1999 году было впервые экспериментально показано участие белка *YabJ* в регуляции биосинтеза пуринов.¹² 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.^{12, 13} Эти исследования проводила та же группа ученых, которая определила структуру белка.

Наконец, в 2012 году, то есть уже после того, как на первом курсе я начала заниматься этим белком, был подробно исследован его гомолог *YjgF* из *Salmonella enterica*.¹⁴ Была подтверждена енамин/имин деаминазная активность и определены механизм реакции (рис. 1) и активный центр белка. По словам авторов статьи, в структуре исследованного ими белка отсутствует определенный аминокислотный остаток, непосредственно участвующий в катализе, и реакция происходит за счет размещения молекулы воды в определенное место активного центра. Остатки активного центра консервативны для всего семейства, в том числе белка *YabJ*, что дает основания полагать, он осуществляет катализ схожим образом. Более того, активный центр гомолога в выравнивании совпадает с предложенной авторами структуры 1QD9 функциональной группой остатков. После определения катализируемой ими реакции все белки семейства

были переименованы в *RidA* (reactive intermediate/imine deaminase A), им был присвоен ЕС 3.5.99.10.

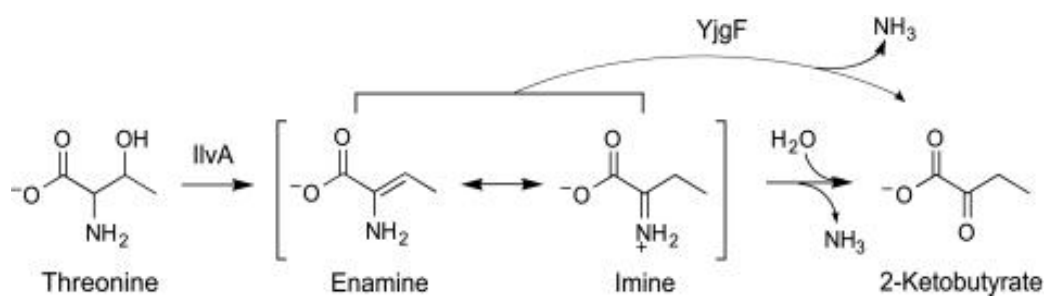


Рис. 1: Реакция, катализируемая семейством белков YjgF.

Результаты

Общая информация о структуре

Публикация структуры 1QD9

Статья, в которой была представлена трехмерная структура белка, на тот момент носящего название YABJ_BACSU (ID P37552), была опубликована в 1999 году в журнале PNAS¹⁵ группой ученых из Университета Пердью,¹⁶ США (С. Синха [S. Sinha], С. Ланг [S. Lange], Г. Залкин [H. Zalkin], Дж. Смит [J. Smith]) и Университета Турку, Финляндия (П. Раппу [P. Rappu], П. Мантсала [P. Mantsala]). По итогам статьи в банк PDB⁹ была добавлена структура 1QD9.

Для проведения успешного эксперимента авторы пользовались опытом кристаллизации растворимого в кислотной среде белка крысы,¹¹ принадлежащего тому же семейству.

Решение фазовой проблемы

Фазовая проблема была решена методом многоволнового аномального рассеяния. Для этого кристаллы белка обрабатывались раствором этилртути (формула приведена на рис. 2 А). Эта молекула ковалентно связывается с цистеинами белка. Атом ртути обладает свойством аномального рассеяния при разных длинах волн рентгеновского излучения, что и 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 (формула 1).

$$\begin{aligned}
|F^+| &\equiv |F_{hkl}| = |F_{\bar{h}\bar{k}\bar{l}}| \\
|F^-| &\equiv |F_{\bar{h}k\bar{l}}| = |F_{h\bar{k}l}| \\
\Delta F^2 &= (|F^+| - |F^-|)^2
\end{aligned}
\tag{1}$$

В цепи белка присутствует только один остаток цистеина (Cys104). С помощью мультиволнового аномального рассеяния были определены координаты трех атомов ртути. По восстановленной функции электронной плотности было подтверждено, что белок обладает некристаллографической симметрией третьего порядка (поворот на 120°), что использовалось при построении модели.

При этом в окончательной модели лишь с цистеинами двух из трех одинаковых белковых цепей (B, C) в тримере связаны молекулы этилртути. Третий определенный атом, согласно модели, представляет собой отдельный ион ртути, взаимодействующий с Cys104 и Met81 цепи A (рис. 2 А). Авторы не приводят этому объяснения. Возможно, по каким-то причинам электронная плотность соответствующего этила не была восстановлена, поэтому авторы статьи решили не включать в окончательную структуру целый лиганд. Однако на карте электронной плотности (рис. 2 Б) действительно не видно свидетельств присутствия других атомов рядом с ионами ртути даже на малых уровнях подрезки.

Указанные в pdb-файле со структурой коэффициенты заполнения не всегда равны единице для молекул этилртути. Для таковых, связанных с цистеинами цепей B и C, представлено по две альтернативные конформации с коэффициентами заполнения 0.73 и 0.15 для цепи B и 0.68 и 0.18 для цепи C, соответственно. На рис. 2 В показан этот участок структуры для цепи B. При этом альтернативной конформации молекулы этилртути соответствует альтернативная конформация (ротамер) остатка Cys104. Коэффициенты заполнения атома SG цистеина во всех случаях равны коэффициентам для связанной конформации этилртути либо больше их. Последний случай соответствует ситуации, когда соединение ртути не присоединилось к белку вовсе. Такое происходило в 12% случаев для цепи B и 14% для цепи C. При этом обе альтернативные конформации цистеинов вполне соответствуют известным ротамерам: углы χ для них равны около 175° и 315° для обеих цепей.

Для цистеина цепи A нет альтернативных конформаций. Однако ион ртути, связанный с ним, может также находиться в двух позициях (рис. 2 Г). В 60% ячеек он, по

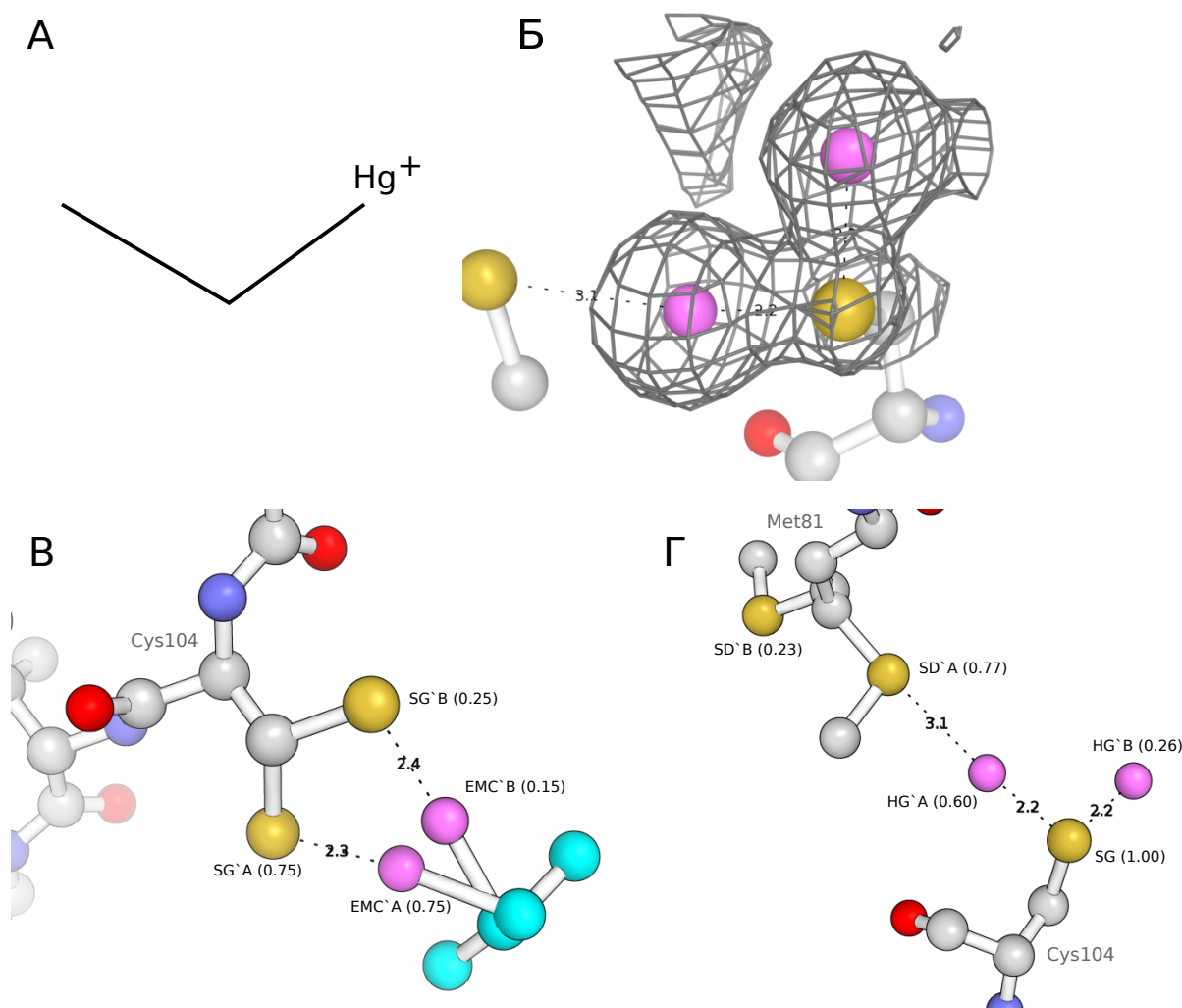


Рис. 2: А: Структурная формула этилртути (ЕМС). Б: Карта электронной плотности около ионов ртути (уровень подрезки 1σ). В: Участок цепи В белка, связанный с молекулой этилртути. Г: Участок цепи А белка, связанный с ионом ртути. Участвующие во взаимодействиях атомы подписаны с коэффициентами заполнения. Атомы ртути покрашены розовым, углероды в этилртути – голубым, сера – желтым, углероды белка – серым.

мнению авторов статьи, взаимодействует также с Met81 той же цепи (который, аналогично, имеет две альтернативные конформации). Действительно, метионин поворачивается к атому ртути (однако минимум в $77-60 = 17\%$ ячеек он имеет такую же конформацию и в его отсутствие). Однако странным остается длина связи между серой метионина и ионом ртути — она составляет 3.1 \AA , что много для этой связи. Еще в 26% ячеек ион ртути связан только с цистеином; в остальных он не был зафиксирован вовсе.

Как упоминалось выше, эксперимент проводился при четырех разных длинах волн рентгеновского излучения. При этом три эксперимента проводились на одном приборе, который позволял получать данные с разрешением до $2.0 - 2.2\text{ \AA}$, а четвертый — на

другом, позволившим 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.

Состав комплекса

По результатам эксперимента 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 (рис. 3). В тримерной структуре β -листы складываются в единый замкнутый β -бочонок, окруженный со всех сторон α -спиралями.

Помимо белка, в структуре присутствуют и низкомолекулярные вещества. Прежде всего это 474 молекулы растворителя (воды). Также с белком связаны две молекулы этилртути и один ион ртути — с помощью аномально рассеивающих атомов решалась фазовая проблема. Кроме того, белок кристаллизовался в кислой среде в присутствии ацетата аммония — растворимость белка из того же семейства в подкисленном растворе и возможность его кристаллизации была показана ранее.¹¹ Поэтому в составе комплекса есть шесть ионов ацетата, по два на каждый мономер.

Измеренные рефлексы

Как говорилось выше, 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 PDB⁹ 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



Рис. 3: Структура 1QD9. Красным покрашены α -спирали, желтым — β -тяжи, зеленым — петли. Треугольником отмечена ось симметрии третьего порядка.

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,¹⁷ 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.

Некристаллографическая симметрия и элементарная ячейка

Структура белка обладает некристаллографической осью симметрии порядка три: 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 6_5 , то есть симметрия кристалла гексагональная (образуется призма с шестиугольником в основании).

Индикаторы качества модели в целом

R-фактор и R-free

Авторы строили модель при помощи программы O.¹⁸ 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. В работе R-фактор не был составным, то есть в нем не учитывалось отклонение длин связей и значений углов (формула 2), однако, как оказалось, авторы накладывали сильные ограничения на длины связей и значения углов.

$$R_{X-ray} = \frac{\sum_{hkl} ||F_{hkl}^{obs} - F_{hkl}^{calc}||}{\sum_{hkl} |F_{hkl}^{obs}|} \quad (2)$$

Затем модель оптимизировалась и уточнялась с помощью программы CNS,¹⁹ 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, при этом R_{free} на 3.1% больше R_{work} , то есть модель вряд ли ”переоптимизирована” под рабочее множество рефлексов.

RMSD

RMSD связей и углов

Помимо простого R-фактора (R_{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. $RMSD_{bonds}$ составило 0.011 Å, $RMSD_{angles}$ — 1.60°, а $RMSD_{dihedrals}$ — 24.6°, что кажется мне вполне хорошими значениями. Также было измерено $RMSD_{improper}$, то есть отклонение от “виртуальных” двугранных углов, которыми описываются, например, плоские группы атомов (такие как карбоксильная, группы с двойными связями и другие) и хиральные группы. Оно составило 0.98°, что также является показателем хорошего качества структуры, тем

более что, по словам авторов, эти параметры не минимизировались в процессе уточнения модели.

Однако в результатах анализа WHAT CHECK отмечено, что такое стандартное отклонение для длин связей меньше обычного ($Z\text{-score} = 0.437$), что может свидетельствовать о применении слишком сильных ограничений на длины связей при построении и подгонке модели. Это могло в конце концов сказаться на качестве восстановленной структуры. Кроме того, аналогичная ситуация с двугранным углом ω — его стандартное отклонение в структуре составило около 2° , что намного меньше среднего по множеству структур 5.5° . Это также может свидетельствовать о слишком сильной подгонке углов.

RMSD между мономерами

Так как ограничения на одинаковую структуру мономеров были сняты на последних итерациях подгонки модели, RMSD между мономерами перестало быть равным нулю и составило 0.3\AA . Это небольшое значение, вполне допустимое для тримера. Разница в строении мономеров могла возникнуть из-за неидеальности кристаллизации и подвижности некоторых элементов структуры. Кроме того, в структуре есть атомы с большим B-фактором, то есть их местонахождение определено неточно (см. раздел). Согласно WHAT CHECK, RMSD между всеми атомами цепей составляет от 0.63 до 0.75\AA для разных пар цепей, а между только $C\alpha$ -атомами — от 0.27 до 0.32\AA . Видимо, в статье авторы привели именно последнее значение.

Карта Рамачандрана

Карты Рамачандрана для всех остатков структуры, построенная при помощи сервисов MolProbability²⁰ и EDS¹⁷ представлена на рис. 4. У этих сервисов различаются границы дозволенной области, кроме того, MolProbability учитывает разницу между "обычными" остатками и остатками пролина, глицина и других уникальных случаев с другим распределением на карте Рамачандрана. Соответственно, по результатам анализа с помощью обоих сервисов различалось количество найденных маргинальных остатков. Согласно EDS, таких остатков 16 (рис. 4 А). Видно, что большая часть из них лежит непосредственно около разрешенной области. По результатам MolProbability (рис. 4 Б), 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.

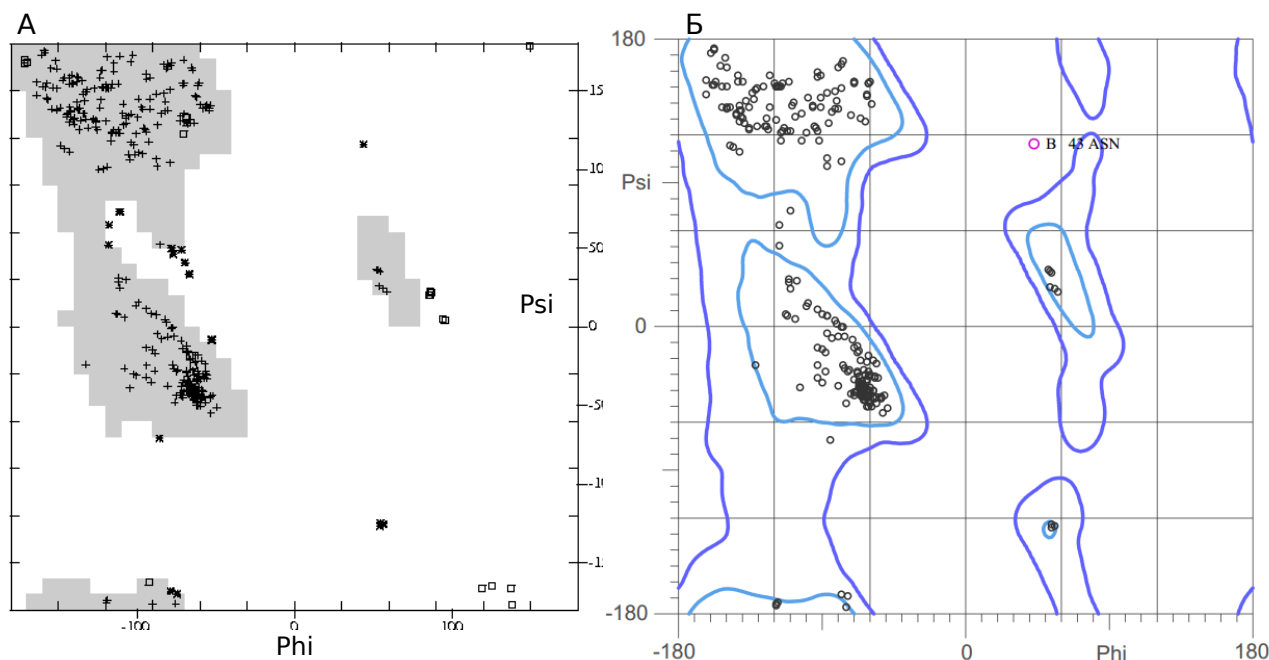


Рис. 4: Карты Рамачандрана для структуры 1QD9 по результатам анализа EDS¹⁷ (А) и MolProbability²⁰ (Б). MolProbability предоставляет также карты Рамачандрана для “нетипичных” остатков (Gly, Pro и т. д.), здесь показана карта для большинства остатков. Видно, что остатки подразделяются на кластеры, соответствующие разным элементам вторичной структуры.

Стоит отметить, что при этом авторы пишут в статье, что маргинальных остатков по значениям углов ϕ и ψ не обнаружено вовсе.

В-фактор

В структуре нет “неопределенных” подвижных петель со слишком размазанной электронной плотностью.

Из всех атомов структуры, принадлежащих белку 4% атомов (116) имеют В-фактор больше 40. Максимальный температурный фактор составляет около 69. При этом В-фактор остатков не коррелирует с вторичной структурой (на рис. 5 показан В-фактор для одной из цепей).

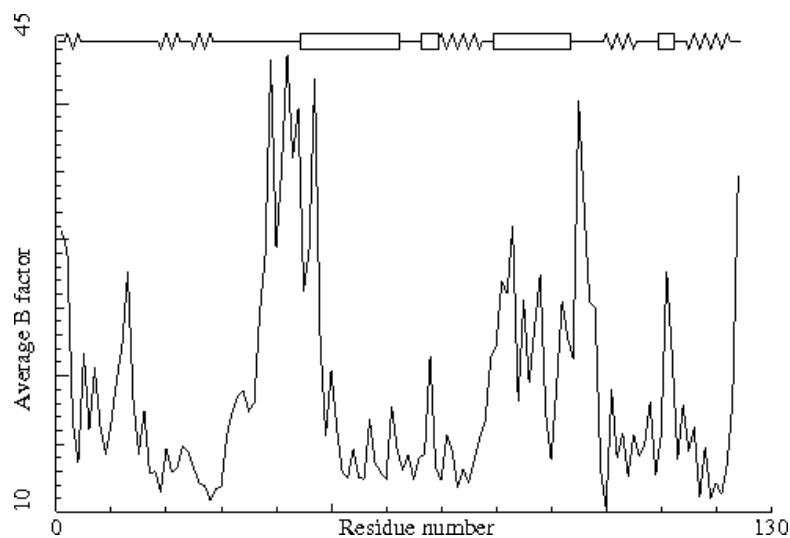


Рис. 5: B-фактор для остатков структуры (цепи A), соотнесенный со вторичной структурой белка. Источник – WHAT CHECK.

Поиск маргинальных остатков

Карта Рамачандрана

Как говорилось выше, по результатам анализа MolProbity²⁰ в структуре есть только один аминокислотный остаток, попадающий в запрещенную область. Это Asn43 цепи В, углы ϕ и ψ которого равны, соответственно, 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.

Конформация боковых цепей

Было найдено всего восемь аминокислотных остатков с конформацией боковых цепей, отличающейся от обычной. Прежде всего это три остатка Asn43 из всех цепей, один из которых уже был отмечен как маргинальный по конформации основной цепи. На рисунке 6 показано, что значения углов χ для этих остатков сильно отклоняются от предпочтительных. При этом аспарагины цепей В и С представлены в структуре в двух альтернативных конформациях с разными коэффициентами заполнения. У Asn43 цепи В при этом обе конформации неблагоприятны, у такого же остатка цепи С — только одна.

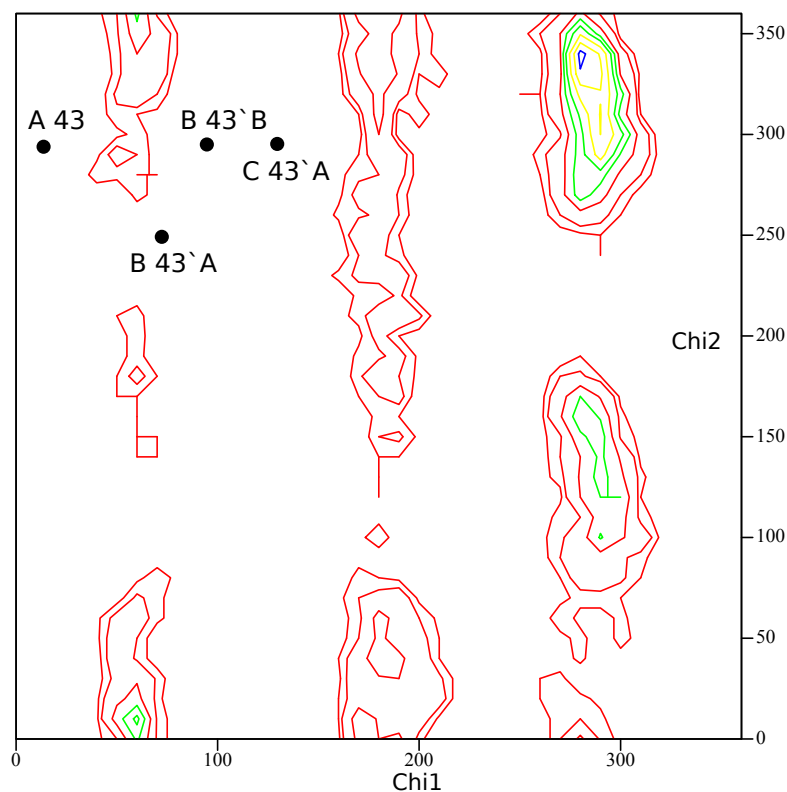


Рис. 6: График предпочтительных углов χ для остатков аспарагина (по данным библиотеки Уппсальского университета²¹) и "нетипичные" остатки Asn43 структуры. Все они выпадают из предпочтительных областей значений углов. Для остатков с двумя альтернативными конформациями (Asn43 цепей B и C) представлены только нетипичные (Asn43`A и Asn43`B цепи B и Asn43`A цепи C).

Кроме того, нетипичной конформацией боковых цепей обладают Glu48 цепей B и C, Val42, Phe84 и Lys125 цепи A.

Двугранные углы

Протокол WHAT CHECK предоставляет информацию о соответствии всех торсионных углов аминокислотного остатка, кроме ω (то есть ϕ , ψ и углов χ) их наиболее распространенным значениям, и высчитывает соответствующий Z-score. В структуре было найдено несколько остатков с Z-score меньше -2, что свидетельствует о сильном отклонении от обычных значений. Это, прежде всего, Val116 из всех цепей (что говорит уже о некоторой закономерности такого отклонения), а также Pro16 и Thr7 цепи B.

RSR фактор

Большие значения RSR фактора говорят о том, что модель плохо вписана в "экспериментальную" электронную плотность. Для удобного анализа используют Z-score, который показывает, насколько плохо остаток вписан в электронную плотность по сравнению со средним среди множества таких же остатков из многих pdb файлов. С

помощью сервиса EDS¹⁷ был получен список остатков с Z-score больше 2.0. Таких нашлось всего шесть. График со значениями Z-score для всех остатков представлен на рис. 7.

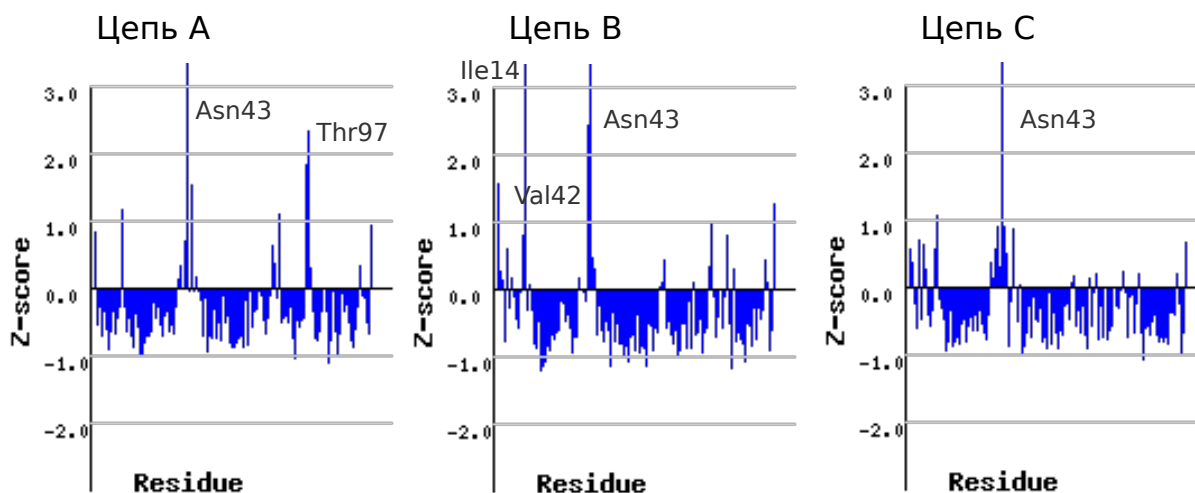


Рис. 7: График Z-score пространственного R-фактора для всех остатков трех полипептидных цепей структуры 1QD9. На графике подписаны остатки с Z-score > 2, говорящем о сильном несоответствии “экспериментальной” электронной плотности и электронной плотности модели.

Наихудший Z-score среди всех остатков структуры — у упомянутых выше аспарагинов всех трех цепей. Значения Z-score для них колеблются около 6, а RSR составляют от 0.36 до 0.48. Кроме того, есть уникальные для каждой цепи маргинальные остатки — это опять же упоминавшийся выше Val42 цепи B, у которого наблюдалось сильное отклонение угла основной цепи и двугранных углов в боковом радикале (Z-score = 2.4), Ile14 той же цепи (Z-score = 3.9) и Thr97 цепи A (Z-score = 2.3).

Неблагоприятное окружение

Инверсия боковой цепи

При анализе структуры сервисом MolProbity²⁰ и изучении протокола WHAT CHECK было найдено шесть аминокислотных остатков с предполагаемой инверсией боковой цепи. Все вышеупомянутые маргинальные остатки являлись таковыми уже после применения обратной инверсии. Если оставить их структуру такой же, какой она представлена в исходном pdb файле, то эти остатки все равно остаются маргинальными. К числу таких остатков принадлежат уже отмеченный как маргинальный Asn43 цепи C, причем обе его альтернативные конформации, Gln32 и Gln52 цепей A, Gln32 и Gln83 цепи B и His9 цепи C.

Отсутствие водородных связей

В структуре есть много полярных остатков, которые не образуют водородных связей внутри белка и при этом не находятся на внешней поверхности белка. Потенциальными акцепторами водородных связей являются остатки Glu117 всех трех цепей. Однако при осмотре этих участков структуры видно, что они образуют водородные связи с водой, формируя поверхность промежутков между мономерами белка. То же самое говорят и авторы в своей статье.

Потенциальных доноров водородных связей в структуре много. Из них в статье отмечаются Asn24 и Arg102 из всех трех цепей, которые, по словам авторов, связывают друг с другом субъединицы белка. Они были отобраны для подробного анализа. Ser67, боковой радикал которого также не образует водородных связей с другими аминокислотами, также был добавлен в сводную таблицу 1.

Отклонение $C\beta$

Был найден только один остаток с отклонением $C\beta$ больше 2 Å— это уже многократно упоминавшийся Asn43 цепи В, причем обе его альтернативные конформации. Других маргинальных остатков по этому параметру в структуре нет.

Планарность ароматических колец

Атомы ароматических колец боковых радикалов аминокислотных остатков (His, Trp, Phe, Tyr,?) должны лежать в одной плоскости — этого требует их sp^2 -гибридизация и обобщенность электронного облака. Кроме того, непосредственно соединенные с ароматической структурой атомы тоже должны лежать в этой плоскости. Однако для His9 цепи А изучаемой структуры это правило нарушается: его $C\beta$ атом отклоняется от плоскости ароматических колец. Однако Z-score = -0.87 (больше -1), то есть такое отклонение нельзя назвать стопроцентно ошибочным (рис. 8). Возможно, из-за излишних ограничений на длины связей или углы либо из-за неточности восстановления электронной плотности и подгонки модели координаты атома $C\beta$ были определены неточно. В сводную таблицу маргинальных остатков этот гистидин не вошел.

Пересечение Ван-дер-Ваальсовых радиусов атомов

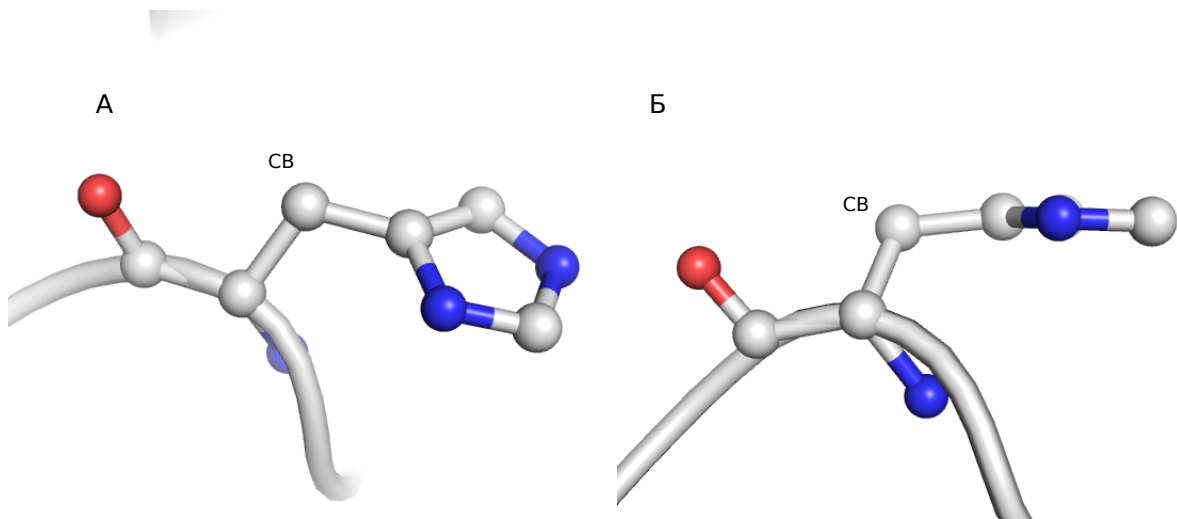


Рис. 8: Остаток His9 цепи А. Рисунки А и Б показывают остаток с двух ракурсов. Как видно, атом СВ (подписан) отклоняется от плоскости кольца бокового радикала.

Атомы в структуре не могут располагаться слишком близко друг к другу. На близких расстояниях на них действуют силы ван-дер-ваальсового отталкивания. Однако в структуре было найдено порядка 20 пар атомов белка, радиусы которых пересекаются больше чем на 0.4 \AA (критерий, применяемый как в WHAT CHECK, так и в MolProbability). Кроме того, есть много перекрытий с молекулами воды или низкомолекулярных веществ. Однако если учитывать лишь атомы с температурным фактором меньше 40, число перекрытий сокращается больше чем вдвое (clashscore = 1.6 ‰). Кроме того, наиболее критические значения пересечений приходится на молекулы воды. Также многие из пересекающихся атомов принадлежат аминокислотным остаткам, уже отобраным как маргинальные по другим признакам (Thr97, Val42 и другие).

Молекулы воды

В протоколе WHAT CHECK была обозначена только одна молекула воды, не образующая водородных связей с белком или другими молекулами воды (НОН657 цепи А). Однако при ближайшем рассмотрении стало ясно, что она вполне может образовывать водородные связи с кислородами С-концевого лизина соответствующей цепи, находящимися на расстоянии 3.3 \AA от нее. Непонятно, почему именно эта молекула была выявлена при анализе, тем более что точно в таком же положении находятся такие же остатки из других цепей.

Мне кажется странным, что абсолютно все молекулы воды, представленные в pdb файле, имеют коэффициент заполнения, равный единице. При этом, однако, около

41% молекул воды имеют температурный фактор больше 40, что свидетельствует о их подвижности и неточной локализации в пространстве.

Сводная таблица маргинальных остатков

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. Маргинальных молекул воды выявлено не было, поэтому в таблице их нет. Для каждого остатка в таблице обозначены критерии выбора и соответствующие значения оценочных функций.

Таблица 1: Сводная таблица маргинальных остатков. Серым фоном помечены остатки, упоминаемые в статье о структуре 1QD9¹⁰

Остаток	Критерий			
	Карта Рама-чандрана и угол τ	Конформация боковых цепей	RSR фактор	Неблагоприятное окружение
Asn43 B	запрещенная область	нетипичные углы χ	Z-score ≈ 6	отклонение $C\beta > 2 \text{ \AA}$
Asn43 AC		нетипичные углы χ	Z-score ≈ 6	нет водородных связей
Val42 A		нетипичные углы χ		
Val42 B	отклонение угла $\tau > 4\sigma$		Z-score = 2.4	
Phe84 A		нетипичные углы χ		
Lys125 A		нетипичные углы χ		
Glu48 BC		нетипичные углы χ		
Val116 ABC		Z-score дву-гр. углов < 2		
Thr7 B		Z-score дву-гр. углов < 2		
Ile14 B			Z-score = 3.9	
Thr97 A			Z-score = 2.3	
Ser67 C				нет водородных связей
Arg102 ABC				нет водородных связей (карман)
Gln52 A				инверсия
His9 C				инверсия

Подробный анализ нескольких маргинальных остатков

Asn43 цепи В

Остатки аспарагина 43 всех трех цепей были несколько раз отмечены как маргинальные — прежде всего из-за нестандартной конформации боковых цепей и плохого пространственного R-фактора. Однако один из них, принадлежащий цепи В, является к тому же единственным остатком в запрещенной области карты Рамачандрана и имеет сильное отклонение положения $C\beta$ -атома. Поэтому он и стал первым претендентом на подробный анализ, хоть и его роль в белке никак не освещена в статье.

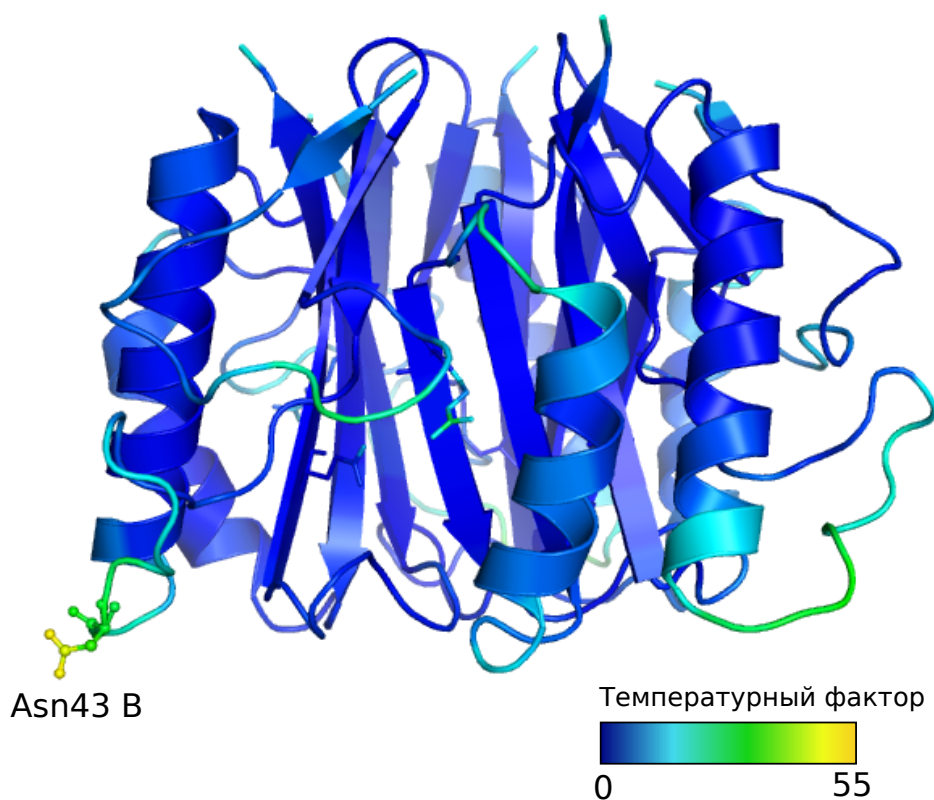


Рис. 9: Расположение остатка Asn43 цепи В в структуре белка. Указанный остаток показан в шаростержневой модели и подписан, вся структура покрашена по В-фактору. В-фактор атомов бокового радикала Asn43 достигает 55.

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. 9). Все атома бокового радикала этого остатка (CB , CG , $OD1$ и $ND2$) имеют температурный фактор больше 40, что свидетельствует о крайне

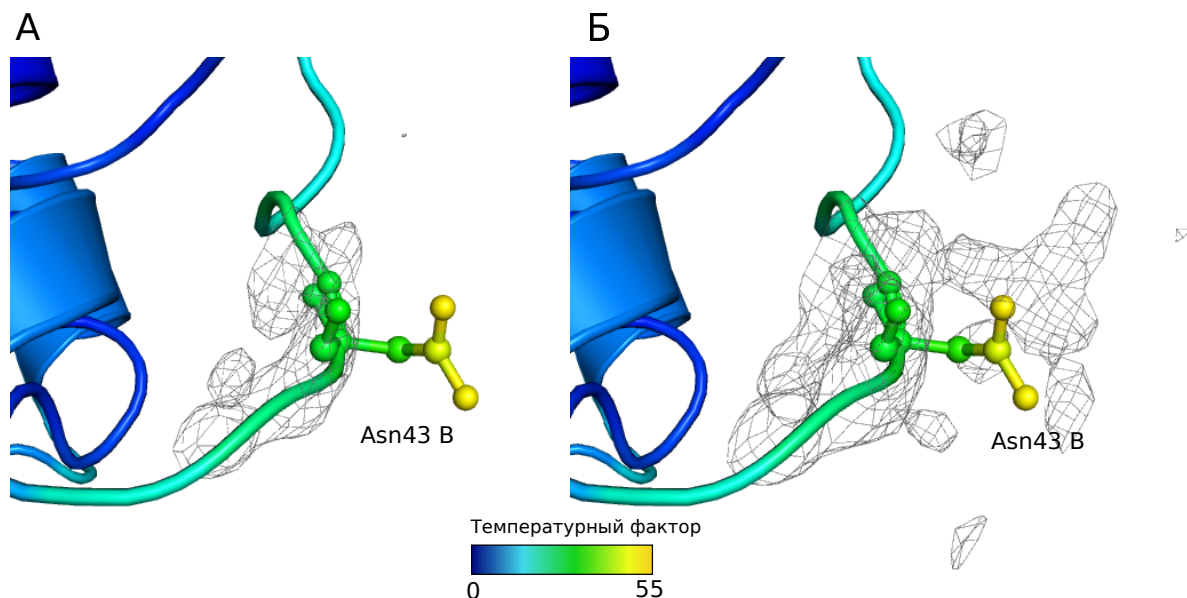


Рис. 10: Карта электронной плотности рядом с остатком Asn43 цепи В. Атомы белка покрашены по В-фактору. А: Уровень подрезки электронной плотности 1σ — различима только основная цепь. Б: Уровень подрезки электронной плотности 0.5σ — видно много шума, однако боковой радикал все еще неразличим. В местах скопления электронной плотности рядом с указанным остатком нет молекул воды, поэтому они, вероятно, представляют собой шум.

неточном определении их местоположения. Этим можно объяснить несоответствие конформации бокового радикала известным ротамерам. Атомы основной цепи также имеют В-фактор, близкий к 40 (от 35 до 38), такая подвижность основной цепи может быть объяснением того, что этот остаток располагается в запрещенной области карты Рамачандрана. Чтобы понять причину большого Z-score пространственного R-фактора, то есть сильного отклонения рассчитанной по модели электронной плотности от “экспериментальной”, была построена полученная с сервера EDS¹⁷ карта электронной плотности рядом с этим остатком (рис. 10). Как оказалось, при уровне подрезки в 1σ на карте электронной плотности различима только основная цепь. Даже при уровне подрезки электронной плотности в 0.5σ , при котором видно много шума, невозможно увидеть ход боковой цепи остатка по электронной плотности.

Вывод: Остаток Asn43 цепи В был очень неточно восстановлен по электронной плотности. Вероятно, это связано с большой подвижностью этого участка полипептидной цепи, располагающегося на поверхности белка, вследствие чего электронная плотность была слишком расплывчата и сигнал был неотличим от шума. При этом появляются сомнения в правильности определения авторами статьи В-фактора данных атомов. Кроме того, непонятно, как вообще определять водородные связи этого

остатка с чем-либо. Это объясняет, почему таковых и не было найдено: этот остаток в представленной модели не связан с водой либо другими аминокислотными остатками ни в своей, ни в соседних ячейках. Аналогичная ситуация обстоит с остатками Asn43 двух других цепей.

Phe84 цепи A

У фенилаланина 84 цепи A конформация боковой цепи не соответствует типичным ротамерам для этого аминокислотного остатка. Он был выбран для детального анализа, так как, вероятно, играет важную роль в обеспечении каталитической активности белка *Yabj*. Когда авторы статьи получили его трехмерную структуру, ни у одного белка из семейства *YjgF/Yer057p/UK114* не были определены экспериментальными методами функциональные группы. Однако среди схожих с *Yabj* по последовательности белков было найдено девять консервативных аминокислотных остатков.¹⁰ Они не составляют единую группу в последовательности, однако в структуре 1QD9 оказались рядом друг с другом между двумя мономерами белка. Кроме того, в структурном гомологе белка *Yabj* хоризмат мутазе *Bacillus subtilis* в этом же сайте находится центр связывания субстрата. Это не является прямым доказательством того, что данные остатки образуют активный центр фермента, однако свидетельствуют о их важной роли в белке. Более того, как выяснилось позже, эти остатки соответствуют активному центру подробно исследованного гомолога белка *YjbF*, что является более прямым указанием на каталитическую функцию этих остатков.

В эту группу “активного центра” белка входит и маргинальный остаток Phe84.

На рисунке 11 показаны значения двугранных углов χ изучаемого остатка на графике распределения этих углов среди большого числа остатков (около 3400) из других pdb файлов. Как видно, угол χ_1 (186°) соответствует одному из популярных значений, а χ_2 (129°) сильно отклоняется от них.

Связи, поворот вокруг которых обозначают двугранные углы χ_1 и χ_2 , показаны на рис. 12 А. Как видно, отклонения от типичных значений χ_2 означают поворот вокруг связи СВ-СG, то есть поворот ароматического кольца относительно остальной части остатка. Как видно по карте электронной плотности на различных уровнях подрезки (рис. 12 Б, В), атомы бокового радикала нельзя из нее однозначно восстановить, хорошо различимы только атомы, принадлежащие основной цепи.

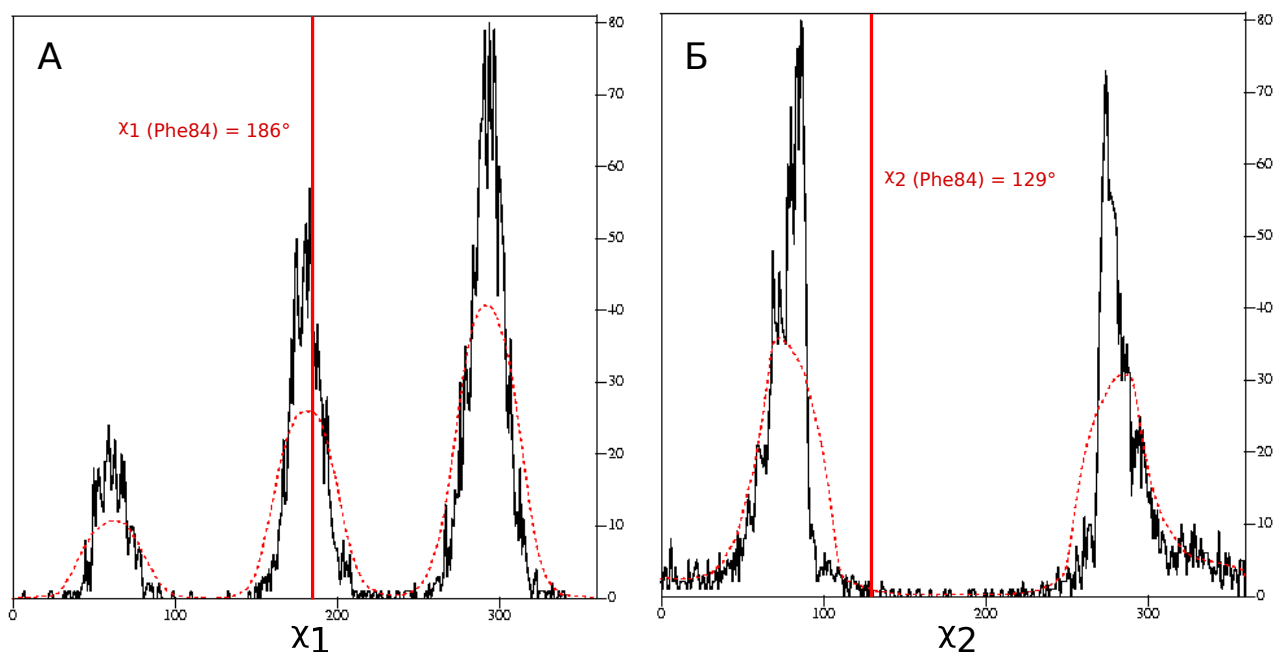


Рис. 11: Двугранные углы χ остатка Phe84 цепи A в сравнении с распределением этих углов среди множества подобных остатков. Значение двугранного угла в структуре отмечено красной линией. А: Угол χ_1 . Б: Угол χ_2 . Источник графиков распределения двугранных углов — сервис Уппсальского университета.²¹

Объяснить такую конформацию боковой цепи фенилаланина трудно. Рядом с ним нет других плоских ароматических колец, с которыми он мог бы вступать в стекинг-взаимодействия. Более того, он окружен полярными остатками, такими как аргинин и аспарагин. Возможно, он входит в состав предполагаемой функциональной группы для стабилизации гидрофобной части субстрата, если она есть, иначе его расположение трудно объяснить. Кроме того, нарушение его конформации может быть связано с тем, что рядом (на расстоянии меньше 4 Å) с ним находятся ионы ртути, появившиеся в структуре из-за использования метода многоволнового аномального рассеяния для решения фазовой проблемы. Возможно, присутствие ртути изменило конформацию не только взаимодействующих с ней цистеина и метионина, но и других лежащих рядом остатков. Белок вряд ли взаимодействует с ионами ртути в клетке, и нельзя сказать, является ли такая конформация остатков нативной.

Arg102

Аргинин 102 всех трех цепей белка также входит в группу консервативных остатков некоторых представителей семейства *YjgF/Yer057p/UK114*, которые, возможно, образуют важную функциональную группу белка. Все три аргинина были отмечены как

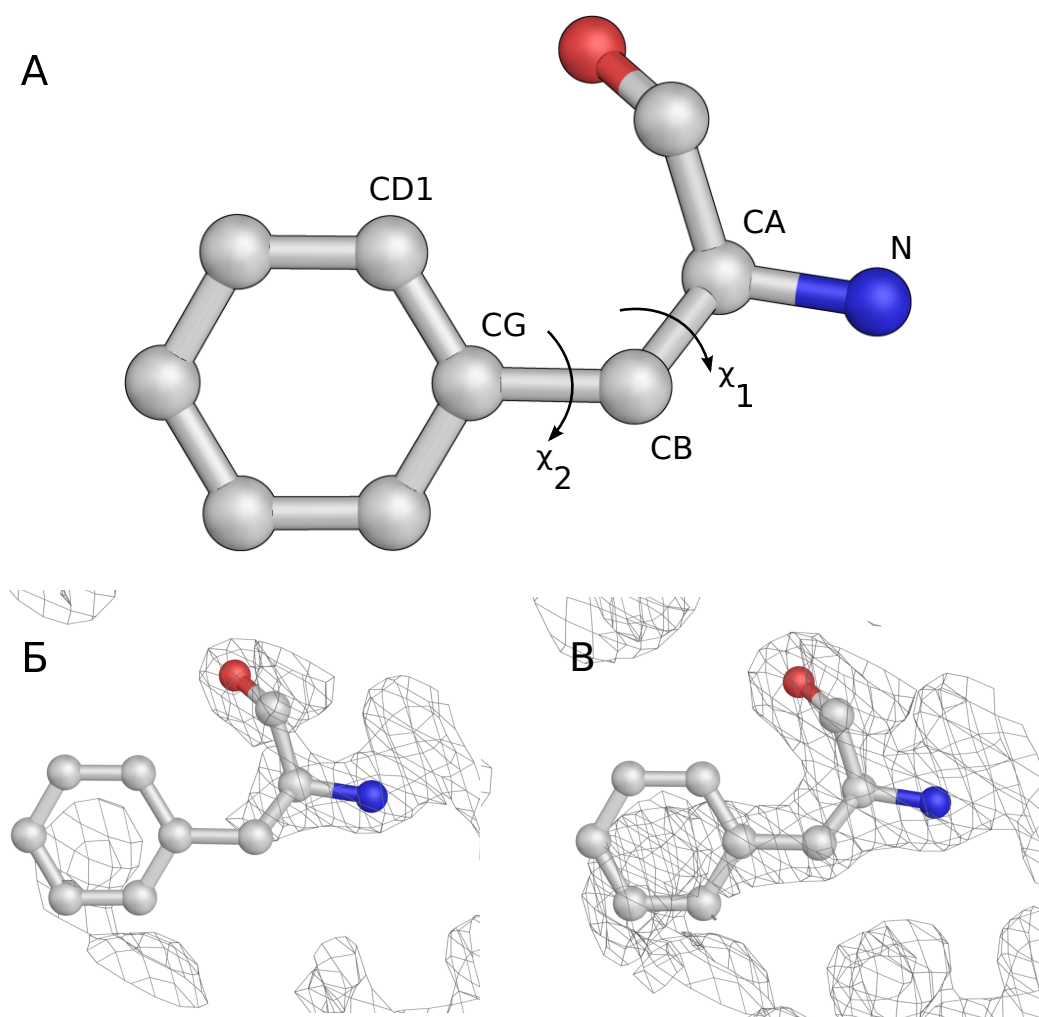


Рис. 12: А: Двугранные углы χ для фенилаланина. Атомы, образующие двугранный угол, подписаны. Б и В: Phe84 с наложенной картой электронной плоскости. Б — уровень подрезки 2σ , В — уровень подрезки 1σ .

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 13. Непонятно, почему по результатам анализа WHAT CHECK не была найдена связь с аспарагином и почему атом NH1 попал в список не образующих водородные связи, однако можно сделать вывод, что не стоит безоговорочно верить результатам автоматизированного анализа структур и что важные маргинальные остатки надо проверять вручную.

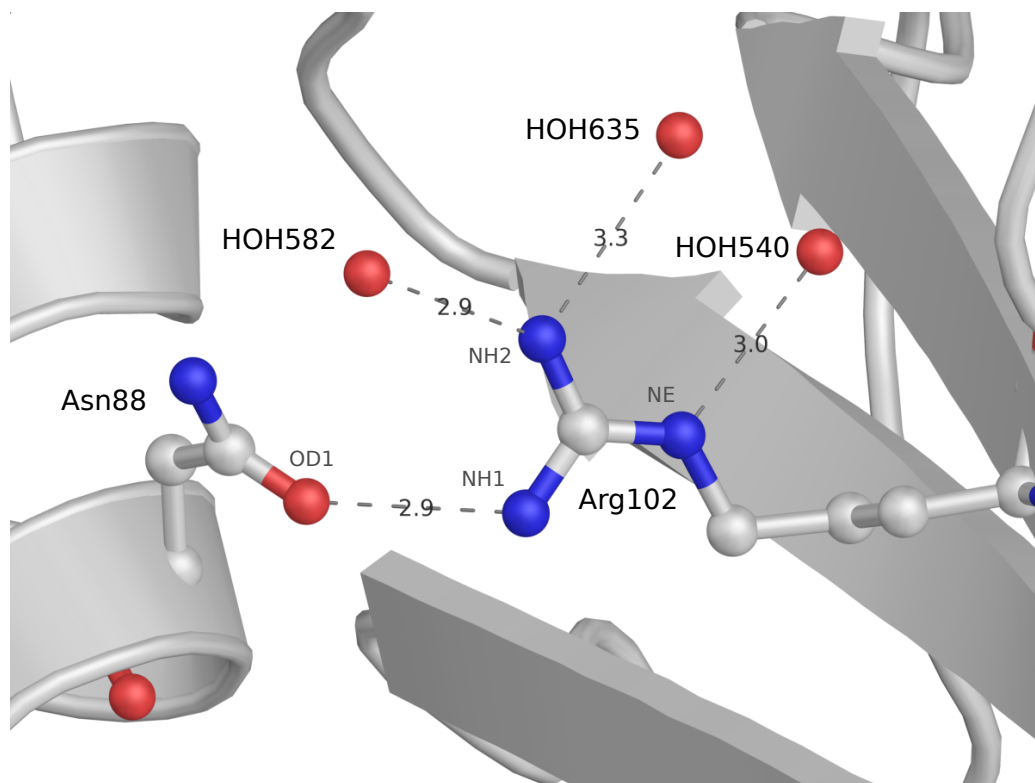


Рис. 13: Водородные связи остатка Arg102 цепи В. Взаимодействующие остатки и их атомы, а также молекулы воды, подписаны. Расстояния между взаимодействующими атомами указаны в ангстремах.

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.

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.

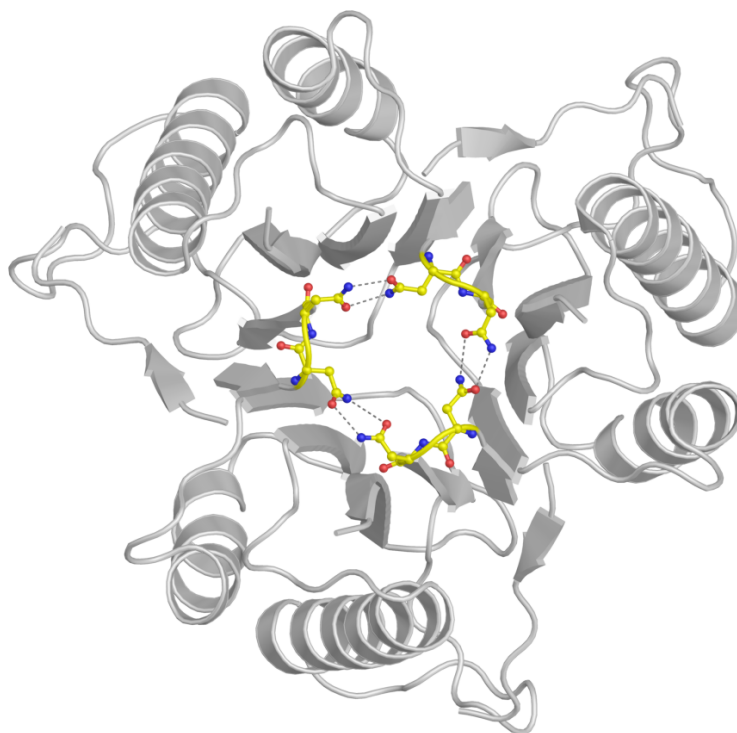


Рис. 14: “Кольцо”, образованное связанными водородными связями остатками Asn24 и Asn25 (показаны желтым цветом) мономеров белка.

Было обнаружено, что боковые радикалы Asn24 всех цепей взаимодействуют с Asn25 из соседней цепи (рис. 14). 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.

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. 15). При этом по карте электронной плотности сложно различить кислород и азот, так что такая инверсия при определе-

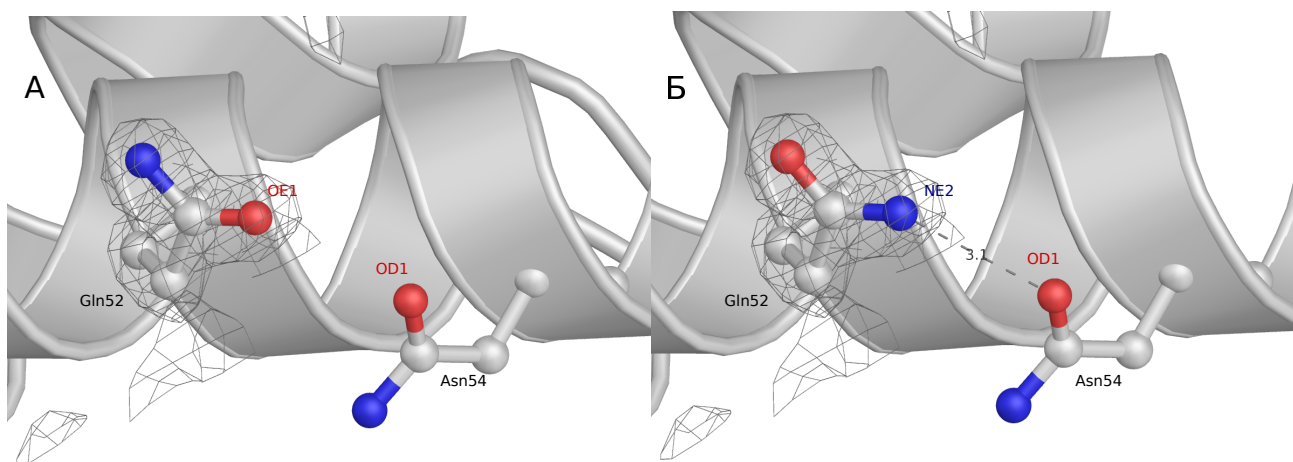


Рис. 15: Остаток Gln52 цепи A с инверсией боковой цепи. Он представлен в шаростержневой модели, как и остаток аспарагина 56 той же цепи, с которым он может образовывать водородную связь при повороте боковой группы. На рисунке показан срез электронной плотности вокруг Gln52 (уровень подрезки — 1.7σ). А: Конформация в структуре 1QD9 (водородной связи не образуется). Б: Конформация после поворота бокового радикала на 180° с помощью MolProbability, при которой образуется водородная связь.

нии конформации остатка вполне могла произойти. Других водородных связей, в том числе с водой, этот остаток глутамина ни в одной конформации не образует.

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 2.

Странно, что посчитанные из исходного pdb файла R_{work} и R_{free} не совпали с указанными авторами и посчитанными EDS.¹⁷ Возможно, в случае PDB_redo подсчитывался составной R-фактор, таким образом, по разным формулам получились разные результаты. В любом случае, оба параметра несколько снизились после оптимизации структуры, при этом разница R_{free} и R_{work} осталась практически без изменения. При этом Z-score от R_{free} уменьшился с 1.25 до -0.93. Этот параметр описывает, как соотносятся друг с другом полученный R_{free} с “предсказанным” некоторым алгоритмом. В идеальном случае Z-score должен быть близок к нулю. При этом его положительные значения свидетельствуют о недостаточной оптимизации или о том, что выбор тестового набора был неслучаен. Если же Z-score меньше нуля, возможно, структура “переподогнана”. Таким образом, можно сделать вывод о том, что после работы PDB_redo R-фактор минимизировался чуть сильнее, однако это не дает уверенности, что сама модель стала лучше.

В структуре, оптимизированной PDB_redo, больше нет маргинальных по углам ϕ и ψ остатков, также были исправлены несоответствия боковых конформаций аминокислот известным ротамерам. Однако не были исправлены случаи инверсии боковой цепи для Asn, Gln и His, RMSD для длин связей и значений углов стало больше.

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 B, 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. 16). При уровнях подрезки электронной плотности больше или равных 1σ электронной плотности, в которую можно было бы вписать боковой радикал

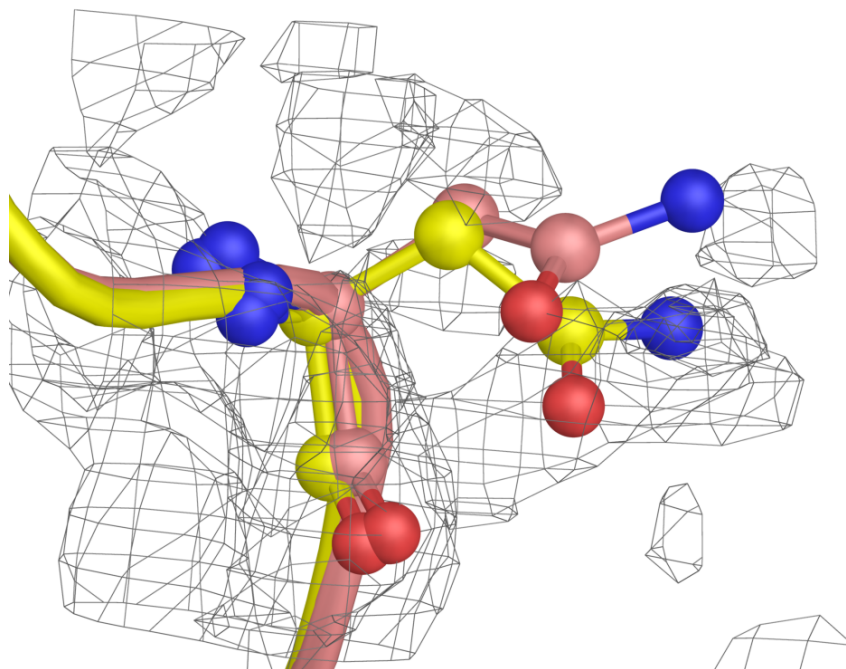


Рис. 16: Различия в конформации боковой группы Asn43 цепи В исходной структуры (показана желтым) и после оптимизации PDB_redo (розовым). Уровень подрезки электронной плотности 0.5σ .

аспарагина, не видно вовсе. При понижении уровня подрезки они видны, но нельзя сказать, настоящий ли это сигнал или шум.

Выводы: 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.

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

	Из заголовка pdb файла	Посчитанные по pdb фай- лу	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	-0.83
Ротамеры (Z-score)		-0.86	-0.70
Конформация основной цепи (Z-score)		0.46	0.51
$RMSD_{bonds}$ Z-score		0.44	0.74
$RMSD_{angles}$ Z-score		0.74	0.86
Незадействованные акцепторы/доноры водородной связи		23	22

Заключение

Из проделанной работы по оценке качества структуры можно сделать вывод о том, что в целом структура решена хорошо. 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. При этом нет оснований полагать, что маргинальность остатков не является ошибкой построения электронной плотности и вписывания в нее модель, а является их особенностью. Возможно, авторы структуры слишком перестарались с соблюдением длин связей и углов и не соблюли правильный баланс между различными критериями при уточнении структуры.

В статье хорошо описаны все структурные особенности белка, которые подтвердились другими исследованиями, в том числе и рентгеноструктурным анализом гомологов²² (эти две структуры были опубликованы одновременно, но разными группами исследователей). До публикации структуры 1QD9 в банке PDB вообще не было представителей семейства *yjgf* — большого и крайне консервативного в различныхкладах живых организмов, так что авторы не могли, например, использовать метод молекулярного замещения для решения фазовой проблемы или удостовериться в правильности разрешенной структуры по гомологии. Более того, по полученной структуре и множественному выравниванию последовательностей гомологов авторы впервые смогли предсказать активный центр белка, расположение которого потом было подтверждено экспериментально на примере близкого гомолога.¹⁴

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. Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin / N. Ogasawara, S. Nakai, H. Yoshikawa et al. // *DNA Res.* 1994. Vol. 1, № 1. P. 1–14.
2. The European Bioinformatics Institute // <http://www.ebi.ac.uk/>.
3. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis* / F. Kunst, N. Ogasawara, I. Moszer et al. // *Nature.* 1997. Nov. Vol. 390, № 6657. P. 249–256.
4. The Rut pathway for pyrimidine degradation: novel chemistry and toxicity problems / K. S. Kim, J. G. Pelton, W. B. Inwood et al. // *J. Bacteriol.* 2010. Aug. Vol. 192, № 16. P. 4089–4102.
5. Comparative genome analysis of *Salmonella Enteritidis* PT4 and *Salmonella Gallinarum* 287/91 provides insights into evolutionary and host adaptation pathways / N. R. Thomson, D. J. Clayton, D. Windhorst et al. // *Genome Res.* 2008. Oct. Vol. 18, № 10. P. 1624–1637.
6. A member of the YER057c/yjgf/Uk114 family links isoleucine biosynthesis and intact mitochondria maintenance in *Saccharomyces cerevisiae* / J. M. Kim, H. Yoshikawa, K. Shirahige et al. // *Genes Cells.* 2001. Jun. Vol. 6, № 6. P. 507–517.
7. The NADP-glutamate dehydrogenase of the cyanobacterium *Synechocystis* 6803: cloning, transcriptional analysis and disruption of the *gdhA* gene / S. Chavez, J. C. Reyes, F. Chauvat et al. // *Plant Mol. Biol.* 1995. Apr. Vol. 28, № 1. P. 173–188.
8. Ribonuclease activity of rat liver perchloric acid-soluble protein, a potent inhibitor of protein synthesis / R. Morishita, A. Kawagoshi, T. Sawasaki et al. // *J. Biol. Chem.* 1999. Jul. Vol. 274, № 29. P. 20688–20692.
9. RCSB Protein Data Bank // <http://pdb.org/>.
10. 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, № 23. P. 13074–13079.
11. Crystallization and preliminary X-ray diffraction studies of perchloric acid soluble protein (PSP) from rat liver / K. Djinovi Carugo, M. Saraste, T. Oka et al. // *Acta Crystallogr. D Biol. Crystallogr.* 1999. Mar. Vol. 55, № Pt 3. P. 667–668.
12. A role for a highly conserved protein of unknown function in regulation of *Bacillus subtilis* *purA* by the purine repressor / P. Rappu, B. S. Shin, H. Zalkin et al. // *J. Bacteriol.* 1999. Jun. Vol. 181, № 12. P. 3810–3815.

13. In vivo effect of mutations at the PRPP binding site of the *Bacillus subtilis* purine repressor / P. Rappu, T. Pullinen, P. Mantsala et al. // *J. Bacteriol.* 2003. Nov. Vol. 185, Nº 22. P. 6728–6731.
14. Conserved YjgF protein family deaminates reactive enamine/imine intermediates of pyridoxal 5'-phosphate (PLP)-dependent enzyme reactions / J. A. Lambrecht, J. M. Flynn, D. M. Downs et al. // *J. Biol. Chem.* 2012. Jan. Vol. 287, Nº 5. P. 3454–3461.
15. Proceedings of the National Academy of Sciences of the United States of America // <http://pnas.org/>.
16. Purdue University // <http://www.purdue.edu/>.
17. Uppsala Electron-Density Server // [http://http://eds.bmc.uu.se/eds/](http://eds.bmc.uu.se/eds/).
18. Improved methods for building protein models in electron density maps and the location of errors in these models / T. A. Jones, J. Y. Zou, S. W. Cowan et al. // *Acta Crystallogr., A, Found. Crystallogr.* 1991. Mar. Vol. 47 (Pt 2). P. 110–119.
19. Crystallography NMR system: A new software suite for macromolecular structure determination / A. T. Brunger, P. D. Adams, G. M. Clore et al. // *Acta Crystallogr. D Biol. Crystallogr.* 1998. Sep. Vol. 54, Nº Pt 5. P. 905–921.
20. MolProbity (Duke University) // <http://molprobity.biochem.duke.edu/>.
21. Uppsala University rotamers library // <http://xray.bmc.uu.se/gerard/supmat/chi.html>.
22. Volz K. A test case for structure-based functional assignment: the 1.2 Å crystal structure of the yjgF gene product from *Escherichia coli* // *Protein Sci.* 1999. Nov. Vol. 8, Nº 11. P. 2428–2437.