

**Московский государственный университет имени  
М. В. Ломоносова**

---

**Факультет биоинженерии и биоинформатики**

**Анализ качества расшифровки структуры  
[4 + 2]-циклазы SpnF (PDB ID: 4PNE)**

*Отчёт студента 4<sup>го</sup> курса*  
Исаева Сергея Валентиновича

Москва

2019

# Содержание

1	<b>Аннотация</b> .....	3
2	<b>Введение</b> .....	3
3	<b>Результаты и обсуждение</b> .....	4
3.1	Общая информация о модели.....	4
3.2	Значения индикаторов качества модели в целом .....	4
3.3	Краткая характеристика обнаруженных маргинальных остатков .....	6
3.4	Анализ некоторых маргинальных остатков .....	7
3.5	Сравнение модели из PDB с моделью из PDB-REDO .....	11
4	<b>Заключение</b> .....	11
4	<b>Литература</b> .....	12

## 1 Аннотация

В ходе работы был произведён анализ модели структуры белка SpnF, в ходе которого было выяснено качество определения структуры, а также произвелось сравнение с оптимизированной структурой из PDB-REDO.

## 2 Введение

[4 + 2]-циклаза SpnF, о которой речь пойдёт в этой работе, — это фермент, способный катализировать реакцию [4 + 2]-циклоприсоединения [1], которая также известна как реакция Дильса-Альдера. Он был найден у бактерии *Saccharopolyspora spinosa*, которая с его помощью может синтезировать спинозин А — тетрациклический природный инсектицид. В отличие от большинства других инсектицидов, спинозины проявляют большую селективность в отношении насекомых-вредителей и меньшую токсичность по отношению к млекопитающим и обитателям водоемов [2].

SpnF не единственный может катализировать такие хитрые реакции [3, 4 и др.]. Однако на фоне остальных белков он выделяется: если все известные ныне ферменты, способные к циклоприсоединению, выполняют множество других функций, то SpnF регулирует ход исключительно реакции Дильса-Альдера (реакцию см. на рис. 1). Фундаментальный интерес SpnF представляет ещё и потому, что механизм осуществления катализа не очень ясен. Механизмы, предложенные на сегодняшний день [5], объясняют превращение переходом через всего одно переходное состояние, в то время как привычная парадигма энзимологии говорит о необходимости большого количества интермедиатов.

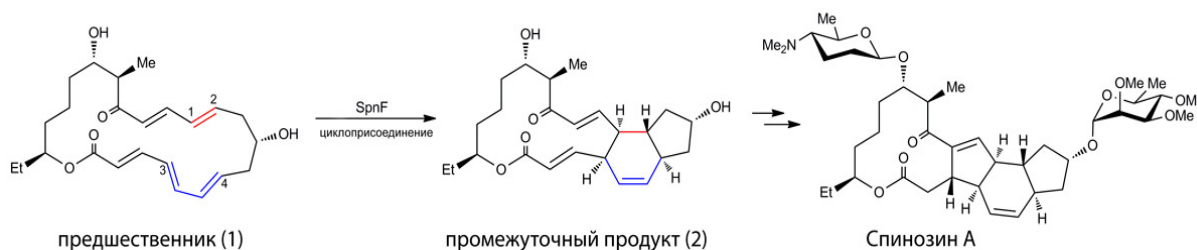


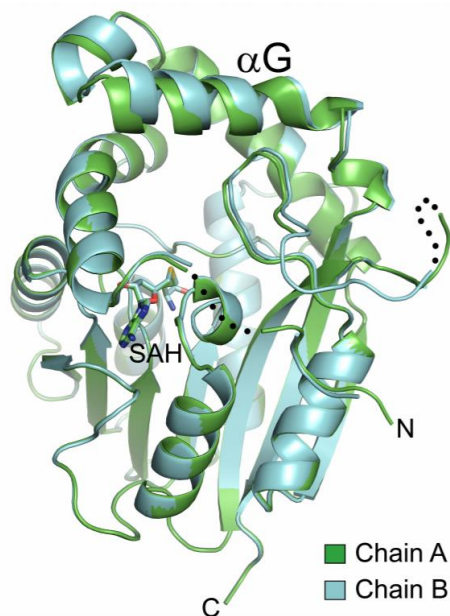
Рисунок 1. Схема реакции, катализируемой ферментом SpnF (по [5])

Расшифровка и анализ структуры белка SpnF может позволить изучить принципиально новые механизмы катализа, которые, возможно, можно будет применить и для инжиниринга уже известных молекулярных машин. Основной задачей отчёта было оценить качество расшифровки структуры, чтобы понимать, насколько можно ей доверять.

## 3 Результаты и обсуждение

### 3.1 Общая информация о модели

Исследуемый белок массой около 32.62 KDa состоит из 302 аминокислотных остатков и включает всего один описанный ранее консервативный домен из базы данных Pfam-A — метилтрансферазный. Структура 4PNE появилась на PDB 18 февраля 2015 года — за месяц до того, как вышла статья авторов этой модели в *Nature Chemical Biology* [6], у *Christopher D Fage, Eta A Isiorho, Yungnan Liu, Drew T Wagner, Hung-wen Liu* и *Adrian T Keatinge-Clay*.



**Рисунок 2.** Структурное выравнивание двух мономеров SpnF. RMSD примерно равен 0.165 Å на 1 357 атомах (по [5])

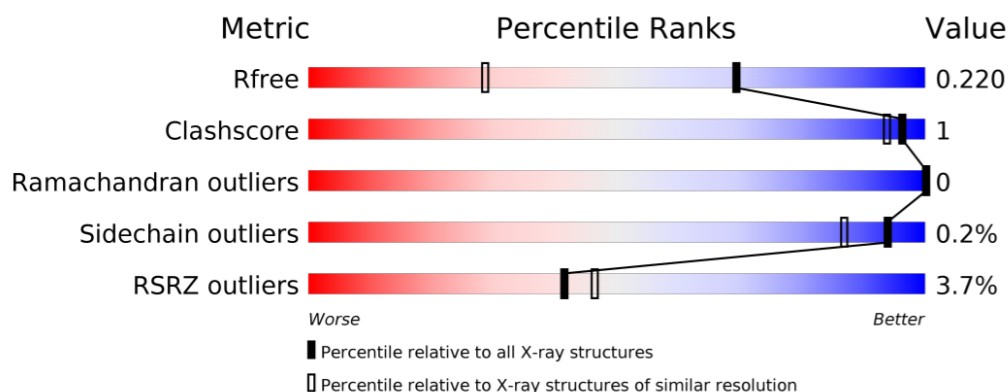
Опубликованная структура содержит две макромолекулы (асимметрическая единица структуры состоит из гомодимера SpnF — структурное выравнивание представлено на рисунке 2.), также два лиганда — малонат и S-аденозил-L-гомоцистеин.

Структура имеет неплохое разрешение в 1.5Å. Получена она была методом рентгеноструктурного анализа; фазовая проблема решалась методом молекулярного замещения — в PDB уже существовала структура с похожей последовательностью (3BUS). Кристаллографическая ячейка модели имеет следующие параметры: углы почти ортогональны ( $\alpha = 90^\circ$ ,  $\beta = 97.49^\circ$ ,  $\gamma = 90^\circ$ ), длины же базисных векторов равны соответственно 41.74 Å, 69.35 Å и 83.63 Å. Эти данные основываются на 74 159 структурных факторах, 5% из которых использовались как контрольные. Полнота набора структурных факторов оценивается EDS в 98%. В одной кристаллографической ячейке находятся 4 молекулы белка. Некристаллографических симметрий нет.

### 3.2 Значения индикаторов качества модели в целом

Посмотрим на метрики качества расшифровки структур, которые предлагает PDB Report. Среди этих метрик основные (см. рис. 3):

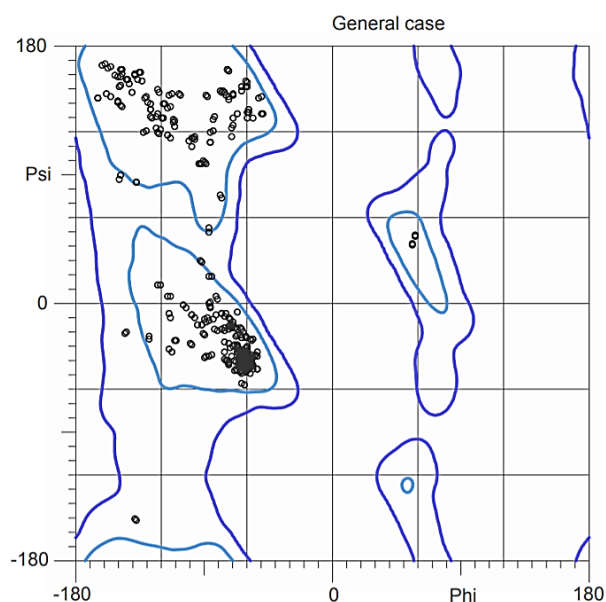
- *R*-фактор — это мера соответствия модели её экспериментальным значениям (разница наблюдаемых и ожидаемых модулей, нормированная на наблюдаемые



**Рисунок 3.** Схематическое изображение основных метрик качества расшифровки структуры 4PNE. Чем ближе засечка к правому концу, тем меньше значение этого параметра даёт нам засомневаться в правильности определения структуры

модули структурных факторов). Характеристика  $R_{free}$ , которая аналогична  $R$ -фактору, считается только на контрольных рефлексах, и она может говорить о переоптимизации модели. Значение  $R$ -фактора для рассматриваемой модели 0.198, а  $R_{free}$  — 0.22. Оба этих параметра соответствуют хорошему качеству структуры; также маленькая разница между этими двумя параметрами (0.022) говорит о том, что модель не переоптимизирована.

- Показатель **перекрывания Ван-дер-Ваальсовых радиусов** (*clashscore*) равен количеству перекрывающихся Ван-дер-Ваальсовых больше, чем на 0.4 Å на 1000 атомов структуры. Для 4PNE этот параметр равен 1, что лучше, чем у 99% структур с таким же разрешением — по данному критерию структура хорошая.
- Карта Рамачандрана (график, в котором каждая точка — одна аминокислота; по оси  $x$  — торсионный угол  $\phi$  остова, по оси  $y$  — угол  $\psi$ ). Распределение точек на карте Рамачандрана является характеристикой энергетической выгодности нахождения какой-либо аминокислоты в её положении в структуре. Соответственно, можно условно определить на карте «запрещённые» области, в которых будут находиться конформации с высокими энергиями. Мера того, насколько много таких аминокислот



**Рисунок 4.** Карта Рамачандрана, построенная с аминокислотами структуры 4PNE. Запрещённые области находятся за синими границами -- в них не попало ни одной аминокислоты

в структуре, — это доля таких остатков (их называют маргинальными) от всего числа аминокислот (**Ramachandran outliers**). В исследуемой структуре нет ни одной аминокислоты в запрещённых областях, что однозначно говорит в её пользу (см. рис. 5).

- Доля **маргиналов по положению в боковой цепи** (ротамеров) (*Sidechain outliers*) — это ещё одна важная характеристика качества структуры. Вообще, ротамеры — это *боковые цепи* в предпочитаемых для данного остатка положениях. Они описываются наборами допустимых значений торсионных углов радикалов. Иначе говоря, если карта Рамачандрана позволяет выявить маргинальные остатки, судя по их поведению в осто́ве, то анализ ротамеров позволяет сделать то же, но уже судя по расположению их радикалов. Отчёт по качеству структуры PDB Report говорит об одном маргинальном остатке (97Gln(B)), инструмент MolProbity — о двух (97Gln(B) и 117Glu(B)). Интересно заметить, что структура, которая анализировалась, содержала информацию о гомодимере, и тот факт, что маргинальные остатки обнаружались только в одной цепи, может говорить о неправильном определении именно её. К тому же, именно регион, содержащий эти остатки, получился наименее консервативным при выравнивании этих двух белков (см. рис. 2). Итоговая доля «плохих» ротамеров составляет (в худшем случае) 0.48%, что не очень хорошо, но и не кричит о плохом качестве.
- *R*-фактор, о котором речь уже шла выше, можно считать не только по всей модели, но и на каких-то конкретным аминокислотных остатках (пространственный *R*-фактор, или RSR). Такая мера качества может помочь нам определить конкретные маргинальные участки в структуре. Но как понять, наблюдаемое отклонение нормально или нет? Для этого существует *z*-score, который является относительной оценкой RSR, то есть показывает, на сколько среднеквадратичных отклонений RSR группы остатков/одного остатка отличается от среднего значения RSR по иным моделям похожего разрешения. В нашем случае порог **RSRZ** было решено оставить на двойке: при таких условиях было найдено 20 сомнительных остатков, которые необходимо в дальнейшем проанализировать более тщательно.

### 3.3 Краткая характеристика обнаруженных маргинальных остатков

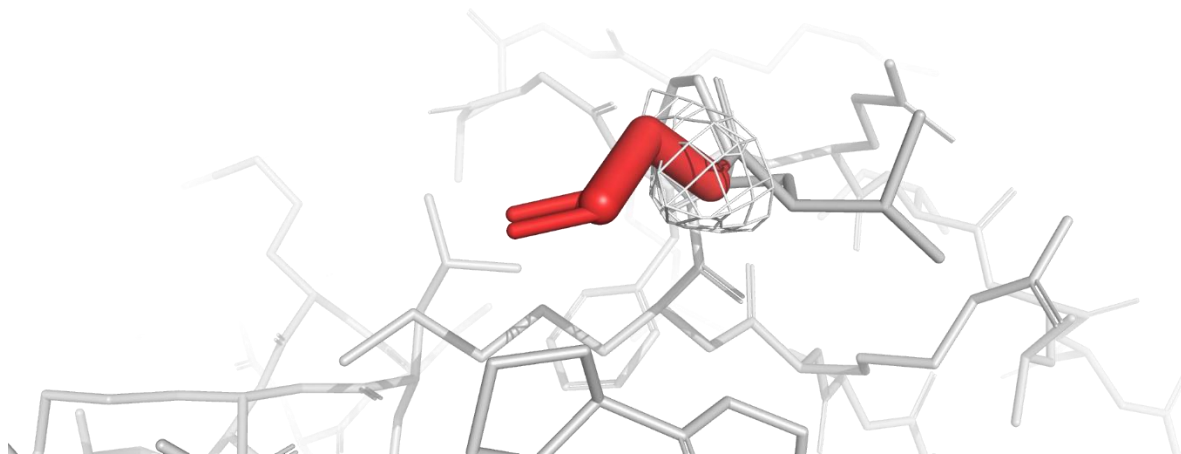
В таблице 1 показаны основные обнаруженные мной маргинальные остатки, которые могут свидетельствовать о недостатках модели.

**Таблица 1.** Перечисление нескольких обнаруженных маргинальных остатков.

Аминокислотные остатки	Критерий
282Gly(B); 282Gly(A); 13Ala(B); 242Gln(B) и проч.	Маргинальные остатки, обнаруженные при помощи анализа RSRZ
85Gln(B) + 86Pro(B); 23Tyr(A) + 26Val(A); 28Pro(A) + 249Tyr(A) и проч.	Перекрытия Ван-дер-Ваальсовых радиусов
97Gln(B); 117Glu(B)	Неблагоприятная конформация боковых цепей

### 3.4 Анализ некоторых маргинальных остатков

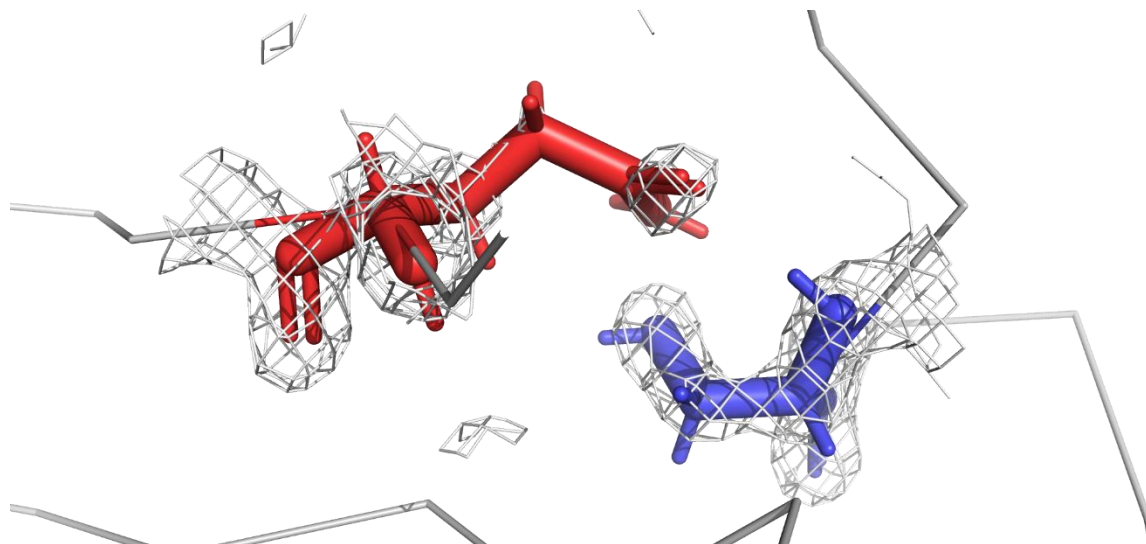
- **282Gly(B)** — этот остаток является С-концевым у всего белка (на изображении он выделен красным, см. рис. 5). Даже на уровне подрезки 1 сложно идентифицировать электронную плотность вокруг этого глицина. По всей видимости, позиция этого остатка легко могла быть определена неправильно, но принципиальной роли это не играет, т. к. функционально он бессмысленен.



**Рисунок 5.** Участок белка с 282Gly. По всей видимости, координаты аминокислотного остатка определены неточно, но это не говорит о качестве структуры в целом

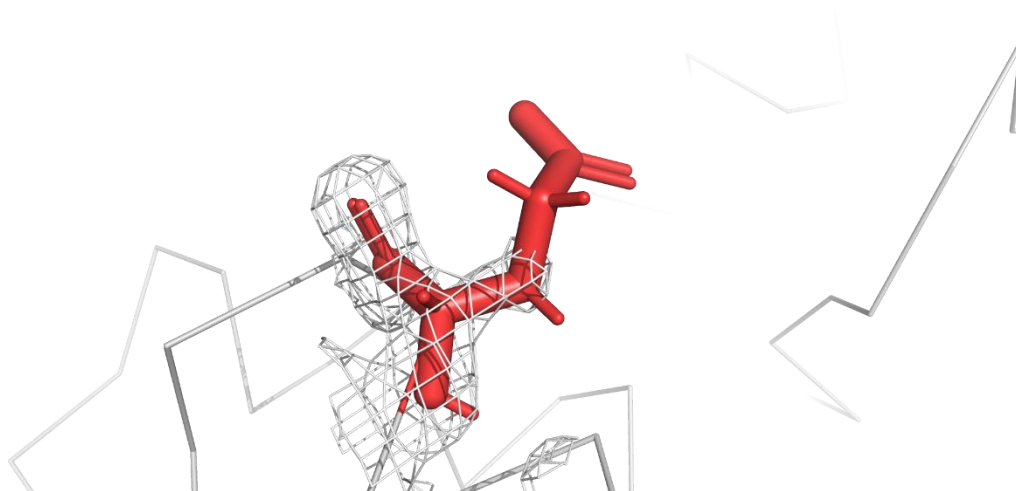
- **97Gln(B)** и **225Ser(B)**. Для 97Gln(B) известно, что в структуре он имеет неблагоприятную конформацию боковой цепи. Для того, чтобы понять, с чем это может быть связано, было решено посмотреть на взаимодействие с соседними структурами. Оказалось, 97Gln(B) близко соседствует с 225Ser(B) из соседней ячейки, и таким образом у нас в структуре есть взаимодействие, которое мы не предполагаем. Электронная плотность вокруг 97 остатка сконцентрирована не

очень чётко, но достаточно, чтобы понять, что, вероятно, определение положения глутамина верно в контексте кристалла (см. рис. 6). Интересно, что этот остаток не посчитался маргинальным в другой цепи гомодимерной структуры. По всей видимости, это как раз объясняется тем, что другая цепь (chain A) в кристалле не взаимодействует с соседними ячейками, и «неестественная» конформация там не формируется.



**Рисунок 6.** 97Gln(B) покрашен красным, 225Ser(B) -- синим. Видно, что радикалы могут взаимодействовать, что может вызывать смещение электронной плотности и вызывать трудности в определении позиции глутамина. Непонятно, нативно ли такое его положение, или же оно возникает только по ходу кристаллизации.

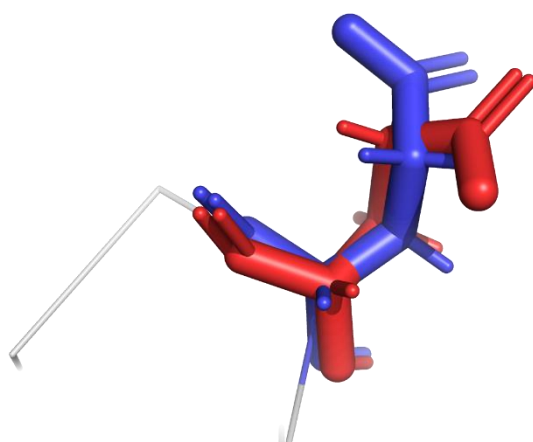
- Ещё один остаток, качество расшифровки которого непонятно, — это **117Glu(B)**. Этот остаток тоже находится на поверхности, однако не взаимодействует с белками из соседних ячеек. Восстановление электронной плотности показало, что она действительно не позволяет точно сказать о расположении атомов в аминокислоте (см. рис. 7).



**Рисунок 7.** Остаток 117Glu(B) с отображённой вокруг электронной плотностью. Видно, что даже на таком уровне подрезки (1.5) сказать что-то о точной геометрии остатка в пространстве сложно

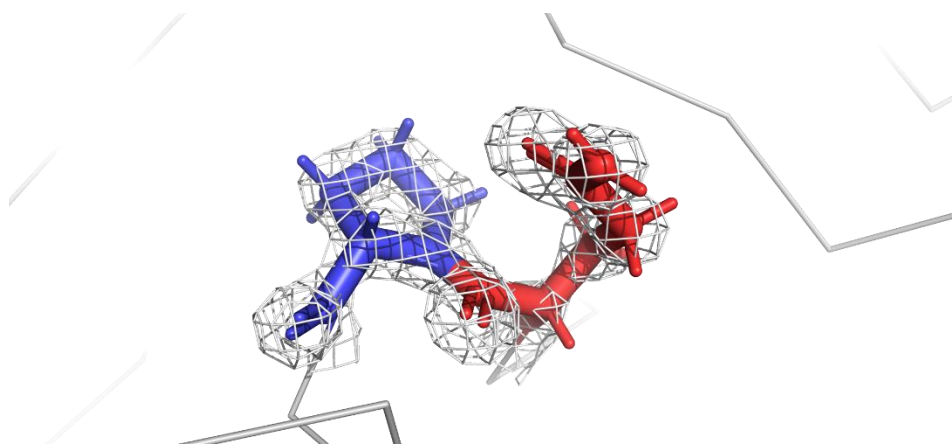


Интересным кажется, что такая же аминокислота из цепочки А не попала в список маргинальных аминокислот. Для того, чтобы понять, в чём между ними разница, сначала та же процедура восстановления электронной плотности была произведена вокруг симметричного остатка в цепи А: оказалось, что там, по всей видимости, информации не больше. После выравнивания этих двух аминокислотных остатков между собой (см. рис. 8) стало понятно, что геометрически они отличаются: видимо, конформация в цепочке А ближе к привычной для глутамата, и она не вызывает подозрений. Есть основания считать, что 117Glu(B) и правда предсказан не очень точно.



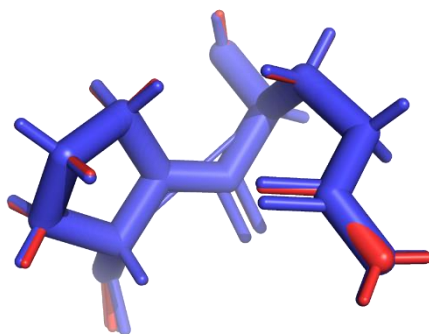
**Рисунок 8.** Попарное выравнивание двух 117 глутаматов из цепочек А и В.

- Далее интересно было проанализировать пару **85Gln(B) + 86Pro(B)**, которые расположены слишком близко друг к другу. Для того, чтобы понять, правда ли это, сначала было решено посмотреть на распределение электронной плотности вокруг остатков. Электронная плотность хорошо позволяет определить координаты атомов (см. рис. 9).



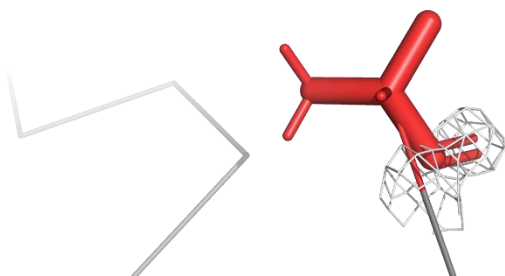
**Рисунок 9.** Остатки 85Gln(B) и 86Pro(B) и электронная плотность, по которой они были восстановлены

Эта структура оказалась консервативной между цепочками: при выравнивании этого участка от цепочки А и цепочки В получается результат с RMSD = 0.062 (см. рис. 10). По всей видимости, расположение остатков определено верно, просто такое случается. Попытки физико-химически объяснить данное явление у меня не пришли к успеху.



**Рисунок 10.** Структурное выравнивание группы остатков 85Gln+86Pro. Расположение атомов в пространстве консервативно между субъединицами гомодимера

- Интересна ситуация и с остатком **13Ala(B)**. Он попал в список маргинальных элементов из-за высокого значения RSR. Если посмотреть на электронную плотность вокруг этого остатка, то станет понятно, почему — по ней сложно восстановить расположение атомов в пространстве (см. рис. 11). Однако стоит отметить, что относительно участка с аминокислотными остатками 8-12 на цепочке В в статье со структурой честно пишут, что не смогли его разрешить (как и несколько других небольших выпетливаний на поверхность белка). В статье также указана спорность определения ранее описанного в отчёте остатка 282Gly(B), что тоже делает авторам честь.



**Рисунок 11.** Неопределённость в определении геометрии остатка 13Ala(B)

### 3.5 Сравнение модели из PDB с моделью из PDB-REDO

В таблице 2 представлено сравнение модели белка SpnF из баз данных PDB и PDB-REDO (базы данных с оптимизированными по сравнению с PDB моделями) — по основным характеристикам.

**Таблица 2.** Сравнение различных мер качества на моделях 4PNE из PDB и PDB-REDO (зелёным обозначены принципиально более «качественные» значения, красным — наоборот)

	<b>PDB</b>	<b>PDB-REDO</b>
<i>R</i> -фактор	0,2009	0,2202
<i>R</i> <sub>free</sub>	0,2226	0,2383
RMS Z-score по длине связей	0,481	0,635
RMS Z-score по углу связей	0,685	0,815
Характеристика карт Рамачандрана	89	88
Характеристика ротамеров	50	53

Как мы можем наблюдать, в основном модель из PDB показывает себя лучше, чем модель из PDB-REDO, что говорит об ответственном отношении к обработке результатов эксперимента PCA.

## 4 Заключение

Модель структуры 4PNE, рассматриваемая в отчёте, расшифрована очень хорошо и неглупо: видно, что авторы статьи удачно выбрали метод борьбы с фазовой проблемой, а также внимательно обрабатывали каждые частные сложности в интерпретации данных. Практически обо всех аминокислотах со спорным структурным статусом прямо обговорено в статье, некоторые недостоверно определившиеся результаты были вырезаны из структуры и вовсе. В результате мы имеем модель с высокими показателями качества, которую можно не боясь использовать в биоинформатических экспериментах с моделированием ферментативной активности.

## 4 Литература

1. Kim, H. J., Rusczycky, M. W., Choi, S., Liu, Y., & Liu, H. (2011). *Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A*. *Nature*, 473(7345), 109–112. doi:10.1038/nature09981
2. Yano BL, Bond DM, Novilla MN, McFadden LG, Reasor MJ. (2002). *Spinosad insecticide: subchronic and chronic toxicity and lack of carcinogenicity in Fischer 344 rats*. *Toxicol Sci*. Feb;65(2):288-98.
3. Oikawa, H., Suzuki, Y., Naya, A., Katayama, K., & Ichihara, A. (1994). *First Direct Evidence in Biological Diels-Alder Reaction of Incorporation of Diene-Dienophile Precursors in the Biosynthesis of Solanapyrones*. *Journal of the American Chemical Society*, 116(8), 3605–3606. doi:10.1021/ja00087a059
4. Stocking, E. M., & Williams, R. M. (2003). *Chemistry and Biology of Biosynthetic Diels–Alder Reactions*. *Angewandte Chemie International Edition*, 42(27), 3078–3115. doi:10.1002/anie.200200534
5. Gordeev, E. G., & Ananikov, V. P. (2015). *Computational Study of a Model System of Enzyme-Mediated [4+2] Cycloaddition Reaction*. *PLOS ONE*, 10(4), e0119984. doi:10.1371/journal.pone.0119984
6. Fage, C. D., Isiorho, E. A., Liu, Y., Wagner, D. T., Liu, H., & Keatinge-Clay, A. T. (2015). *The structure of SpnF, a standalone enzyme that catalyzes [4 + 2] cycloaddition*. *Nature Chemical Biology*, 11(4), 256–258. doi:10.1038/nchembio.1768
7. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., ... Richardson, D. C. (2009). *MolProbity: all-atom structure validation for macromolecular crystallography*. *Acta Crystallographica Section D Biological Crystallography*, 66(1), 12–21. doi:10.1107/s0907444909042073