

Московский государственный университет  
имени М.В. Ломоносова

Факультет биоинженерии и биоинформатики

Отчет по структурной биоинформатике студента четвертого курса  
Михаила Александровича Молдована

**Качество расшифровки структуры белка Tie2 (PDB-идентификатор  
2RY5) методом рентгеноструктурного анализа**

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

Москва, 2016

## **Оглавление**

Введение .....	3
Поиск маргинальных остатков .....	7
Более глубокий анализ пяти маргинальных остатков .....	16
Сравнение с моделью из PDB_REDO .....	20
Выводы.....	23
Список литературы .....	24
Ссылки .....	25

## **Введение**

### **Общая информация о белке Tie-2**

Развитие сердечно-сосудистой системы включает дифференцировку эндотелиальных клеток из предшественников-ангиобластов и их последующие миграции, рост и прорастание. В норме, оно протекает через два относительно различных процесса: васкулогенез – формирование зачатков крупных кровеносных сосудов и ангиогенез – формирование рабочих крупных сосудов, а также мелких. Во взрослом организме васкулогенез не активен, а ангиогенез используется для заживления ран.

Также, ангиогенез задействован на последних стадиях развития злокачественных опухолей, которым для ускоренного роста необходимы питательные вещества и кислород. Вместе с тем, без активации ангиогенеза, опухоль не может давать метастазы.

В работах на мутантных мышах было показано, что активация ангиогенеза идет через ангиопоэтины (белки Ang), для которых был найден рецептор Tie-2 [1,8]. Мыши с нарушенной экспрессией белков Ang или Tie-2, демонстрировали ослабленное развитие сосудистой ткани, а позже было установлено, что гиперэкспрессия белка Tie-2 ассоциирована с многими типами рака [2], в особенности с высокоинвазивными и агрессивными раками.

Собственно рецепторы ангиопоэтинов состоят из белков Tie-1 и Tie-2, которые являются рецепторными тирозинкиназами. При этом было показано, что взаимодействует со всеми четырьмя ангиопоэтинами именно Tie-2, в то время, как для белка Tie-1 лигандов найдено не было.

Для того, чтобы понять, как именно происходит взаимодействие ангиопоэтинов с белком Tie-2, командой ученых из университета Вирджинии под руководством Димитара Николова [2] было проведено кристаллографическое исследование структуры Tie2 без лиганда и структуры комплекса Tie-2 с ангиопоэтином-2.

Соответственно, в результате работы в PDB было опубликовано две структуры. Далее рассматривается только структура белка Tie-2 без ангиопоэтина, точнее структура его внеклеточного домена.

### **Измеренные рефлексы**

В ходе одного эксперимента было получено 34586 рефлексов при длине волны в 1.215 Å. В оригинальном cif-файле все эти рефлексы лежат в виде структурных факторов (F) и интенсивностей (I), которые связаны уравнением  $I = F^2$ . При этом, 1527 интенсивностей имеют отрицательные значения, а соответствующие им значения F и стандартные отклонения для F принимаются за 0. Вместе с этим, полнота данных, получаемая по интенсивностям и структурным факторам, различна. Так, 28039 значений F превышают три своих стандартных отклонения, что соответствует значению 81.0%. Но то же самое можно сказать про всего 24998 значений I, что соответствует значению 72.28%. Среднее отношение величины не принятых за ноль структурных факторов к стандартному отклонению составляет 13.05, для интенсивностей – 13.46. Таким образом, исходные данные в среднем имеют не очень высокое качество, сделано достаточно большое количество недостоверных измерений.

Скорее всего, это вынудило авторов взять неполный набор рефлексов. Было взято 26227 рабочих рефлексов и 2832 тестовых. Какие именно рефлекссы были выбраны для построения модели, нигде не указано.

Согласно сервису EDS, полнота данных в диапазоне 46.74-2.80 Å составляет 94.9%, а заявленное разрешение структуры составляет 2.90 Å. Согласно этому показателю, данные хорошие, и настоящее разрешение структуры лежит действительно в районе 2.9 Å, то есть разрешение атомарное.

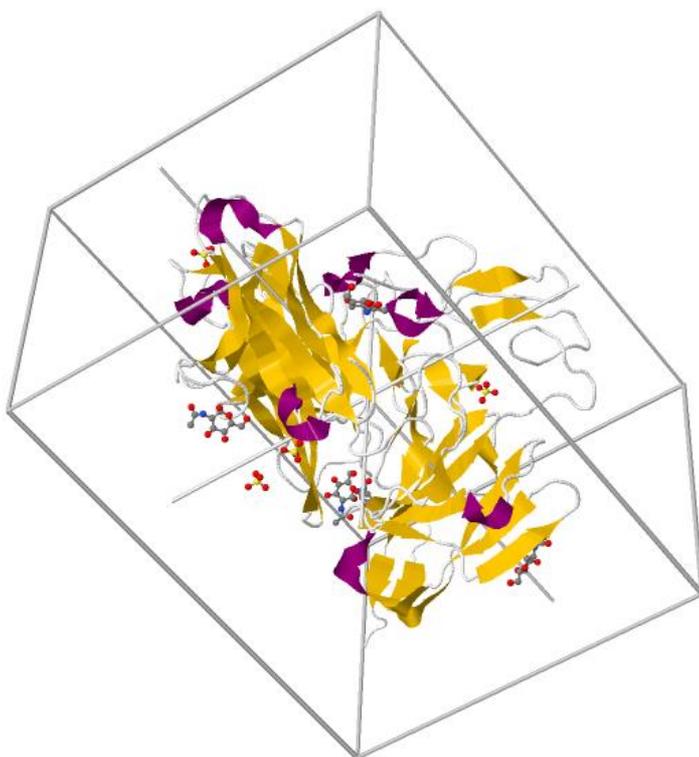
Согласно же PDB-файлу, диапазон данных был 20-2.8 Å, а полнота данных – 93%. В статье же указано третье значение полноты – 96.9% в диапазоне 2.8-50.0 Å.

### Параметры кристаллографической ячейки

Поле CRYST1 в PDB-файле выглядит следующим образом:

```
CRYST1 114.653 114.653 113.774 90.00 90.00 90.00 P 41 21 2 8
```

Это значит, что в элементарной ячейке кристалла содержится 8 молекул, а в ассиметрической единице – одна молекула (рис. 1). В данном случае ассиметрическая единица соответствует биологической.



**Рисунок 1.** Ассиметрическая единица структуры 2GY5, соответствующая биологической единице. Изображение получено на сайте RCSB [11]

Интересно, что для этой структуры приводится несколько наборов значений параметров кристаллографической ячейки.

В cif-файле, лежащем в PDB, дается два альтернативных набора значений параметров a, b и c:

114.653 114.653 113.774 (первый набор)

114.456 114.456 113.689 (второй набор)

В статье же приводится третий набор, полученный, согласно авторам, после процедур улучшения структуры:

114.900 114.900 113.89

При этом в PDB-файле и на сервисе EDS указан только первый набор.

Значения углов альфа, бета и гамма везде одинаковые и все они равны 90 градусов.

### **Решение фазовой проблемы**

Фазовая проблема была решена методом изоморфного замещения: в структуру белка было введено 12 тяжелых атомов, положения которых были определены с помощью SnB.

### **Дополнительная обработка и оценка качества структуры авторами**

Модель была построена с помощью пакета O [3]. Последующая обработка, включавшая молекулярную динамику и минимизацию энергии сворачивания, была проведена с помощью пакета CNS [4]. Дальнейший анализ модели программой PROCHECK показал приемлемое качество полученной структуры.

Как уже было сказано, из исходных 34586 рефлексов было отобрано 26227 рабочих и 2832 тестовых. По рабочим рефлексам была построена модель, которая потом была улучшена, и были посчитаны параметры ее качества, указанные в таблице 1.

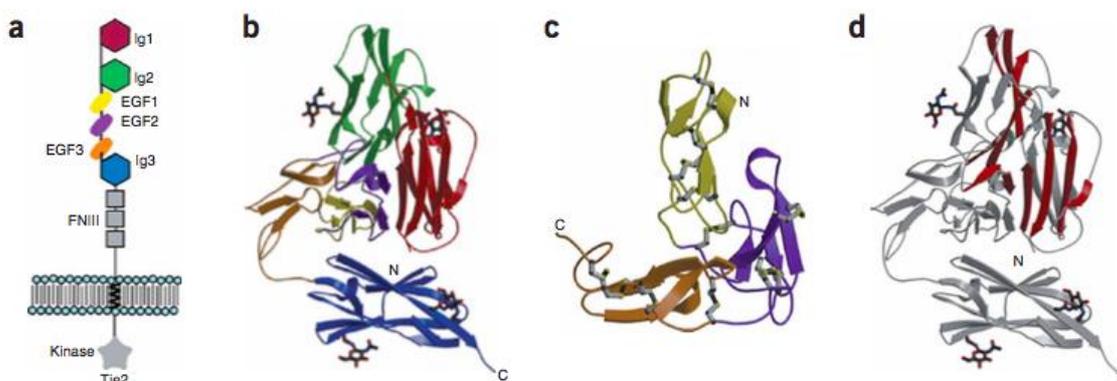
**Таблица 1.** Значения параметров, характеризующих качество модели

<b>Параметр</b>	<b>Значение, указанное в статье</b>	<b>Значение, указанное в EDS/PDB</b>
I/σI	12.8	-
Полнота данных	96.9 (диапазон 50.0-2.8 Å)	94.9 (диапазон 46.74-2.8)
R <sub>work</sub> (%)	24.0	24.0
R <sub>free</sub>	29.3	29.4
R <sub>work</sub> – R <sub>free</sub>	5.3	5.4
Средний температурный фактор (белок)	29.8	-
Средний температурный фактор (лиганды)	66.8	-
Дисперсия длин связей (Å)	0.008	-
Дисперсия величин углов	1.63	-
Real-space R	-	0.219

Интересно, что значения, приведенные в файле PDB и в статье, несколько разнятся, но разница не принципиальна. Указанное R<sub>free</sub> говорит о высоком качестве модели и о чистоте кристаллографов.

## Состав комплекса и домены

Кристаллическая структура показала, что внешний домен белка Tie2 имеет размеры 90X65X50 Å. Он содержит три домена Ig и три домена EGF (epitelium growth factor), которые компактно свернуты (рис. 2).



**Рисунок 2.** Структура лиганд-связывающего региона белка Tie2. (a) Схематично показанная доменная организация. Ig – иммуноглобулин-подобные домены, EGF – домены рецептора эпителиального фактора роста, FN III – повторы, похожие на фибронектин типа 3. (b) Внеклеточный лиганд-связывающий домен белка Tie2. Структура покрашена по доменам, цвета соответствуют таковым на рисунке (a). (c) Участок структуры, соответствующий EGF. (d) лиганд-связывающая область белка, представленная большим (11 цепей) и сильно изломанным β-листом.

В статье Бартона с коллегами [2] было показано, что белок Tie2 связывает все четыре ангиопоэтина примерно одинаково. Последовавшие эксперименты с направленным мутагенезом позволили определить конкретные аминокислоты, участвующие во взаимодействии. Также была показана интересная архитектура самого белка, в которой три иммуноглобулиновых домена и три домена EGF сворачиваются друг относительно друга, создавая непрерывный бета-лист из одиннадцати цепей. При этом создается большая поверхность, заходящая как бы внутрь белковой глобулы.

## Публикация структуры в PDB и в PDB\_REDO

О публикации структуры 2RY5, вместе со структурой 2RY7, было объявлено в статье «Crystal structures of the Tie2 receptor ectodomain and the angiopoietin-2-Tie2 complex.», которая вышла в журнале nature structural and molecular biology 13 июня 2006 года.

Авторы: Barton WA, Tzvetkova-Robev D, Miranda EP, Kolev MV, Rajashankar KR, Himanen JP, Nikolov DB.

Структура была опубликована в PDB девятого мая 2006 года, авторы указаны те же.[12]

Файл структурных факторов был опубликован шестого июня 2006 года и подвергся стандартизации формата в 2013 году.

Запись о структуре в PDB\_REDO появилась 18 июня 2013 года. Валидация проводилась программой WHAT\_CHECK [5], и авторы опубликовали полный отчет о валидации [13].

## Поиск маргинальных остатков

### Работа с сервисом MolProbity

Для оценки качества структуры по параметрам, характеризующим качество геометрического расположения остатков, я воспользовался сервисом MolProbity [14]. В онлайн-форме можно задать как PDB-идентификатор, так и загрузить пользовательский PDB-файл. Далее проводится процедура восстановления положений атомов водорода, которая может включать (а может и не включать – это решает пользователь) инверсию остатков аспарагина, глутамина и гистидина, которая проводится, если положение этих остатков в файле хуже своей альтернативы в смысле таких параметров, как сила перекрывания атомных орбиталей соседних остатков и вероятность водородной связи в каждой конформации. Чем сильнее перекрытие – тем сильнее межатомное отталкивание, и тем структура менее стабильна, а, значит, и менее вероятна. Очевидно также, что чем выше вероятность образования водородной связи у структуры, тем она более вероятна. В данном случае инверсия не проводилась с целью наиболее полным образом характеризовать отрицательные стороны изучаемой структуры. Однако было проведено поверхностное сравнение параметров структуры без инверсий и структуры с инверсиями (см. далее).

Общая характеристика выдачи программы MolProbity приведена в таблице 2.

Можно заметить, что у этой структуры достаточно большой процент «плохих» с точки зрения карт Рамачандрана и статистики по ротамерам остатков. Также наблюдаются средние значения таких параметров, как качество положения C $\alpha$ -атомов и значения углов. Однако в структуре отсутствуют «плохие» связи, практически не наблюдается излишнего перекрывания атомных орбиталей и программа MolProbity дает структуре в целом положительную оценку, лежащую в 70 перцентили общего распределения.

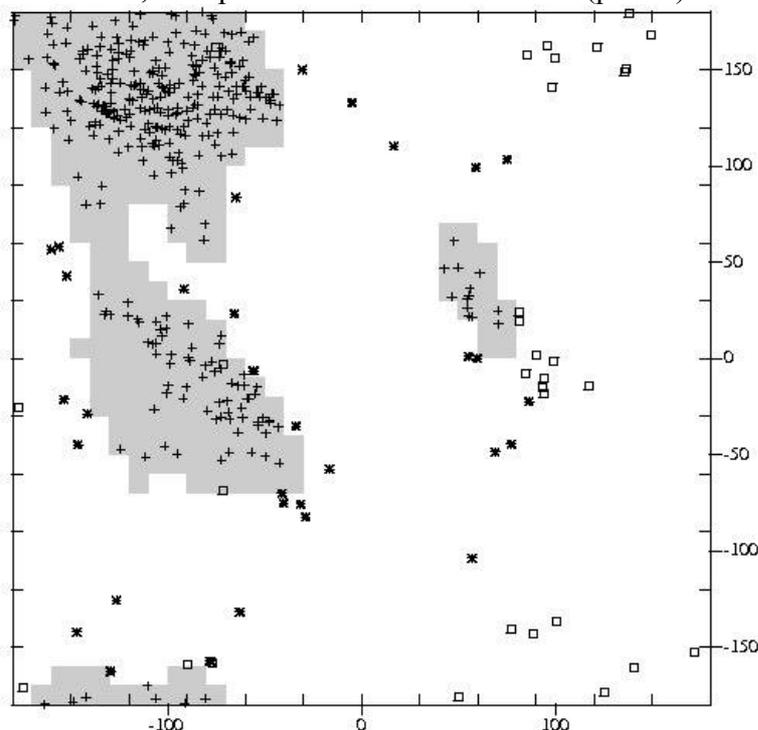
**Таблица 2.** Общая характеристика выдачи программы MolProbity

All-Atom Contacts	Clashscore, all atoms:	19.57		95 <sup>th</sup> percentile* (N=97, 2.90Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	24	6.56%	Goal: <0.3%
	Favored rotamers	314	85.79%	Goal: >98%
	Ramachandran outliers	16	3.80%	Goal: <0.05%
	Ramachandran favored	372	88.36%	Goal: >98%
	MolProbity score <sup>^</sup>	3.00		70 <sup>th</sup> percentile* (N=3760, 2.90Å ± 0.25Å)
	C $\beta$ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 3476	0.00%	Goal: 0%
	Bad angles:	7 / 4728	0.15%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 29	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	16	3.82%	Goal: <1.0%
	CA Geometry outliers	4	0.95%	Goal: <0.5%

Далее описывается поиск маргинальных остатков по различным параметрам. Можно все это не читать, так как вся информация собрана в таблице 4.

## Маргинальные остатки на карте Рамачандрана

На сервисе EDS [15] лежит карта Рамачандрана для структуры 2GY5 с указанными остатками, которые в нее не вписываются (рис. 3).



**Рисунок 3.** Карта Рамачандрана с сервиса EDS. Серым показаны регионы, содержащие 99% остатков

На сервисе MolProbity можно получить целый набор карт Рамачандрана, на которых различаются 6 случаев, для которых было показано, что в них статистика углов  $\phi$  и  $\psi$  должна быть разная [6], благо большой размер банка PDB позволяет ее собрать:

1. Цис-пролин
2. Транс-пролин
3. Глицин
4. Остатки, после которых идет пролин
5. Валин и изолейцин
6. Все остальные случаи

Сервис MolProbity находит в 2 раза меньше (16) маргинальных остатков (рис. 4), чем указано на сервисе EDS, но зато с большей достоверностью.

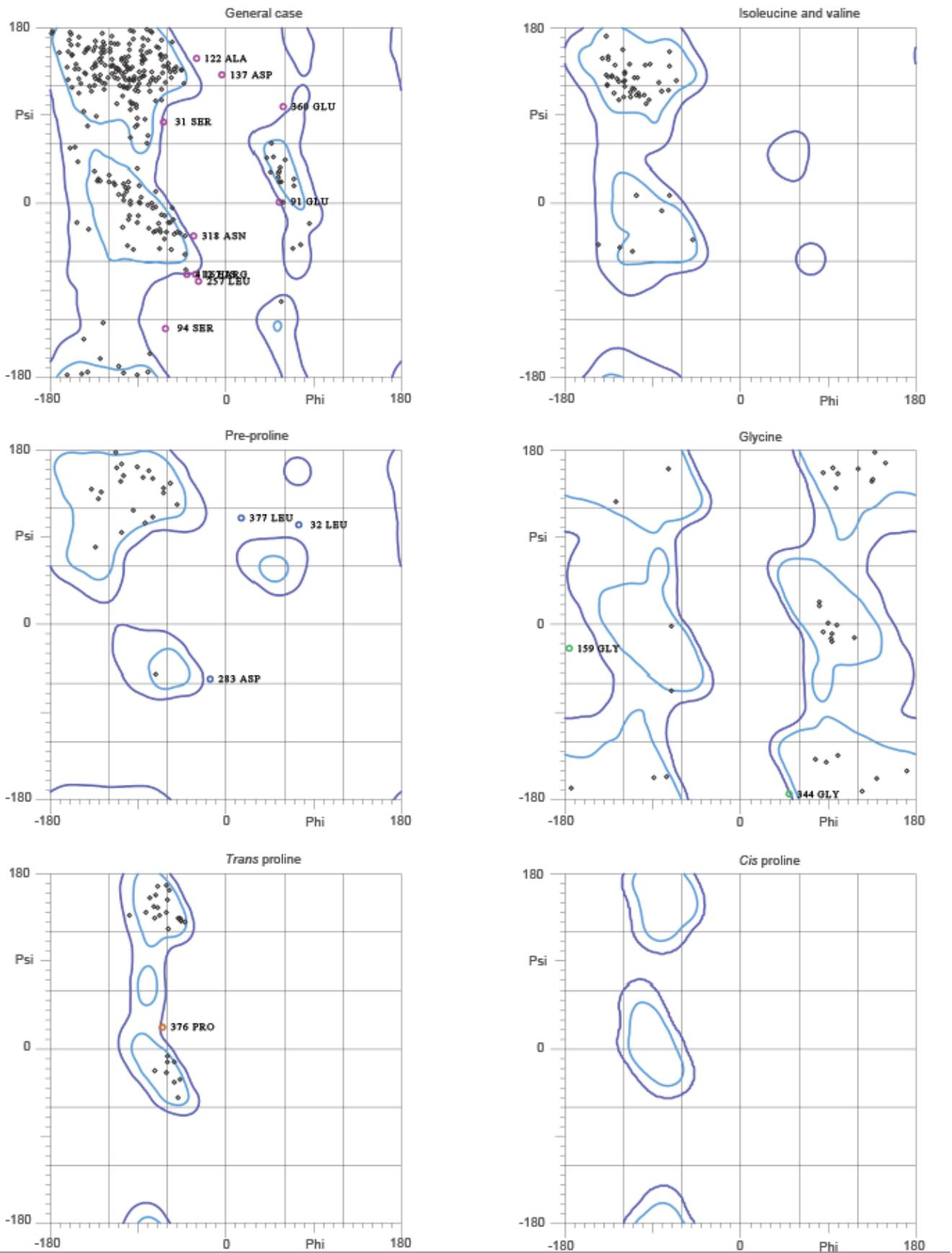
Следующие остатки, присутствующие в качестве маргинальных и в выдаче MolProbity и на EDS, обращают на себя внимание:

- SER31 – лежит на границе внешней области значений угла  $\phi$
- LEU32 – пред-пролиновый остаток, оба угла которого значительно отклоняются от пиков их распределений, при этом имеет плохую геометрию  $C\alpha$ -атома и перекрывается с орбиталью ASN30
- GLU91 – Лежит на границе внешней области для угла  $\phi$ , при этом имеет плохую геометрию  $C\alpha$ -атома и перекрывается с ASP37
- SER94 – сильно отклоняется по углу  $\phi$ , также имеет плохое значение угла  $\chi$  и немного отклоняется по положению  $C\alpha$ -атома
- ASP137 – сильно отклоняется по углу  $\phi$ , немного отклоняется по положению  $C\alpha$ -атома, а также имеет очень нехарактерный угол N-CA-C ( $4.1\sigma$ )

- TRP375 – немного отклоняется по углу  $\phi$ , перекрывается с LEU377
- PRO376 – транс-пролин, немного отклоняется по углу  $\phi$ , также перекрывается с TRP375 и имеет плохую геометрию  $C\alpha$ -атома
- LEU377 -- – пред-пролиновый остаток, оба угла которого значительно отклоняются от пиков их распределений, перекрывается с TRP375, имеет плохую геометрию  $C\alpha$ -атома.

Можно заметить, что на последовательности эти остатки имеют тенденцию группироваться – позиции 31-32, 375-377 и в какой-то степени 91-94.

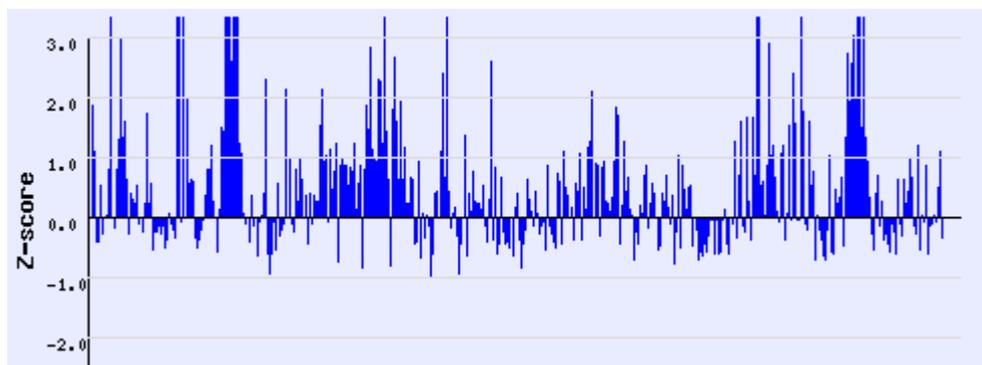
Также, при изучении выдачи MolProbity, становится видно, что большинство остатков, маргинальных по одному параметру, маргинальны и по какому-нибудь другому.



**Рисунок 4.** Карты Рамачандрана с сервиса MolProbity. Светло-синим показаны регионы, содержащие 98% остатков, а темно-синим – 99.8%

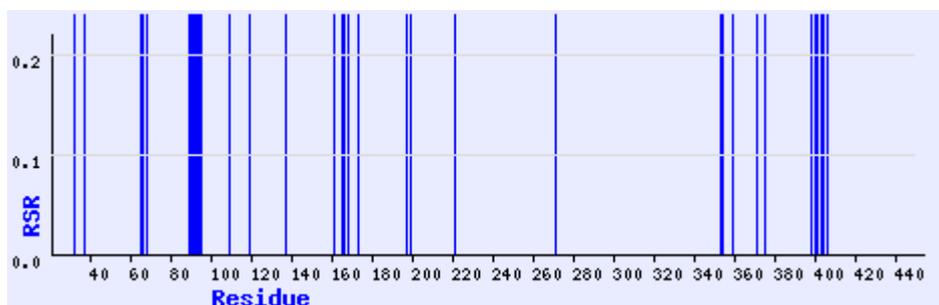
## Маргинальные остатки по RSR

RSR (real-space R) – параметр, характеризующий качество вписания модели в “экспериментальную” электронную плотность. Можно рассматривать RSR отдельных остатков, и чаще всего работают не с RSR непосредственно, а с Z-score, которые вычисляются по всем структурам с похожим разрешением. На сервисе EDS доступны как график Z-score по последовательности (рис. 5), так и отдельный график остатков, Z-score которых превышает порог в  $2\sigma$  (рис. 6).



**Рисунок 5.** Значение Z-score по последовательности.  $Z\text{-score} = (RSR - \langle RSR \rangle) / \sigma$ , где  $\langle RSR \rangle$  считается по всем структурам с похожим разрешением, а  $\sigma$  – стандартное отклонение массива этих разрешений. Взято из [16].

Для поиска маргинальных остатков, впрочем, удобнее представление данных по типу того которое изображено на рисунке 6.



**Рисунок 6.** Положения остатков, чей Z-score превосходит 2. Количество таких остатков на данной структуре составляет 8%. Взято из [17]

В принципе, здесь прослеживается та же закономерность, что и на картах Рамачандрана: маргинальные остатки имеют тенденцию располагаться рядом на последовательности. Также, среди остатков с плохим RSR, находятся практически все маргиналы по карте Рамачандрана. Так, например, остатки SER94 и GLU91 являются частью самой большой группы с плохим RSR: LYS89-LYS95.

## Маргинальные остатки по ротамерам

Ротамеры – это характерные положения боковых цепей, численно выражаемые в совокупности углов. В выдаче программы MolProbity есть данные по ротамерам, и она нашла 24 ротамера, значительно отклоняющихся от своих характерных значений.

Рассмотрим три остатка, маргинальных по ротамерам.

- LYS95 – нехарактерный угол  $\chi$ , нехарактерная геометрия  $C\alpha$ -атома, а также плохой RSR (см. предыдущий пункт).
- THR115 – нехарактерный угол  $\chi^+$  перекрытие с ASN30
- GLU360 – нехарактерный угол  $\chi$  + нехарактерная геометрия  $C\alpha$ -атома

Глядя на данные, можно сказать, что плохие ротамеры достаточно сильно ассоциированы с нехарактерной геометрией C $\alpha$ -атома.

### **Маргинальные остатки по положениям C $\alpha$ -атомов**

На сервисе MolProbity качество положения C $\alpha$ -атомов оценивается как внутренней программой, так и программой CaBLAM [18]. Внутренняя программа нашла 4 нестандартных геометрии C $\alpha$ -атомов, а CaBLAM – 16.

Рассмотрим 7 примеров:

- LEU32 – уже упоминался как маргинальный на карте Рамачандрана
- PRO126 – маргинал по CaBLAM, также сильно перекрывается с MET118, а также имеет очень нехарактерный угол N-CA-C (4.9 $\sigma$ )
- ALA186 – геометрия сильно отклоняется по данным внутренней программы (0.375%)
- CYS287 – геометрия отклоняется по данным внутренней программы (0.04%)
- TYR307 -- геометрия сильно отклоняется по данным внутренней программы (0.375%) + нехарактерный угол N-CA-C (4.9 $\sigma$ )
- PRO376 – уже упоминался как маргинальный на карте Рамачандрана
- PRO377 – уже упоминался как маргинальный на карте Рамачандрана

Интересно, что остатки ALA186 и CYS287 маргинальны только по этому параметру.

### **Анализ структуры с разрешенными перегруппировками остатков**

#### **Asn/Gln/His**

Сервис MolProbity предлагает 11 инверсий остатков ASN/GLN/HIS:

- 97 ASN
- 138 ASN
- 163 HIS
- 221 ASN
- 234 HIS
- 297 GLN
- 319 ASN
- 380 ASN
- 399 ASN
- 438 ASN

Однако только для остатков ASN319 и ASN380 говорится, что инверсия необходима. Для остальных инвертированное положение лишь немного лучше исходного.

Остатки ASN319 и ASN380 в дальнейшем рассматриваться не будут, так как они маргинальны фактически только по этому параметру. Хотя следует упомянуть, что у остатка ASN380 наблюдается очень слабое перекрытие с SER405.

Рекомендуемые инверсии были сделаны, и на полученном файле еще раз были запущены упоминавшиеся ранее программы пакета MolProbity. Полученная статистика показана в таблице 3.

**Таблица 3.** Общая характеристика выдачи программы MolProbity по PDB-файлу, в который были внесены рекомендуемые инверсии.

All-Atom Contacts	Clashscore, all atoms:	18.37	95 <sup>th</sup> percentile* (N=97, 2.90Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	26	7.10%	Goal: <0.3%
	Favored rotamers	310	84.70%	Goal: >98%
	Ramachandran outliers	16	3.80%	Goal: <0.05%
	Ramachandran favored	372	88.36%	Goal: >98%
	MolProbity score^	3.00	70 <sup>th</sup> percentile* (N=3760, 2.90Å ± 0.25Å)	
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 3476	0.00%	Goal: 0%
	Bad angles:	7 / 4728	0.15%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 29	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	16	3.82%	Goal: <1.0%
	CA Geometry outliers	4	0.95%	Goal: <0.5%

Сравнивая таблицу 3 с таблицей 2, можно заметить, что изменилось только два параметра: параметр, характеризующий невыгодное перекрывание орбиталей и количество плохих ротамеров. При этом, если после инверсий орбитали стали немного меньше перекрываться, то количество плохих ротамеров возросло на 2.

В таблице 4 собрана информация по остаткам, маргинальным по рассмотренным параметрам.

**Таблица 4.** Характеристика выбранных 15-ти маргинальных остатков. Количество плюсов указывает на вырожденность того или иного отклонения. Жирным отмечены остатки, подробно рассмотренные далее.

Остаток	Карты Рамачандрана	Геометрия Cα-атомов	Плохие ротамеры	Плохие значения RSR	Дополнительно
SER31	+	+			Sterical clash
LEU32	+	+		++	Sterical clash
<b>GLU91</b>	+	+		+++	<b>Sterical clash</b>
<b>SER94</b>	++	+	+	++	
LYS95		+	+	++	Sterical clash
THR115			+		Sterical clash
<b>PRO126</b>		+			<b>Strong sterical clash, bad N-CA-C angle</b>
ASP137	++			+	
ALA186		+++			
CYS287		++			
TYR307		+++			Weak sterical clash
<b>GLU360</b>		+	+		
TRP375	+			+++	Sterical clash
<b>PRO376</b>	+	++			<b>Sterical clash</b>

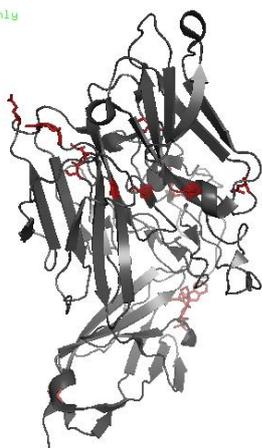
## Расположение маргинальных остатков на структуре

Ранее было сделано наблюдение, что маргинальные остатки имеют некоторую тенденцию группироваться. Теперь можно проверить две гипотезы, которые из этого вытекают:

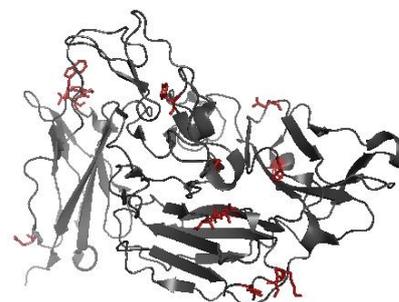
1. Возможно, существует плохо расшифрованная пространственная область в молекуле белка, и там группируются маргинальные остатки
2. Маргинальные остатки приурочены к какому-то структурному элементу, например-к неструктурированным петлям.

На рисунке 7 изображена структура 2GY5 с помеченными маргинальными остатками из таблицы 4.

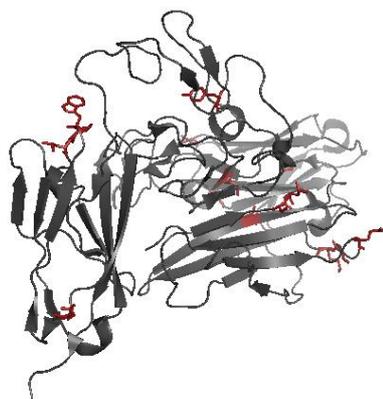
For Educational Use Only



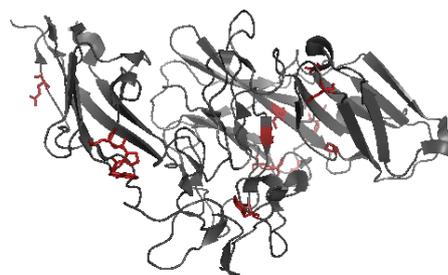
For Educational Use Only



For Educational Use Only



For Educational Use Only

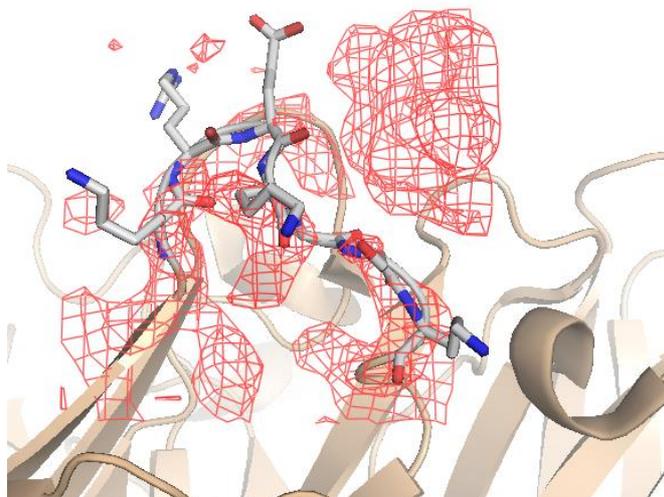


**Рисунок 7.** Маргинальные остатки в структуре 2GY5. Маргинальные остатки из таблицы 4 помечены красным и у них изображены боковые группы.

Видно, что вторая гипотеза частично подтверждается – маргинальных остатков действительно больше в петлях. Но есть они и в структурированных участках – в бета-тяжах, но располагаются в районах изломов  $\beta$ -складок. Как уже было отмечено, Tie2 связывает лиганд  $\beta$ -листом сложной формы, в котором много изломов. Поэтому эти остатки могут и не быть дефектами структуры.

Теперь рассмотрим две области, в которых наблюдается больше всего маргинальных остатков – остатки 89-95 и остатки 375-377.

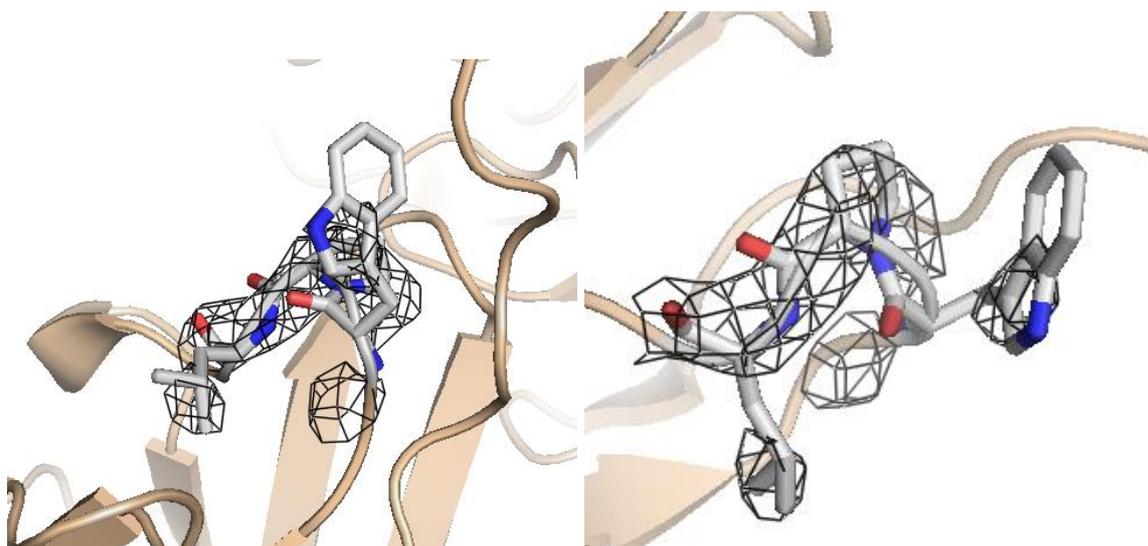
Остатки 89-95, для каждого из которых наблюдаются очень плохие значения RSR, принадлежат петле, соединяющей два бета-тяжа (рис. 8).



**Рисунок 8.** Изображение электронной плотности участка 89-95 на уровне подрезки  $1\sigma$ . Видно, что электронная плотность не охватывает половину боковых групп и прерывается по ходу цепи. Также видна большая часть электронной плотности, которая не соответствует ничему.

Видно, что структура вписана в электронную плотность достаточно плохо, а само качество электронной плотности в этом участке оставляет желать лучшего. На уровне подрезки  $2\sigma$  не видно ничего. На уровне же  $1\sigma$  наблюдается уже некоторое количество «мусора», при этом неохваченными остаются многие боковые цепи.

Далее рассмотрим область 375-377, каждая аминокислота в которой лежит в маловероятной области карты Рамачандрана и испытывает стерические затруднения. Электронная плотность этого участка показана на рисунке 9. Заметим, что эта область соответствует резкому повороту цепи между двумя  $\beta$ -листами, а остатки вписаны в электронную плотность достаточно хорошо. На уровне подрезки  $2\sigma$  хорошо видна цепь и некоторые боковые атомы. На уровне  $1\sigma$  петля полностью видна. Но об этом пойдет речь при рассмотрении остатка PRO376.

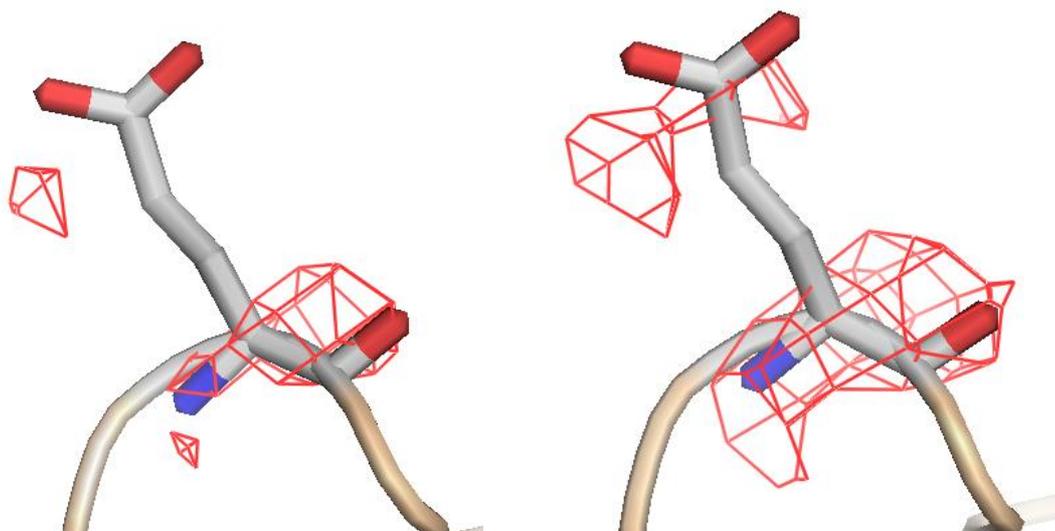


**Рисунок 9.** Электронная плотность участка 375-377 на уровне подрезки  $2\sigma$

## Более глубокий анализ пяти маргинальных остатков

### Анализ электронной плотности остатка GLU91

На рисунке 10 показан глутамат в положении 91 при уровнях подрезки  $1\sigma$  и  $0.75\sigma$ . Видно, что боковая цепь совершенно не охватывается функцией электронной плотности и поэтому невозможно нормально вписать в нее остаток. Отсюда неправильные значения углов на карте Рамачандрана, стерические затруднения и высокий RSR. Также следует отметить, что этот остаток располагается на вершине петли, и, возможно поэтому его положение в кристалле не было строго фиксировано.



**Рисунок 10.** Электронная плотность остатка GLU91 при уровнях подрезки:  $1\sigma$  (слева) и  $0.75\sigma$  (справа).

#### Значения показателей:

Перекрытие с LYS92: 0.51 Å.

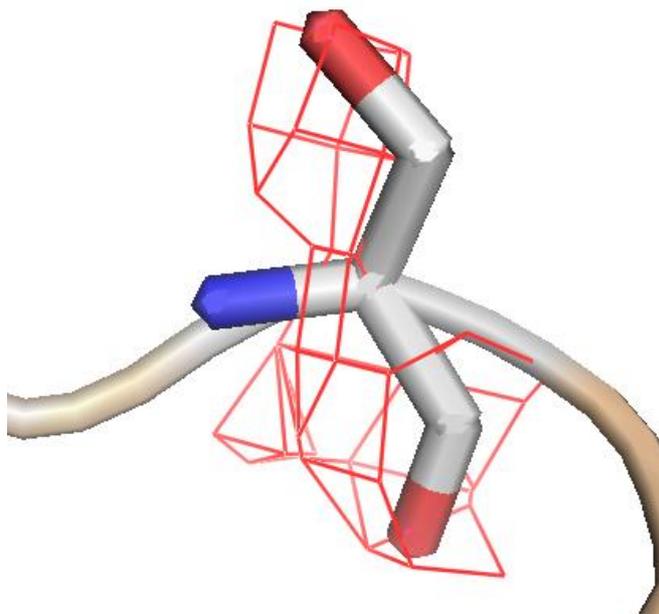
На карте Рамачандрана лежит в области: 0.04%

Перцентиль характерной геометрии  $C\alpha$ -атома (CaBLAM-score): 1.463%

Zscore(RSR): 7.27

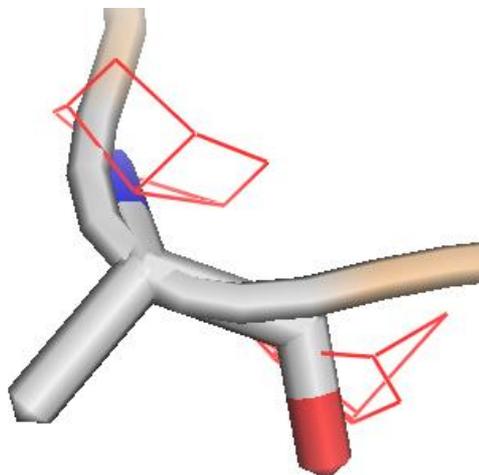
### Анализ электронной плотности остатка CYS94

У этого остатка (см. табл. 4) очень много «минусов» с точки зрения валидации: он располагается вдали от границ предпочтительных областей карт Рамачандрана, у него нехарактерная геометрия  $C\alpha$ -атомов, плохие ротамеры и высокий RSR. Лежит он в той же петле, что и предыдущий рассмотренный остаток. Электронная плотность CYS94 показана на рисунке 11.



**Рисунок 11.** Электронная плотность остатка CYS94 на уровне подрезки  $1\sigma$ .

Видно, что электронная плотность покрывает остаток лучше, чем в предыдущем случае, но все же плохо: Cβ атом выдается из «рукава» электронной плотности, да и вообще остаток вписан не лучшим образом. Возможным объяснением может быть тот факт, что рядом лежащий аланин (рис. 12) практически не окружен электронной плотностью. В результате неправильно был определен ход цепи и положение цистеина.



**Рисунок 12.** Электронная плотность 93-го аланина на уровне подрезки  $0.75\sigma$ . Видно, что качество электронной плотности не позволяет достоверно вписать остаток в нее.

**Значения показателей:**

На карте Рамачандрана лежит в области: 0.0%

Перцентиль характерной геометрии Cα-атома (CaBLAM-score): 2.125%

Перцентиль характерной геометрии угла  $\chi_1$ : 1.8%

Zscore(RSR): 4.51

### Анализ электронной плотности остатка PRO126

Согласно сервису MolProbity, пролин 126 имеет сильные стерические затруднения и нехарактерное значение угла N-CA-C. Рассмотрим его электронную плотность (рис. 13).

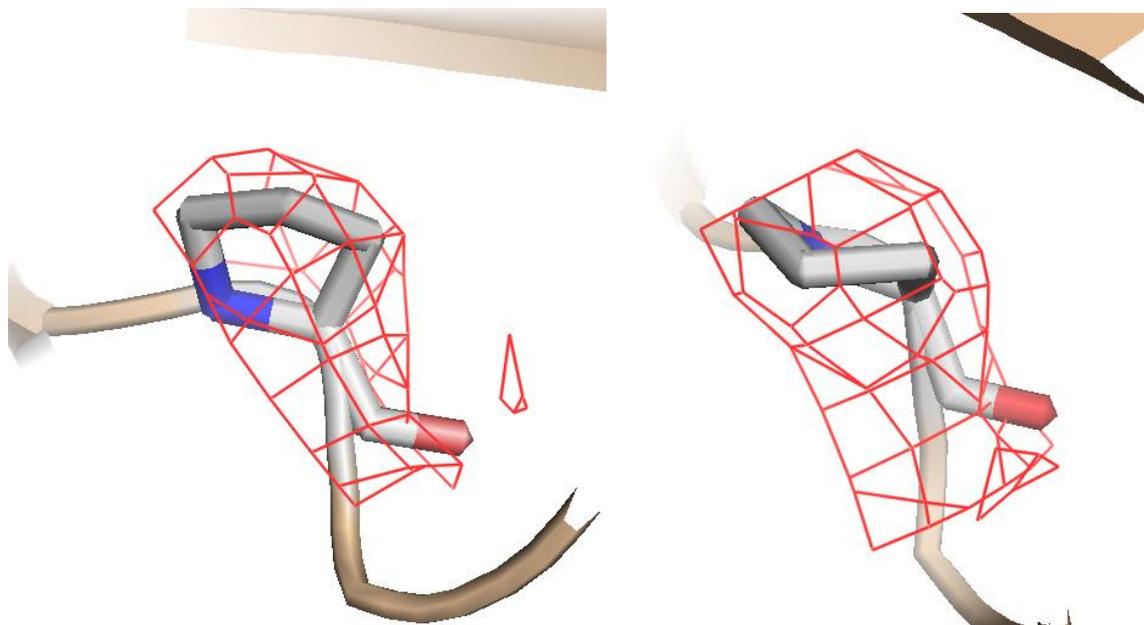


Рисунок 13. Электронная плотность PRO126 на уровне подрезки  $1.5\sigma$ .

Видно, что в принципе остаток вписан в электронную плотность хорошо, но есть возможность немного развернуть, так, чтобы атом кислорода не «выпадал» из электронной плотности. Возможно в этом случае изменятся углы и остаток перестанет быть маргинальным. Вообще, положение остатка в петле дает кристаллографам большую свободу, и углы можно сделать более правдоподобными.

#### Значения показателей:

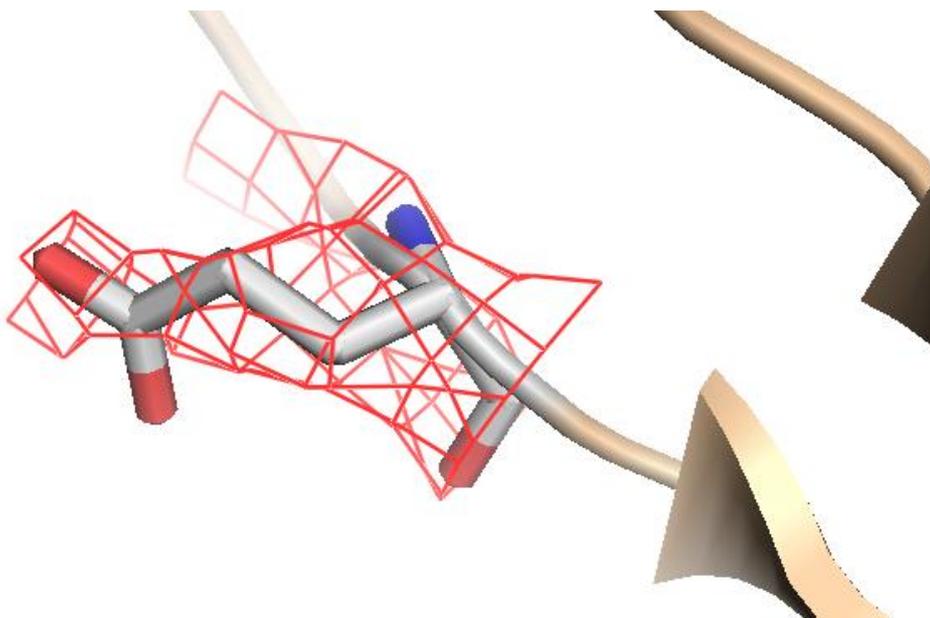
Перекрытие с MET118:  $0.95 \text{ \AA}$

Перцентиль характерной геометрии  $C\alpha$ -атома (CaBLAM-score): 0.4%

Z-score характерной геометрии угла N-CA-C:  $4.6\sigma$

### Анализ электронной плотности остатка GLU360

Остаток GLU360 попал в список маргинальных из-за плохих ротамеров и нехарактерной геометрии  $C\alpha$ -атома. Как и предыдущий, он лежит в петле, и, опять же, как и предыдущий, он окружен электронной плотностью достаточно хорошо, но все же недостаточно хорошо, чтобы точно определить положения всех атомов. Как и в предыдущем случае, остаток можно привести к более выгодному положению, его «пошевелив» в электронной плотности.



**Рисунок 14.** Электронная плотность остатка GLU360 на уровне подрезки  $1.5\sigma$

**Значения показателей:**

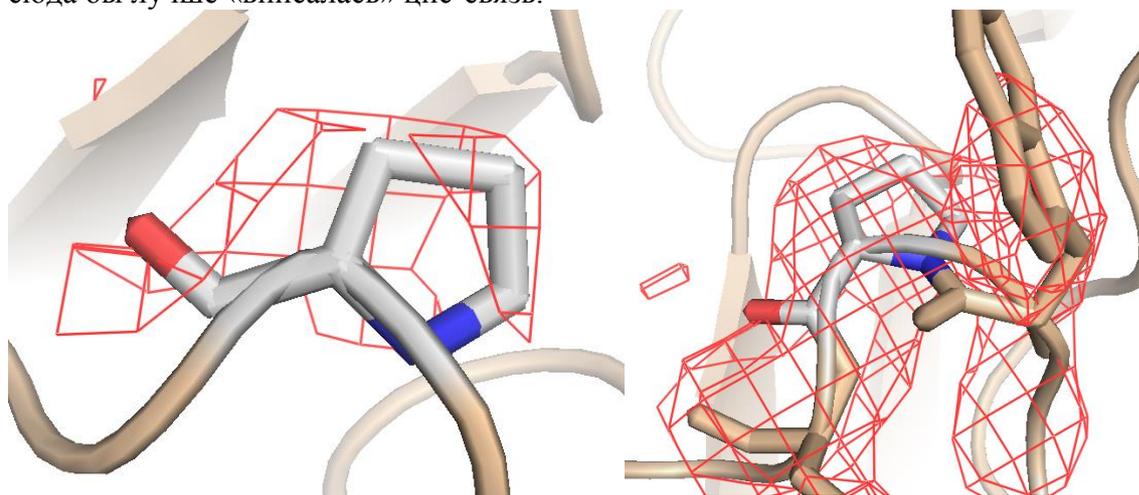
Перцентиль характерной геометрии  $C\alpha$ -атома (CaBLAM-score): 0.177%

Перцентиль характерной геометрии угла  $\chi$ : 0.0%

На карте Рамачандрана лежит в области: 0.2%

**Анализ электронной плотности остатка PRO376**

376-й пролин лежит на вершине уже рассмотренной петли 375-377. Эта петля делает резкий поворот как раз на уровне этого остатка (рис. 15). При этом сама петля определена достаточно хорошо, но в нее плохо вписан участок между TRP375 и PRO376 – предсказано, что это – транс-пептидная связь. Однако, как мне кажется, сюда бы лучше «вписалась» цис-связь.



**Рисунок 15.** Электронная плотность остатка PRO376 на уровне подрезки  $1.5\sigma$  (слева) и электронная плотность петли 375-377 (справа).

**Значения показателей:**

Перцентиль характерной геометрии  $C\alpha$ -атома (CA GEOM-score): 0.122%

На карте Рамачандрана лежит в области: 0.08%

Перекрывание с TRP375: 0.62 Å

## Сравнение с моделью из PDB\_REDO

В 2013 году с помощью программы WHAT\_CHECK была проведена валидация структуры 2GY5. Получившаяся в результате структура лежит в PDB\_REDO вместе с отчетом о валидации, там же ее можно и скачать.

По итогам построения структуры заново, можно сказать, что Rfree и Rwork практически не изменились, а разность между ними даже немного возросла. Также не удалось значительно уменьшить количество остатков в неблагоприятных областях карты Рамачандрана, количество необычных ротамеров и количество невыгодных орбитальных перекрытий. При этом Z-score длин связей и величин углов даже немного вырос.

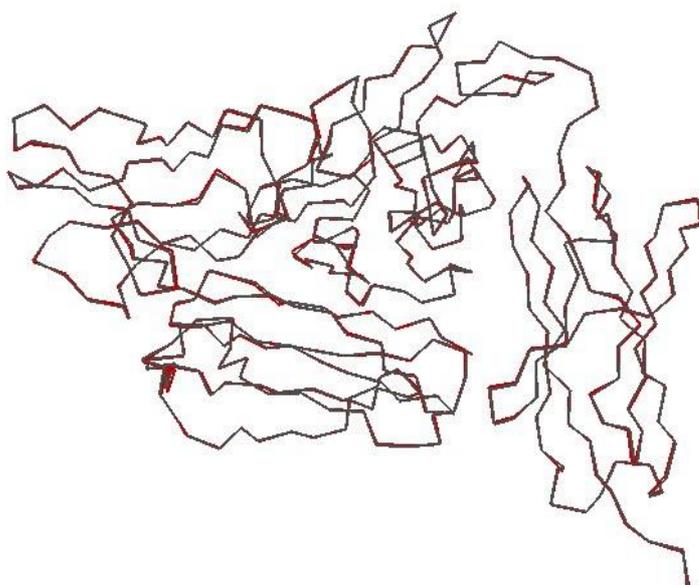
Впрочем, валидация оказалась успешной в смысле RSR и RSCC. Для 34 остатков RSCC стал выше и лишь для одного – ниже, и для 9 остатков RSR стал ниже, и ни для одного не повысился.

Небольшое отступление: RSCC (real space correlation coefficient) – это мера похожести электронной плотности в модели и электронной плотности, получаемой в эксперименте. Он считается как:

$$RSCC = \frac{\sum |\rho_{obs} - \langle \rho_{obs} \rangle| * \sum |\rho_{calc} - \langle \rho_{calc} \rangle|}{(\sum |\rho_{obs} - \langle \rho_{obs} \rangle|^2 * \sum |\rho_{calc} - \langle \rho_{calc} \rangle|^2)^{0.5}}$$

Больше параметров приведено в протоколе валидации [19]

Для начала, можно сравнить две структуры, просто наложив их. Для наложения структур я использовал сервис FATCAT, осуществляющий гибкое выравнивание последовательностей [7]. Структуры очень похожи, но не идентичны – RMSD составило 0.21 Å, а на наложении видно, что некоторые Cα-атомы все же отстоят в разных структурах (рис. 16).



**Рисунок 16.** Наложение структуры 2GY5 (красная) и соответствующей структуры из PDB\_REDO (серая).

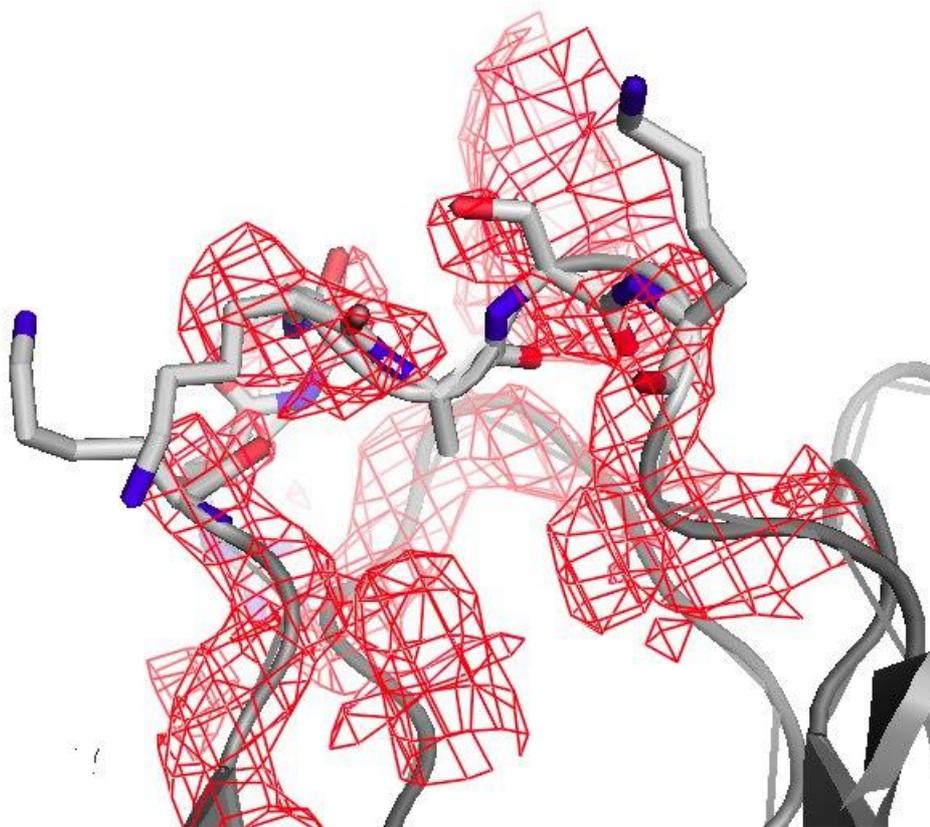
Далее воспользуемся сервисом MolProbity для оценки качества структуры из PDB\_REDO. Результаты получаются неоднозначные (таблица 5).

**Таблица 5.** Общая характеристика выдачи программы MolProbity по файлу из PDB\_REDO

All-Atom Contacts	Clashscore, all atoms:	10.38	97 <sup>th</sup> percentile* (N=141, 2.80Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	28	7.63%	Goal: <0.3%
	Favored rotamers	310	84.47%	Goal: >98%
	Ramachandran outliers	9	2.14%	Goal: <0.05%
	Ramachandran favored	388	92.16%	Goal: >98%
	MolProbity score <sup>^</sup>	2.69	84 <sup>th</sup> percentile* (N=4482, 2.80Å ± 0.25Å)	
	Cβ deviations >0.25Å	2	0.51%	Goal: 0
	Bad bonds:	0 / 3477	0.00%	Goal: 0%
	Bad angles:	9 / 4729	0.19%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 29	0.00%	Expected: ≤1 per chain, or ≤5%
	Twisted Peptides:	1 / 422	0.24%	Goal: 0
Low-resolution Criteria	CaBLAM outliers	19	4.53%	Goal: <1.0%
	CA Geometry outliers	3	0.72%	Goal: <0.5%

С одной стороны, уменьшилось почти вдвое количество маргиналов по картам Рамачандрана, уменьшилось число невыгодных орбитальных перекрытий и улучшился параметр MolProbity score (теперь он в 84-й перцентили, вместо 70-й). С другой стороны, увеличилось число плохих ротамеров, нехарактерных углов и Ca-атомов с нехарактерной геометрией.

Наконец, посмотрим на самую плохо разрешенную часть в исходной структуре – петлю 89-95 – в структуре из PDB\_REDO (рис. 17).



**Рисунок 17.** Петля 89-95 структуры из PDB\_REDO, вписанная в электронную плотность при уровне подрезки  $1\sigma$ .

Здесь, как можно заметить, ситуация все-таки несколько лучше, чем на рисунке 8. Основная цепь лучше вписывается, пептидные связи лежат внутри нее. Но все равно, многие боковые группы не покрыты электронной плотностью.

## Выводы

В результате проделанной работы можно сказать, что в целом данная структура разрешена хорошо. В ней нет явно нелогичных деталей (как, например, заряда в гидрофобном ядре), все домены и элементы вторичной структуры определены правильно и с хорошим разрешением. Все минусы структуры относятся к более низкому уровню организации – к конформации аминокислот, которая, к сожалению, во многих случаях определена плохо.

Плохое определение конформаций аминокислот вытекает, скорее всего, из не очень качественного построения электронной плотности. Действительно, авторы отбросили несколько тысяч рефлексов, нигде не указав, по каким именно рефлексам строилась их модель. Когда публиковалась структура в PDB\_REDO, ее автор построил модель по всем рефлексам, получив значения  $R_{work}$  и  $R_{free}$  сильно отличающиеся от тех, что написаны в статье. Больше была и их разность. Скорее всего, авторы построили модель по всем рефлексам, а затем выбросили часть, которая ее ухудшает. Впоследствии авторы статьи применили молекулярную динамику для улучшения уже построенной структуры.

Судя по тому, что некоторые части структуры имеют невысокое качество как в улучшенной модели, так и в модели PDB\_REDO, можно предположить, что эта структура в принципе сложна для построения модели и что в ней действительно есть (и, возможно, много) аминокислоты с нестандартными значениями различных структурных параметров. На это указывают наличие петель между  $\beta$ -листами, которые резко поворачиваются больше, чем на 180 градусов, изломы  $\beta$ -структуры и нехарактерные углы поворота  $\beta$ -листов. Все это довольно необычно для белков и, в принципе, встречается достаточно редко [6].

Следует добавить, что структура из PDB\_REDO оказалась лучше исходной фактически только в смысле параметров RSR и RSCC, а такие локальные атрибуты структуры, как углы поворота боковых цепей (ротамеры) и углы связей в этой структуре менее достоверны.

Несмотря на изъяны в структуре, ее авторам удалось по ней предсказать аминокислоты, участвующие в связывании лиганда. Последующие эксперименты только подтвердили их предположения. Поэтому, эту структуру вполне можно использовать для исследований свойств белка Tie2.

## Список литературы

1. Markus Ramsauer and Patricia A. D'Amore. Getting Tie(2)d up in angiogenesis J. Clin. Invest. 110:1615–1617. (2002)
2. Barton WA, Tzvetkova-Robev D, Miranda EP, Kolev MV, Rajashankar KR, Himanen JP, Nikolov DB. Crystal structures of the Tie2 receptor ectodomain and the angiopoietin-2-Tie2 complex. Nat Struct Mol Biol. 2006 Jun;13(6):524-32. Epub 2006 May 28.
3. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A 47, 110–119 (1991).
4. Brunger, A.T. et al. Crystallography and NMR system (CNS): A new software system for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 54, 905–921 (1998)
5. Who checks the checkers? Four validation tools applied to eight atomic resolution structures. EU 3-D Validation Network. J Mol Biol. 1998 Feb 20;276(2):417-36.
6. Алексей Финкельштейн, Олег Птицын. Физика белка. Курс лекций с цветными и стереоскопическими иллюстрациями и задачами, 2002 год
7. Ye Y, Godzik A. FATCAT: a web server for flexible structure comparison and structure similarity searching. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W582-5.
8. Loughna, S., and Sato, T.N. 2001. Angiopoietin and Tie signaling pathways in vascular development. Matrix Biol. 20:319–325.
9. Joosten RP, Salzemann J, Bloch V, Stockinger H, Berglund AC, Blanchet C, Bongcam-Rudloff E, Combet C, Da Costa AL, Deleage G, Diarena M, Fabbretti R, Fettahi G, Flegel V, Gisel A, Kasam V, Kervinen T, Korpelainen E, Mattila K, Pagni M, Reichstadt M, Breton V, Tickle IJ, Vriend G. PDB\_REDO: automated re-refinement of X-ray structure models in the PDB. J Appl Crystallogr. 2009 Jun 1
10. Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. 2004 Jul 1;32

## **Ссылки**

11. <http://www.rcsb.org/pdb/explore/jmol.do?structureId=2GY5&opt=3>
12. <http://www.rcsb.org/pdb/explore.do?structureId=2gy5>
13. [http://www.cmbi.ru.nl/pdb\\_redo/gy/2gy5/wc/index.html](http://www.cmbi.ru.nl/pdb_redo/gy/2gy5/wc/index.html)
14. <http://molprobity.biochem.duke.edu/>
15. <https://eds.bmc.uu.se/cgi-bin/eds/rama?pdbCode=2gy5>
16. <https://eds.bmc.uu.se/cgi-bin/eds/zscore?2gy5>
17. <https://eds.bmc.uu.se/cgi-bin/eds/sign?2gy5&sco=2.0>
18. [https://www.phenix-online.org/documentation/reference/cablam\\_validation.html](https://www.phenix-online.org/documentation/reference/cablam_validation.html)
19. [http://www.cmbi.ru.nl/pdb\\_redo/gy/2gy5/index.html](http://www.cmbi.ru.nl/pdb_redo/gy/2gy5/index.html)
20. <http://www.rcsb.org/pdb/explore/materialsAndMethods.do?structureId=2GY5>
21. [http://ftp.wwpdb.org/pub/pdb/validation\\_reports/gy/2gy5/2gy5\\_full\\_validation.pdf](http://ftp.wwpdb.org/pub/pdb/validation_reports/gy/2gy5/2gy5_full_validation.pdf)
22. <http://www.rcsb.org/pdb/explore/explore.do?structureId=2gy5>