

Практикум № 6. Валидация РСА-структуры

Иосиф Финкельберг

1 Введение

В этом практикуме мы работали с PDB-структурой **lsbp**. Эта структура была получена с помощью рентгено-структурного анализа (РСА) в работе Jing J. He и Florante A. Quioco 1993-го года [1]. **lsbp** является структурой периплазматического сульфат-связывающего ABC-транспортера (**Sulphate Binding Protein, SBP**) гамма-протеобактерии *Salmonella typhimurium*. SBP прочно и специфично связывает сульфат-ионы [2]. Именно эта его особенность и заинтересовала авторов работы [1]: они использовали SBP в связанном с заряженным субстратом (ионами сульфата) состоянии, как хорошую модель для изучения электростатических взаимодействий. Авторы предложили схему, объясняющую взаимодействия субстрата с белком (см. рис 1 и 2). В непосредственном связывании субстрата (иона сульфата) задействованы 7 водородных связей. Находящиеся недалеко от связанного иона SO_4^{2-} положительно заряженные аминокислотные остатки (Arg, His) участвуют в электростатическом взаимодействии с ним. Кроме того, авторы замечают, что в сторону иона SO_4^{2-} направлены своими N-концами три альфа-спирали, что, возможно, тоже частично компенсирует отрицательный заряд субстрата (такая модель альфа-спиралей называется “macro-dipole model”).

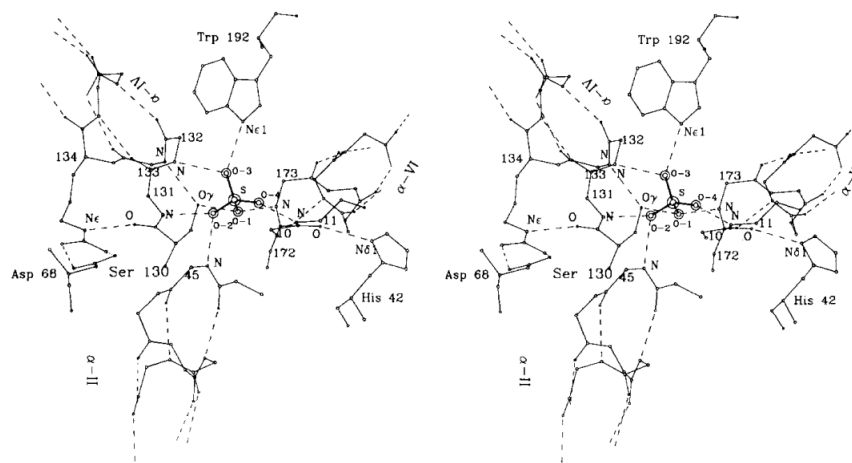


Рис. 1: Предложенная в статье [1] схема связывания SBP с субстратом (ионом SO_4^{2-}). Показаны три альфа-спирали, NH-группы остовов которых образуют водородные связи с кислородами субстрата. Спираль α -VI образует две водородные связи с атомами кислорода SO_4^{2-} . Остатки Trp-192 и Ser-130 образуют по одной водородной связи с субстратом своими боковыми цепями. Еще одну водородную связь с субстратом образует азот остова остатка Gly-131, который находится у границы альфа-спирали IV. Итого, в связывании субстрата непосредственно задействованы 7 водородных связей.

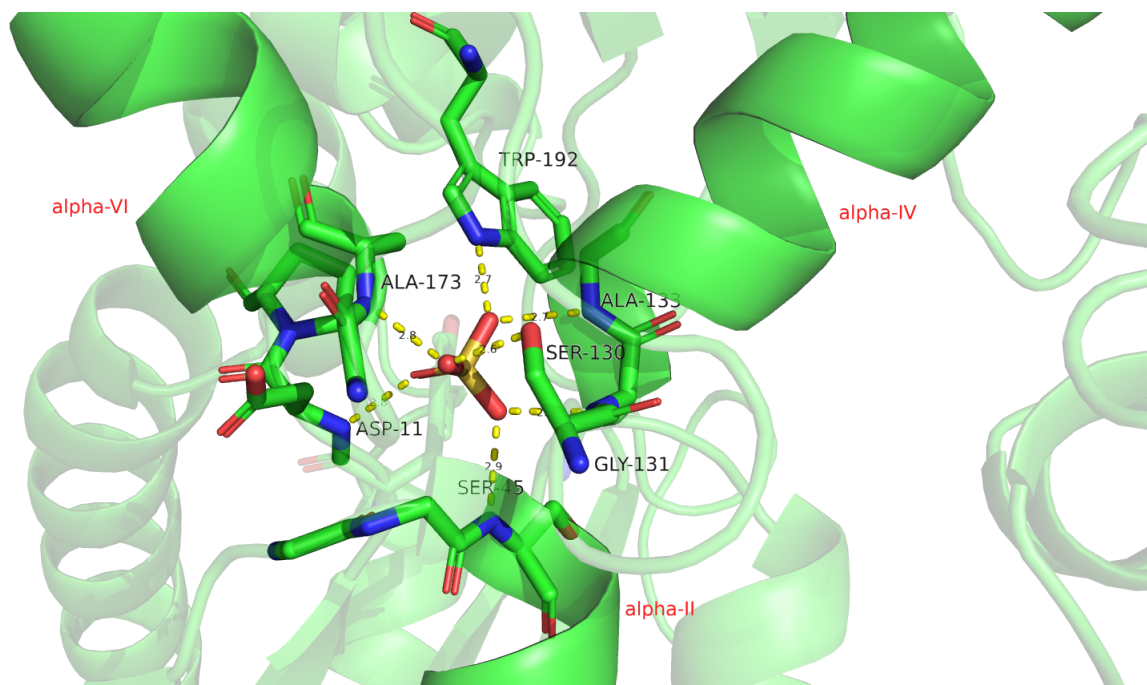


Рис. 2: Предложенная в статье [1] схема связывания SBP с субстратом (ионом SO_4^{2-}): изображение в rumpol. Видим, что ключевыми для связывания с субстратом остатками можно в первом приближении считать: Asp-11, Ser-45, Ser-130, Gly-131, Ala-133, Ala-173, Trp-192.

Для дальнейших анализов констатируем здесь, что в первом приближении для взаимодействия с субстратом важными можно считать 7 остатков: Asp-11, Ser-45, Ser-130, Gly-131, Ala-133, Ala-173, Trp-192 (именно эти остатки образуют водородные связи с субстратом непосредственно). При этом, как отмечено в статье, для связывания необходимы также и некоторые другие остатки, которые способствуют электростатическим взаимодействиям (компенсируют отрицательный заряд сульфата): в первую очередь это Arg-134 и His-42. Однако эти остатки находятся от субстрата на расстоянии порядка 7-8 Å, и можно предположить, что небольшие погрешности в определении их положения будут уже не так критичны для понимания и изучения функции белка.

2 Задание 1. Общий анализ структуры

В этом задании мы проанализировали данные о качестве структуры из **отчета о валидации**, приведенного на PDB-странице для нашего белка. К сожалению, многие данные о качестве структуры оказались недоступны. Так, в отчете об эксперименте не приведены метод решения фазовой проблемы

и число рефлексов (см. рис 3).

Refinement

Statistics									
Diffraction ID	Structure Solution Method	Resolution (High)	Resolution (Low)	Number Reflections (R-Free)	Percent Reflections (Observed)	R-Factor (Observed)	R-Work	R-Free	Mean Isotropic B
X-RAY DIFFRACTION		1.7				0.178			

Рис. 3: Скриншот **PDB-страницы** с отчетом об эксперименте. Видим, что данные о решении фазовой проблемы и числу рефлексов недоступны. Они отсутствуют также в **отчете о валидации**.

В таблице 1 мы привели важные на наш взгляд экспериментальные характеристики структуры, которые были указаны в отчете о валидации.

Таблица 1. Основные характеристики РСА-расшифровки

Группа симметрии	P 21 21 21
Параметры ячейки: a, b, c, α , β , γ	40.77Å 49.37Å 141.03Å 90.00° 90.00° 90.00°
Метод решения фазовой проблемы	Не указан
Число рефлексов	Не указано
Разрешение	9.98 — 1.70 Å
Полнота данных	89.7% для разрешения 1.70 Å
Среднее значение В-фактора по всем атомам (Å ²)	14.0

Группа симметрий указана по орторомбической сингонии (одной из кристаллографических классификаций). В данном случае мы имеем дело с примитивной (простой, P) решеткой Бравэ. Ее диаграмма приведена на рис. 4.

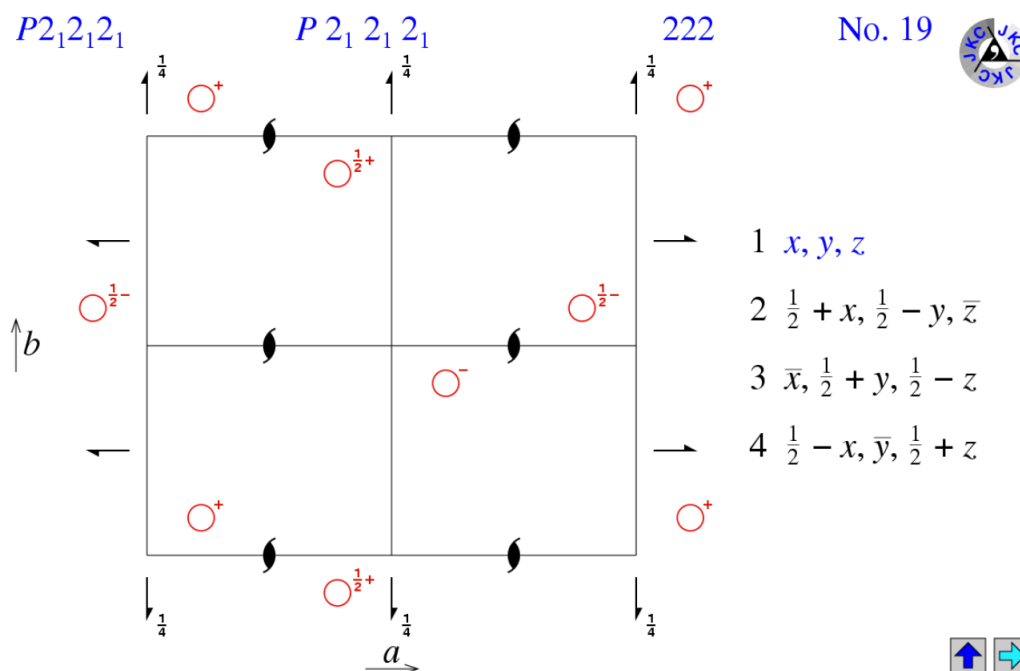


Рис. 4: Диаграмма орторомбической сингонии $P2_12_12_1$, взятая [отсюда](#). Используемые обозначения приведены [здесь](#). Справа приведены 4 оператора (включая тождественный), которые соответствуют преобразованиям симметрии.

Кристаллографическая решетка является прямоугольным параллелепипедом (см. таблицу 1). В основании лежит почти что квадрат, высота в 3.5 раза длиннее стороны основания. Разрешение структуры, указанное авторами составляет 1.7 Å. При таком разрешении полнота разрешенных гармоник составляет около 90% , что является довольно хорошим показателем.

Затем мы проанализировали метрики качества структуры в целом. Для этого мы сперва нашли значение R-фактора (значение R_{free} фактора, к сожалению, опять же, оказалось не приведено). Оно оказалось равным 0.178. Как мы узнали на семинарах, "хороший" R-фактор не должен превышать значения (разрешение / 10). В данном случае, $R\text{-factor} = 0.178 > 0.17 = \frac{1.7}{10}$. Т.е. R-фактор недостаточно мал, чтобы про него можно было сказать, что он "хороший". Он, однако, находится не слишком далеко от границы. Мы также бегло рассмотрели некоторые другие метрики, краткая сравни-

тельная информация о которых приведена в pdb-отчете о валидации (см. рис. 5).

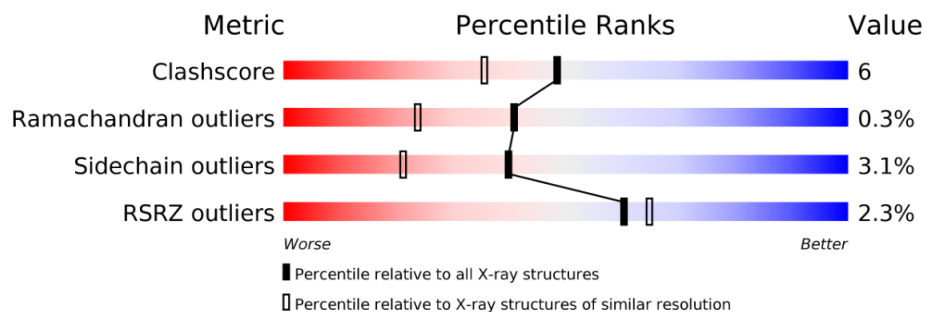


Рис. 5: Таблица с основными метриками качества 1sbr структуры, взятая из PDB-отчета о валидации. Незакрашенные прямоугольники обозначают процентиль в группе значений метрики для структур с тем же разрешением (1.7 Å). Закрашенные прямоугольники – процентиль в группе значений метрики для вообще всех полученных с помощью PCA PDB-структур.

Мы видим, что данная структура вписывается в электронную плотность (ЭП) лучше, чем половина всех PCA PDB-структур, и даже лучше, чем половина структур с разрешением 1.7 Å (см. строку с RSRZ outliers: оба прямоугольника соответствуют квантилю уровня 50-60%).

По остальным приведенным в таблице метрикам значения для нашей структуры близки к медианным значениям по всем PCA-структурам, однако довольно значительно уступают значениям соответствующих метрик для структур с таким же заявленным разрешением: 1.7 Å. Можно предположить, что большинство структур с таким хорошим разрешением были получены в течение последнего десятилетия. Соответственно, неудивительно, что они будут превосходить по различным показателям структуру 1sbr, полученную около тридцати лет назад. Подробнее мы разберем недостатки нашей структуры в следующем разделе. Сейчас же напоследок визуально рассмотрим нашу структуру и ее ЭП в ruro1 (см. рис. 6).

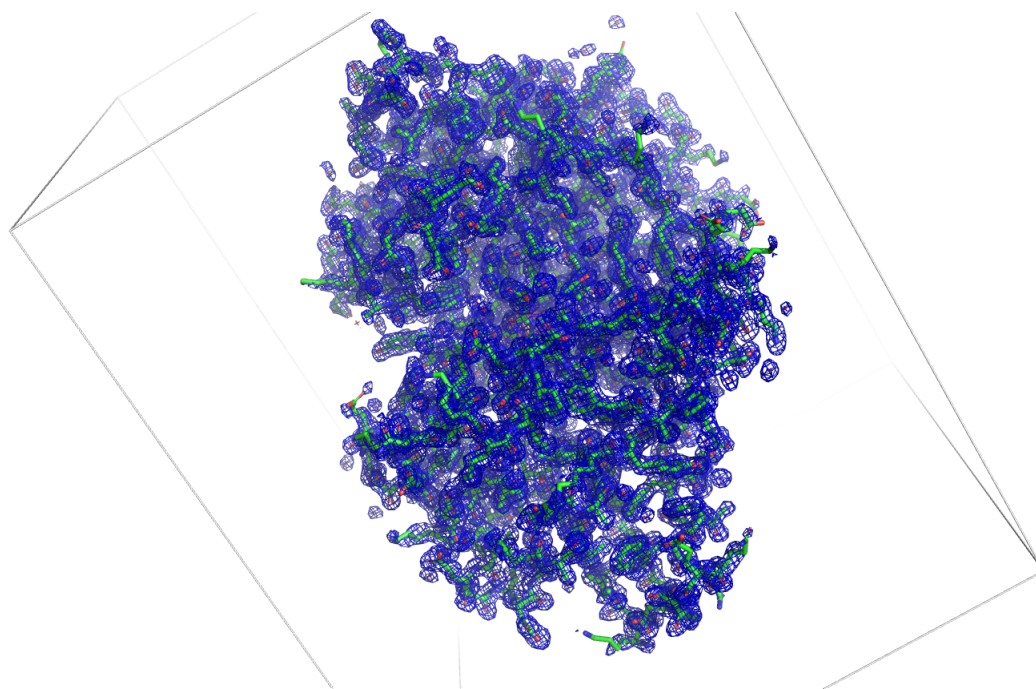


Рис. 6: Структура 1sbr с наложенной картой электронной плотности (выбран уровень подрезки $\sigma = 1$).

Мы видим, что ЭП подтверждает почти что все участки структуры. Она отсутствует только вокруг боковых цепей 3-6 аминокислотных остатков.

3 Валидация структуры с помощью сторонних сервисов.

Для валидации нашей структуры мы воспользовались сервисом [MolProbity](#) (формально можно было бы не проводить самим анализ структуры с помощью этого сервиса, так как результаты итак уже даны в PDB-отчете о валидации). Мы воспользовались методом Analyze geometry without all-atom contacts (при этом мы использовали настройки по умолчанию), реализованном на этом сервисе. Сперва мы сделали это, не добавляя предварительно водороды.

Protein Geometry	Poor rotamers	13	5.10%	Goal: <0.3%
	Favored rotamers	223	87.45%	Goal: >98%
	Ramachandran outliers	1	0.33%	Goal: <0.05%
	Ramachandran favored	302	98.37%	Goal: >98%
	Rama distribution Z-score	-0.96 ± 0.45		Goal: abs(Z score) < 2
	Cβ deviations >0.25Å	10	3.47%	Goal: 0
	Bad bonds:	4 / 2486	0.16%	Goal: 0%
	Bad angles:	66 / 3376	1.95%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 12	0.00%	Expected: ≤ 1 per chain, or ≤ 5%
Additional validations	Tetrahedral geometry outliers	2		

Рис. 7: Выдача сервиса MolProbity (Analyze geometry without all-atom contacts) для структуры 1sbp.

Мы видим, что некоторые из основных метрик качества достаточно сильно отклоняются от целевых значений: это число маргинальных и “хороших” ротамеров, а также углы и длины межатомных связей.

Затем мы добавили водороды в MolProbity (Add Hydrogens) и посмотрели на возможные инверсии остатков глутамина, аспарагина, гистидина.

Chain	Res#	Alt	Res ID	Orig	Flip	Flip-Orig
A	4		GLN	-4.5	-0.54	3.96
A	35		ASN	-5.1	-0.86	4.24
A	48		GLN	-0.67	0.13	0.8
A	110		HIS	-0.65	0.87	1.52
A	147		HIS	-4.7	0.54	5.24
A	165		ASN	-1.8	1.3	3.1
A	258		ASN	-3.2	2	5.2
A	306		GLN	-0.85	0.45	1.3

Рис. 8: Выдача сервиса MolProbity (Add Hydrogens) для структуры 1sbp.

Итак, оказывается, что около восьми остатков, вероятно, вписаны в ЭП неверно (с инверсией). Мы приняли изменения, которые предлагал нам MolProbity, и посмотрели, как изменились метрики качества (см. рис 9).

All-Atom Contacts	Clashscore, all atoms:	4.79		95 th percentile* (N=819, 1.70Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	13	5.10%	Goal: <0.3%
	Favored rotamers	224	87.84%	Goal: >98%
	Ramachandran outliers	1	0.33%	Goal: <0.05%
	Ramachandran favored	302	98.37%	Goal: >98%
	Rama distribution Z-score	-0.96 ± 0.45		Goal: abs(Z score) < 2
	MolProbity score [^]	1.79		73 rd percentile* (N=9248, 1.70Å ± 0.25Å)
	Cβ deviations >0.25Å	10	3.47%	Goal: 0
	Bad bonds:	4 / 2486	0.16%	Goal: 0%
	Bad angles:	66 / 3376	1.95%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 12	0.00%	Expected: ≤ 1 per chain, or ≤ 5%
Additional validations	Tetrahedral geometry outliers	2		
	Waters with clashes	6/138	4.35%	See UnDowser table for details

Рис. 9: Выдача сервиса MolProbity (Add Hydrogens) для структуры 1sbr: посчитаны все метрики для структуры после добавления водородов.

Видим, что фактически все метрики остались прежними. Единственное, после добавления водородов появились Water clashes, которых прежде MolProbity не замечал.

Интересно, что маргиналы согласно MolProbity нашлись и среди функционально важных аминокислотных остатков: Asp-11 (poor rotamer), His 42 (bond angles), Trp-192 (Cβ-deviation). Впрочем отклонения от нормы для этих остатков достаточно небольшие, и, кроме того, точное положение остатка His 42 не является ключевым для связывания субстрата (см. выше). Кроме остатков Asp-11, Trp-192 субстрат связывают еще 5 других остатков, так что, можно предположить, что общая картина связывания субстрата остается валидной несмотря на возможные обнаруженные неточности.

Мы, впрочем, решили проверить это утверждение, посмотрев, насколько хорошо ЭП структуры описывает функционально важные остатки (см. рис. 10).

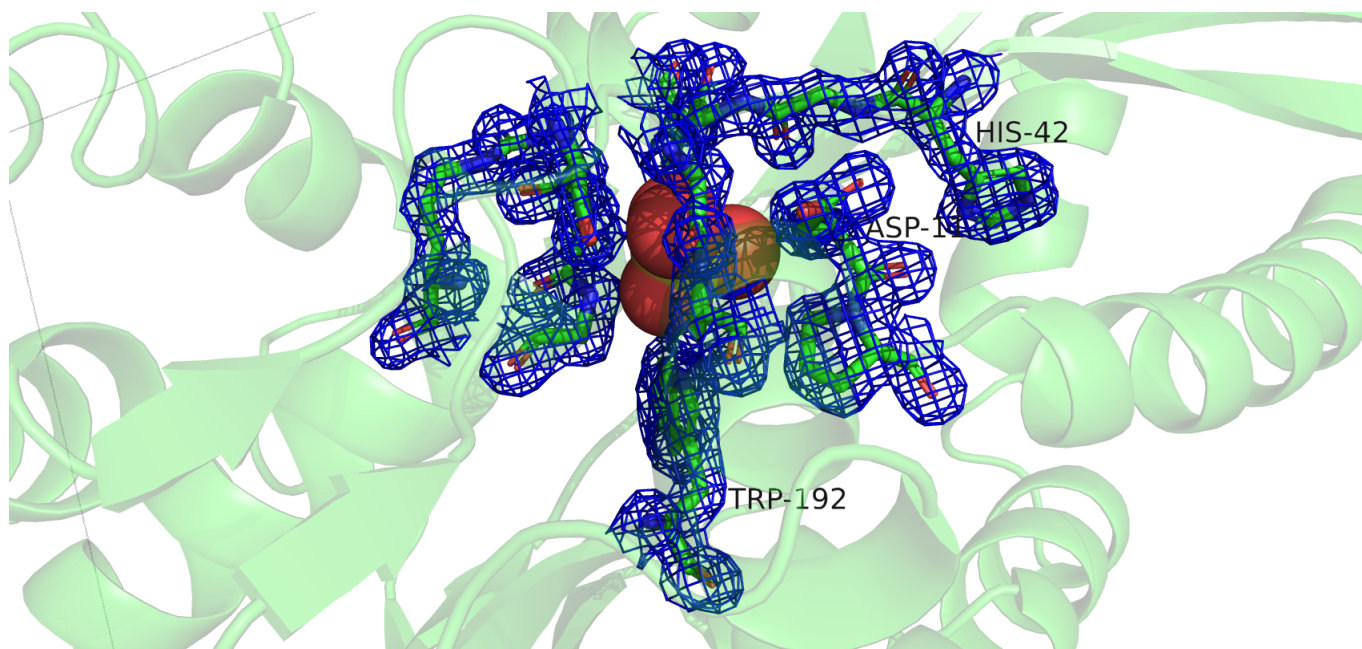


Рис. 10: Остатки, важные для связывания субстрата SO_4^{2-} . Подписаны три остатка, которые с точки зрения MolProbity оказались маргинальными: Asp-11, His-42, Trp-192. Показана ЭП, соответствующая функционально важным остаткам, выбран уровень подрезки $\sigma = 1$. Видим, что ЭП хорошо подтверждает положение всех показанных остатков.

Мы видим, что все остатки идеально вписываются в ЭП. Таким образом, структуру в самом деле можно использовать для изучения функции белка (связывания ионов сульфата).

Затем мы решили воспользоваться сервером [PDB-redo](#), чтобы внести нашу структуру поправки. Поправленную PDB-redo структуру мы проверили с помощью сервера MolProbity (предварительно добавив к ней водороды и инвертировав остатки GLN, ASN). На этот раз MolProbity сообщил только о возможных инверсиях в остатках A-19 GLN и A-258 ASN (сравни с рис. 8). Выдача сервера MolProbity приведена на рис. 11.

All-Atom Contacts	Clashscore, all atoms:	1.45	99 th percentile* (N=819, 1.70Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	4	1.56%	Goal: <0.3%
	Favored rotamers	244	95.31%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	303	98.70%	Goal: >98%
	Rama distribution Z-score	-0.56 ± 0.45		Goal: abs(Z score) < 2
	MolProbity score [^]	1.03		100 th percentile* (N=9248, 1.70Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 2490	0.00%	Goal: 0%
	Bad angles:	1 / 3380	0.03%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 12	0.00%	Expected: ≤ 1 per chain, or ≤ 5%
Additional validations	Chiral volume outliers	0/367		
	Waters with clashes	0/129	0.00%	See UnDownser table for details

Рис. 11: Выдача MolProbity для структуры 1sbr, которую мы исправили с помощью сервера PDB-Redo и инвертировали, добавили водороды с помощью MolProbity.

Ну что тут скажешь, вот теперь это уже совершенно другое дело (ср. рис. 11 и рис. 9)! Показатели структуры значительно улучшились. Остались только 4 остатка – маргинала (предположительно, неправильные ротамеры). Эти остатки – это Thr-32, Leu-65, Gln-108, Thr-303, для нашего белка они не являются функционально важными.

Мы решили также визуально оценить различия в положении остатков до и после поправки с помощью серверов PDB-redo и MolProbity. Для этого мы выбрали 5 остатков, которые перестали быть маргинальными после исправления структуры (см. таблицу 2).

Таблица 2. Всякие там маргинальные элементы.

Остаток (пара остатков)	Признак, по которому является маргиналом	Источник данных о маргинальности
Ile-3(HG21) : Val-238(HG11)	CLASH	MolProbity
Lys-80	Poor rotamer	MolProbity
Asp-79 (CB-CG-OD1)	Bond angles	MolProbity
Glu-253 (CD-OE2)	Bond length	PDB отчет о валидации
Leu-203	RSRZ	PDB отчет о валидации

Рис. 12: Выбранные нами 5 маргинальных остатков.

Мы совместили в Pymol изначальную (с добавленными сервером MolProbity водородами) и исправленную структуры и сравнили положения остатков до

/ после исправления (см. рис 13, 14, 15, 16, 17).

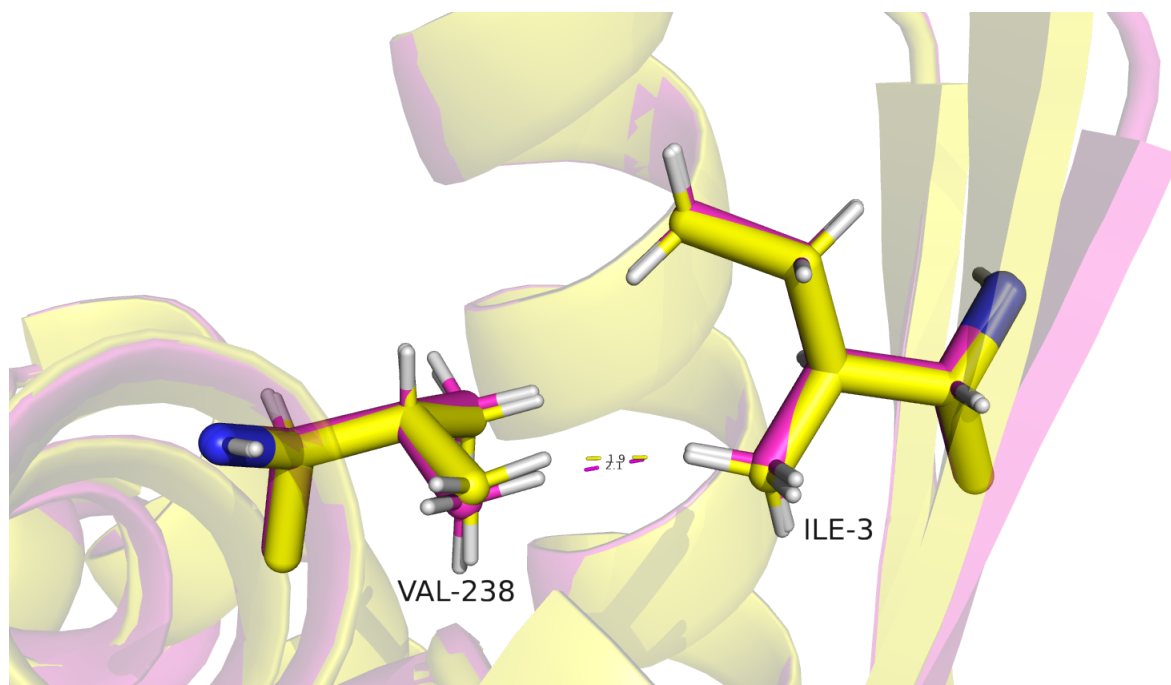


Рис. 13: Розовым показана структура после исправления, желтым – до. К обеим структурам добавлены водороды.

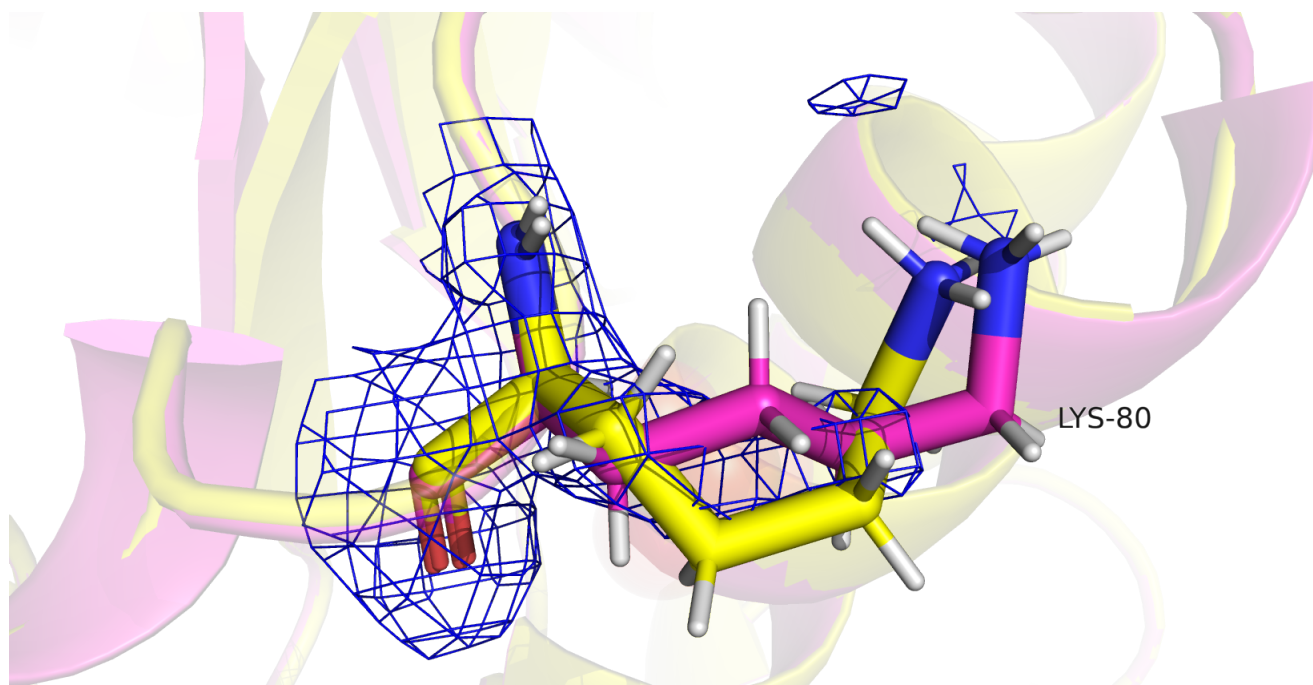


Рис. 14: Розовым показана структура после исправления, желтым – до. К обеим структурам добавлены водороды.

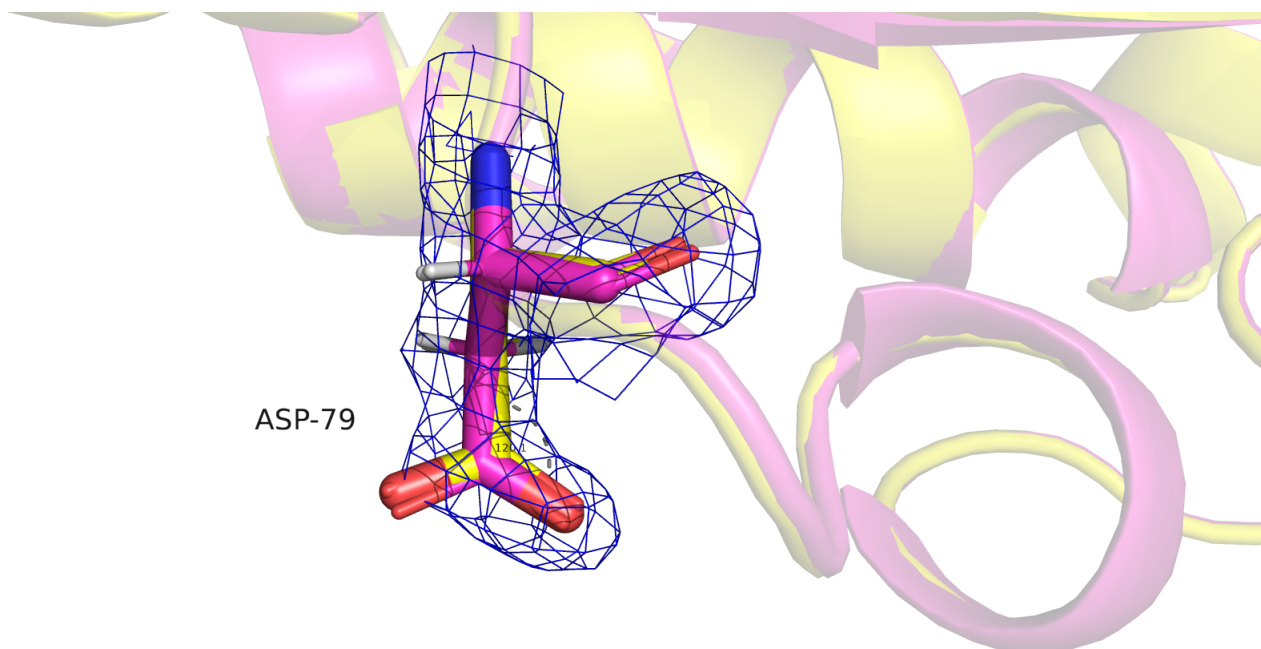


Рис. 15: Розовым показана структура после исправления, желтым – до. К обеим структурам добавлены водороды.

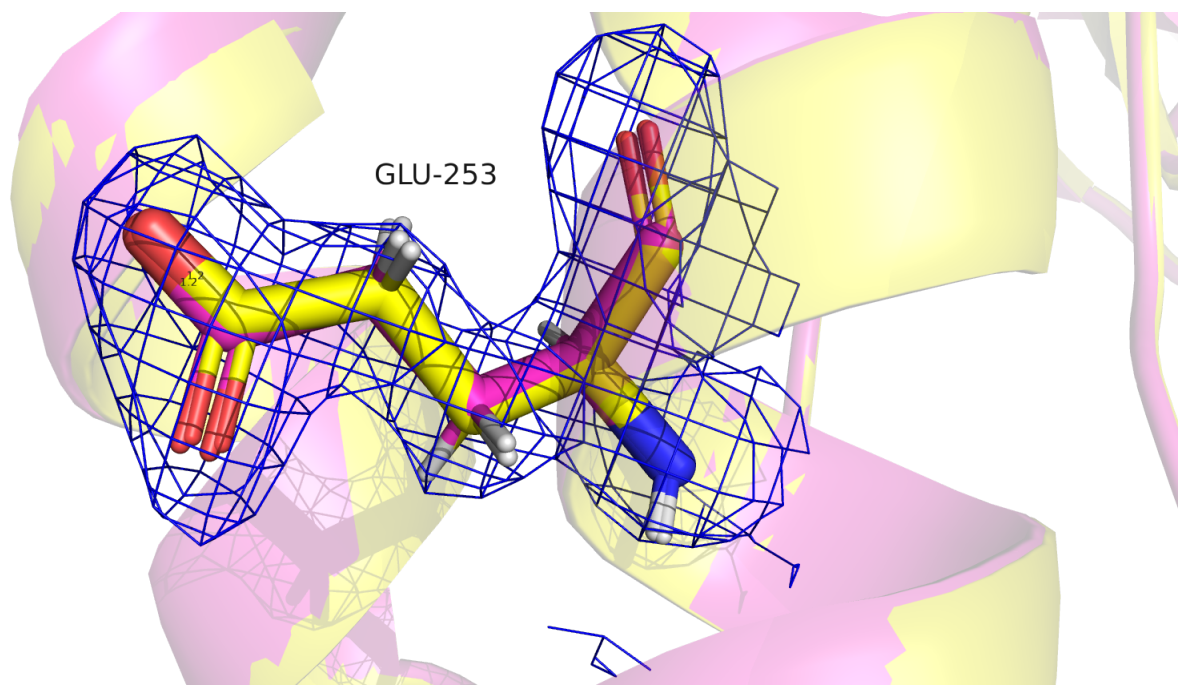


Рис. 16: Розовым показана структура после исправления, желтым – до. К обеим структурам добавлены водороды.

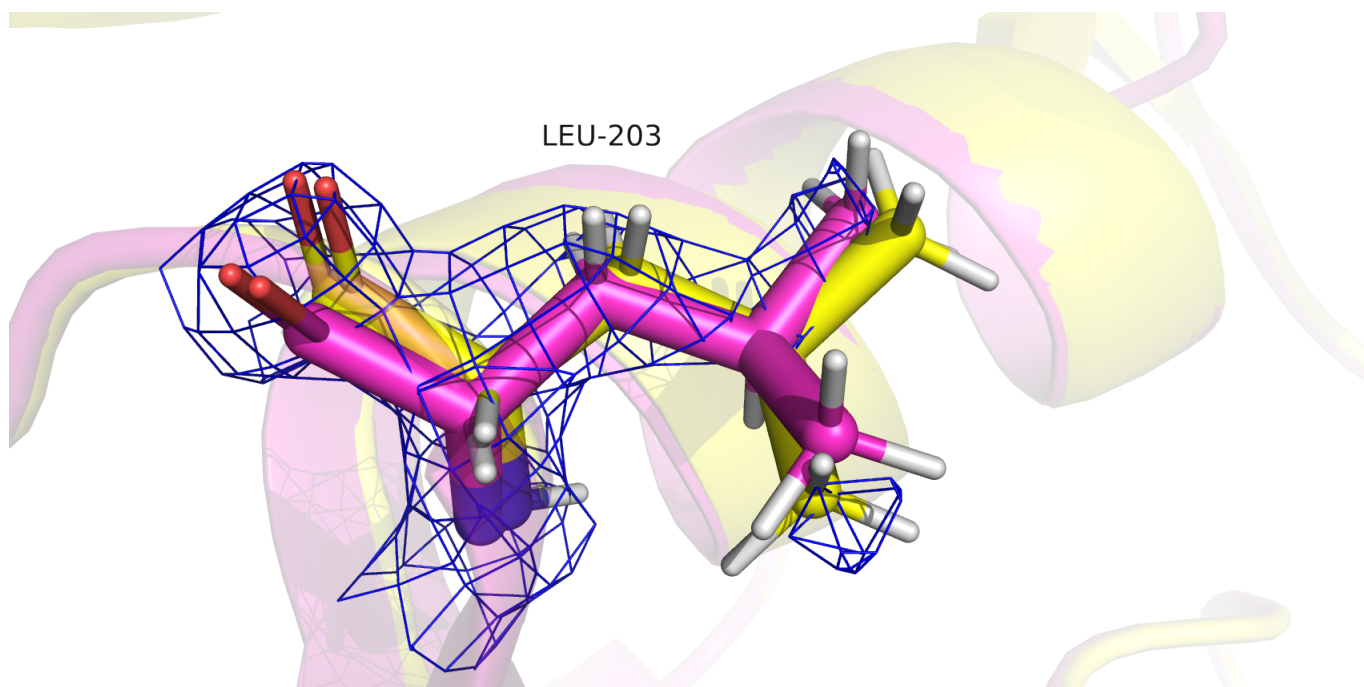


Рис. 17: Розовым показана структура после исправления, желтым – до. К обоим структурам добавлены водороды.

Как мы уже заключили выше для изучения свойств белка (связывания ионов серы) вполне пригодна и изначальная структура. Все поправки, которые были внесены касаются функционально незначимых остатков. Вряд ли они имеют большой интерес для исследователя. Кроме того, мы видим, что иногда поправки и не вполне осмысленные – так, например, на рис. 17 можно заметить, что после исправления структуры остаток Leu-203 стал даже хуже вписываться в электронную плотность (хотя именно в несоответствии ЭП и состояла маргинальность Leu-203 в изначальной структуре).

4 Сессии PyMol

<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac6/session0.pse>
<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac6/session1.pse>
<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac6/session2.pse>

5 ССЫЛКИ

[1] He and Quioco, 'Dominant role of local dipoles in stabilizing uncompensated charges on a sulfate sequestered in a periplasmic active transport protein' 1993 Protein Science

[2] Langridge et al., 'Sulfate-Binding Protein from Salmonella typhimurium: Physical Properties' 1970 Science