

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

Отчет по качеству расшифровки структуры
ванадий-зависимой бромопероксидазы 5LPC

**студентки 4 курса
Березиной Ксении**

Москва, 2016

Аннотация

Провели анализ качества расшифровки структуры 5LPC. Использовали базы данных с параметрами оценки структуры, нашли маргинальные по разным признакам остатки, а также рассмотрели подробно несколько маргинальных остатков. Сравнили модель белка из PDB с оптимизированной моделью из PDB_redo.

Введение

Галогеновый заместитель часто является необходимым в разных органических структурах в клетке. Галогенирующие ферменты включают, в частности, ванадий-зависимые галопероксидазы. Эти ферменты имеют ванадат в качестве простетической группы и катализируют окисление галогенид-иона в присутствии пероксида водорода согласно реакции: $\text{H}_2\text{O}_2 + \text{X}^- + \text{H}^+ \rightarrow \text{H}_2\text{O} + \text{HOX}$, где X^- – Cl^- , Br^- или I^- [1].

На конец декабря 2016 года в базе данных PDB (*Protein Data Bank*) представлено только пять структур ванадий-зависимых галопероксидаз: три эукариотических белка из водорослей и два прокариотических. Структура 5LPC ванадий-зависимой бромпероксидазы цианобактерии *Acaryochloris marina* получена рентген-структурным анализом и имеет разрешение 3.10 Å [2]. Эта структура интересна, поскольку (1) является второй полученной бактериальной ванадий-зависимой галопероксидазой; (2) ванадий-зависимые галопероксидазы стабильны при высоких температурах и в органических растворителях, могут быть базой для разработки галогенирующих ферментов для биоинженерии [2].

Задача этой работы – оценить качество расшифровки структуры бромпероксидазы 5LPC, изучив параметры структуры, маргинальные остатки и оптимизированную структуру в PDB_redo.

Результаты и обсуждение

Общая информация о модели

Структура ванадий-зависимой галопероксидазы 5LPC содержит сам белок (одну цепь длиной 639 остатков) и ион фосфата. Белок является 12-мером (рис. 1).

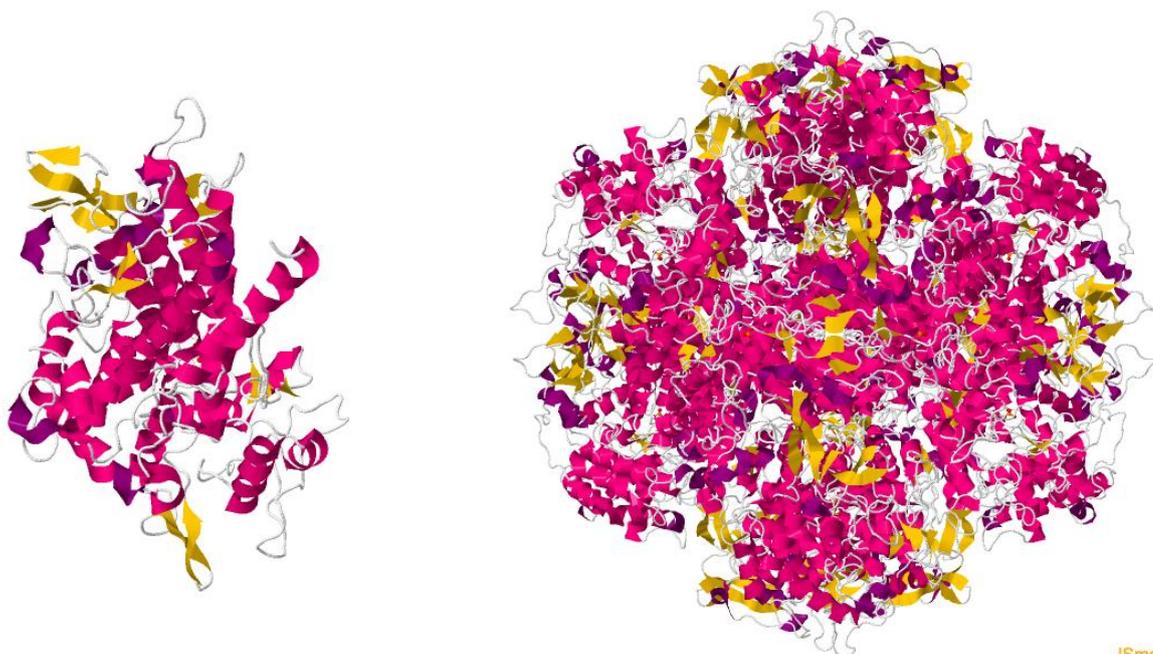


Рис. 1. Ванадий-зависимая галопероксидаза 5LPC: ассиметрическая ячейка и биологическая единица. *Источник: PDB*

Структура занесена в базу PDB в 2016 году авторами Frank A., Groll M. В ходе рентген-структурного анализа фазовую проблему решили молекулярным замещением. 5LPC имеет следующие параметры эксперимента:

- Число измеренных рефлексов: 21555.
- Разрешение структуры: 3.10 Å при полноте набора структурных факторов 95.4 %.
- Диапазон разрешения структурных факторов: 14.96 - 3.10 Å.
- Параметры кристаллографической ячейки: $a=306.16$ Å, $b=306.16$ Å, $c=306.16$ Å; $\alpha=90.00$, $\beta=90.00$, $\gamma=90.00$.

Некристаллографической симметрии в ассиметрической ячейке нет.

Значения индикаторов качества модели в целом

- R-фактор: 0.209.
R-фактор – интегральная мера различия структурных факторов модели и эксперимента: он показывает, насколько расшифрованная структура способна предсказать наблюдаемые результаты.
- R-free: 0.249.
R-free рассчитывается так же, как R-фактор, но по контрольным структурным факторам и окончательной модели, то есть является контролем качества модели. Отличным считается $R\text{-free} < 20\%$ и $(R\text{-free} - R) < 10\%$. Можно считать, что R-free модели 5LPC хороший.
- Для дальнейшего выяснения качества воспользовались сервисом MolProbity, причем при подаче включали атомы водорода.
В целом, оказалось, что у модели отличный clashscore (число недопустимых наложений атомов из-за Ван-дер-Ваальсовых сил), 100-ый перцентиль. Нет недопустимых валентных углов, и нормальное число разрешенных картой Рамачандрана остатков (96%). Однако зашкаливает число недопустимых ротамеров по углам χ : 6% (норма $< 0.3\%$) и маргиналов в карте Рамачандрана.
- Карта Рамачандрана показывает разрешенные углы ϕ и ψ для остатков цепи. В нашей структуре 5LPC 4 остатка лежат вне допустимой области (рис. 2), это 0.65% от всех остатков. Для структур с высоким разрешением этот показатель составляет $< 0.05\%$, поэтому можно сказать, что в этой структуре слишком много маргинальных остатков.

5lpcFH.pdb, model 1

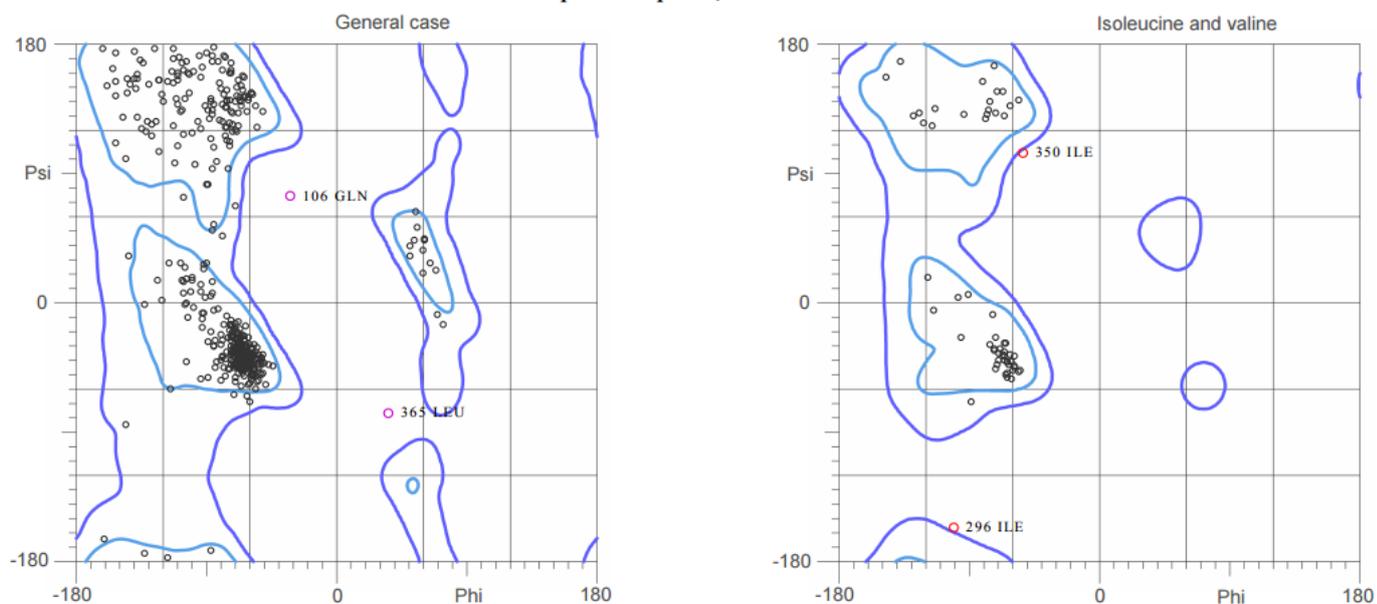


Рис. 2. Карта Рамачандрана структуры 5EJX. Цветные кружочки обозначают маргинальные остатки. Добавлены атомы водорода и все предложенные инверсии Asn, Gln, His. *Источник:* <http://molprobitry.biochem.duke.edu/>

Маргинальные остатки

В таблице 1 приведены некоторые маргинальные остатки структуры. Их искали с помощью сервисов MolProbity (при подаче структуры включили атомы водорода) и EDS. Некоторые признаки маргинальности стоит уточнить. Маргинальность по углам ϕ и ψ связана с картой Рамачандрана, запрещенные ротамеры определяются по углам χ боковых связей. RSR – это пространственный R-фактор для всех остатков и молекул растворителя, который оценивает благоприятность окружения. Худшие остатки по этому параметру определили по пикам Z-score и *significant regions* в EDS.

Табл. 1. Некоторые маргинальные остатки структуры 5LPC. Источник: MolProbity и EDS.

№	Остаток	Причина маргинальности
1	GLN106	В запрещенной области карты Рамачандрана
2	ILE296	В запрещенной области карты Рамачандрана
3	ILE350	В запрещенной области карты Рамачандрана
4	LEU365	В запрещенной области карты Рамачандрана
5	SER47	Неблагоприятная конформация боковых цепей (poor rotamere)
6	THR21	Неблагоприятная конформация боковых цепей (poor rotamere)
7	HIS290	Необходима инверсия боковой цепи
8	ASP42	RSR фактор
9	GLN106	RSR фактор

Надо сказать, что в структуре нет молекул растворителя, поэтому с молекулами воды поработать не удалось. Также при поиске маргиналов не удалось воспользоваться PDBReport, так как в нем нет информации о белке 5LPC (видимо, из-за того, что структура недавняя).

Анализ маргиналов

Бромопероксидаза 5LPC имеет два важных сайта: активный сайт, где происходит превращение субстрата, и сайт димеризации [2]. В активном центре на месте ванадата закристаллизовался стерически похожий фосфат. Аминокислоты активного сайта и цистеины, образующие дисульфидный мостик в сайте димеризации, показаны на рисунке 3. Мономер белка также стабилизируется дисульфидным мостиком по CYS206-CYS308.

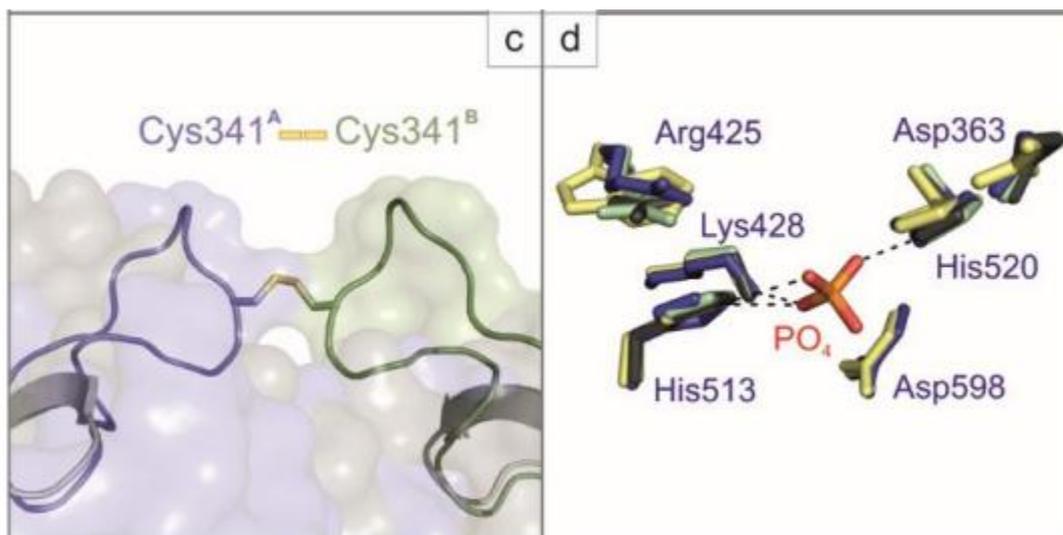


Рис. 3. c) сайт димеризации 5LPC d) активный сайт 5LPC. [2]

Зная эти особенности белка, можно рассмотреть некоторые маргинальные остатки и оценить, являются ли они особенностью структуры или ошибкой эксперимента.

- Остаток LEU365 имеет несколько признаков маргинальности:
 - (1) запрещенные углы по карте Рамачандрана: 36.9, -77.7;
 - (2) запрещенная конформация боковых цепей: углы χ_1 29.8, 160.1;
 - (3) показатель CaBLAM (смотрит на правильное положение CA атомов с точки зрения принадлежности вторичной структуре в эксперименте с разрешением $>2.5 \text{ \AA}$ [3]): 0.025%

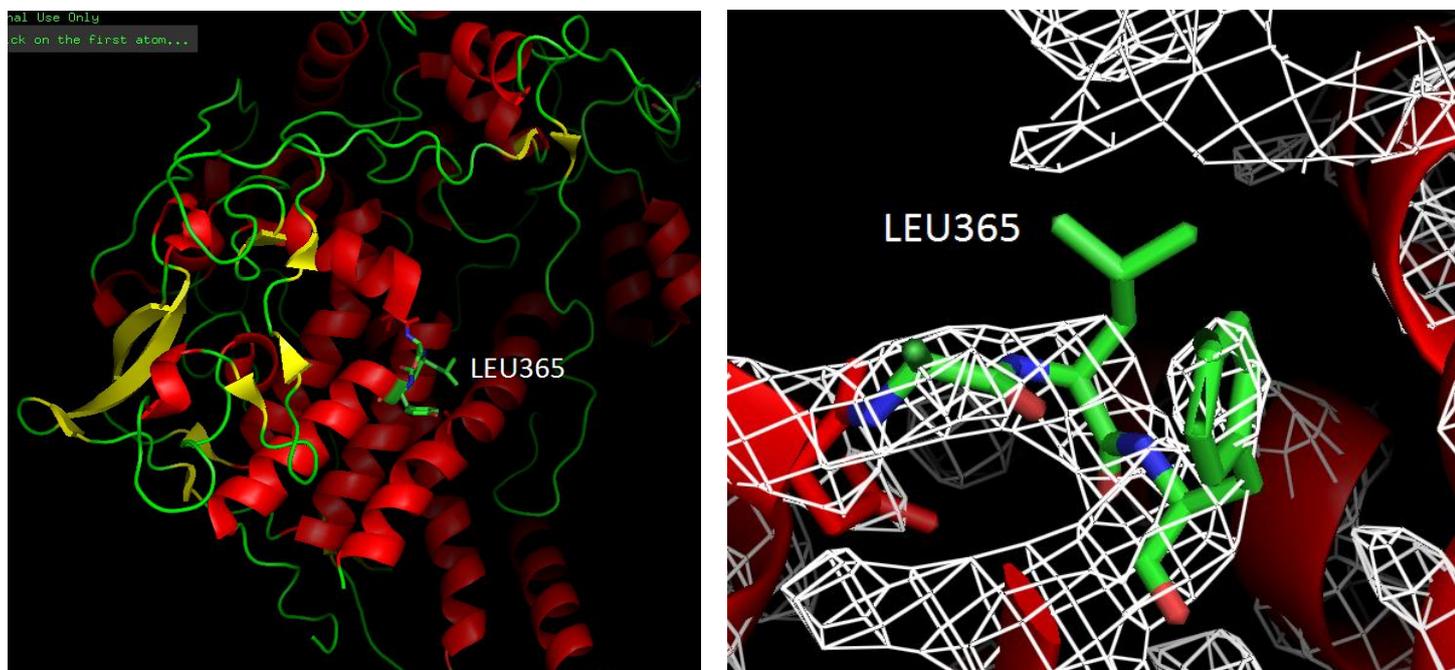


Рис. 4. Маргинальный остаток LEU365. Положение на краю глобулы; совмещение с ЭП: level=1.

Лейцин располагается на краю белковой глобулы и находится в петле, соединяющей две спирали. При совмещении структуры с картой электронной плотности оказалось, что на месте радикала облака электронной плотности практически нет (рис. 4). Тогда визуализировали соседние в кристалле структуры. Одна из структур оказалась близка к области, где лежит наш остаток (рис. 5). На расстоянии меньше 4 \AA от гидрофобного лейцина находятся кислороды THR411 и остова PRO407. Возможно, за счет гидрофобного отталкивания они могли повлиять на конформацию лейцина. Все же, я думаю, что недостаточно оснований полагать, что этот маргинальный остаток – особенность структуры. Скорее всего, маргинальность LEU365 – ошибка эксперимента.

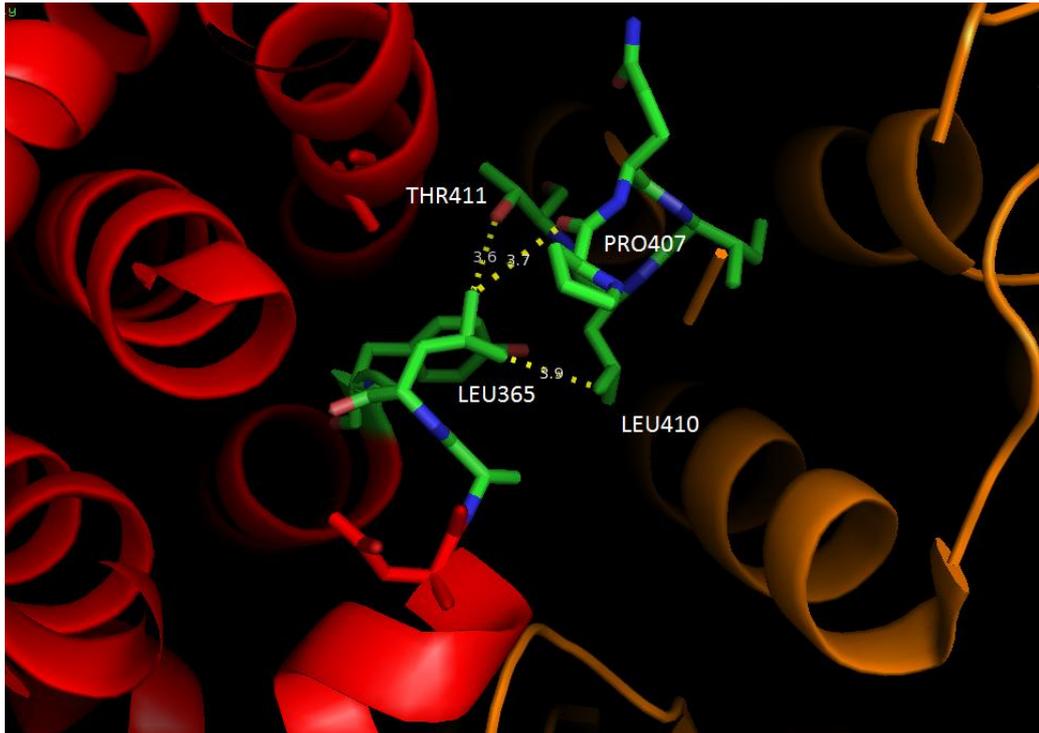


Рис. 5. Расстояние от LEU365 до ближайших остатков белка, соседнего в кристалле (сосед оранжевый).

- В структуре 5LPC многие из маргинальных остатков располагаются на периферии. Например, среди четырех остатков с плохим RSR-фактором ($Z > 2$), оценивающим окружение, три лежат на самом краю глобулы (рис. 6). Причем не понятно, чем обусловлена такая маргинальность. Например, GLN106 является маргиналом по двум параметрам: (1) имеет запрещенные углы по карте Рамачандрана: -32.6, 75.8; (2) большое значение RSR-фактора: $Z=2.8$.

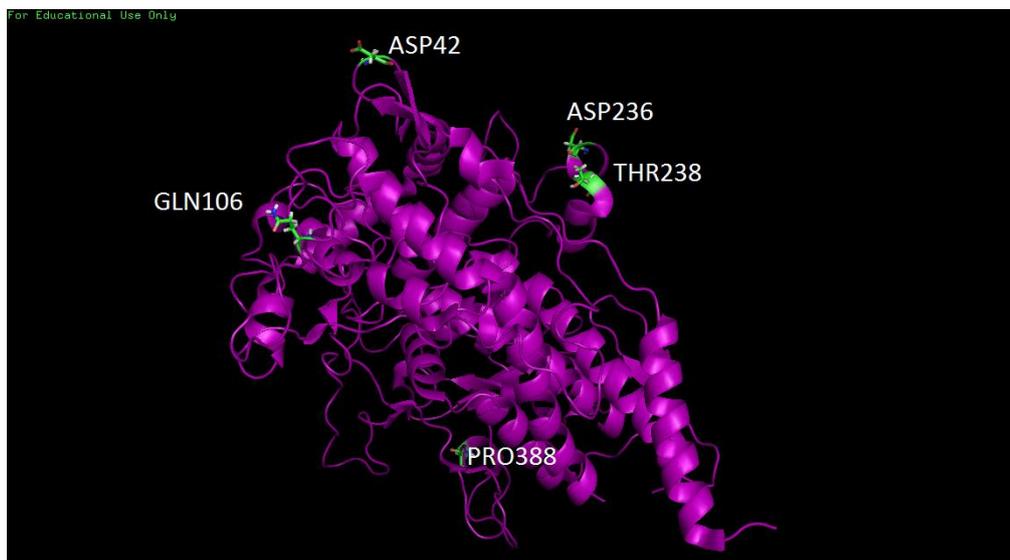


Рис. 6. Остатки с плохим окружением ($Z > 2$).

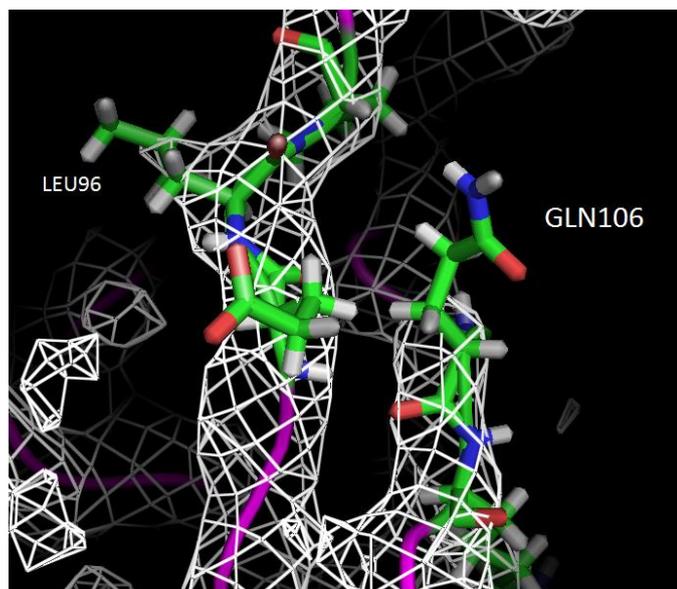


Рис. 7. Маргинал по карте Рамачандрана GLN106. ЭП: level=1.

Если визуализировать электронную плотность, оказывается, что облака электронной плотности на месте радикала почти нет (рис. 7). Каких-то особенностей структуры рядом с этим глутамином нет. Можно сказать, что причина маргинальности GLN106 – ошибка расшифровки.

- GLN483 маргинальный по двум параметрам:
 - (1) запрещенная конформация боковых цепей: углы χ_1 220.6, 74.5, 186.4 (0.2%);
 - (2) пересечение Ван-дер-Ваальсова радиуса водорода HG2 с HG21 THR313: clash 0.54 Å.

Глутамин находится в структуре непосредственно около CYS308, который образует с CYS206 дисульфидный мостик (рис. 8). Эта связь является решающей в формировании третичной структуры мономера белка [2]. Возможно, кислород боковой цепи глутамина отталкивается от кислорода остова CYS308. Таким образом, у глутамина (1) изменяется конформация боковых цепей на нетипичную; (2) Ван-дер-Ваальсов радиус одного из водородов перекрывается с радиусом водорода близкого треонина. Возможно, такая неблагоприятная конформация глутамина должна присутствовать в структуре потому, что белку обязательно иметь мостик по CYS308-CYS206.

Маргинальность GLN483 может быть особенностью белка, а не ошибкой.

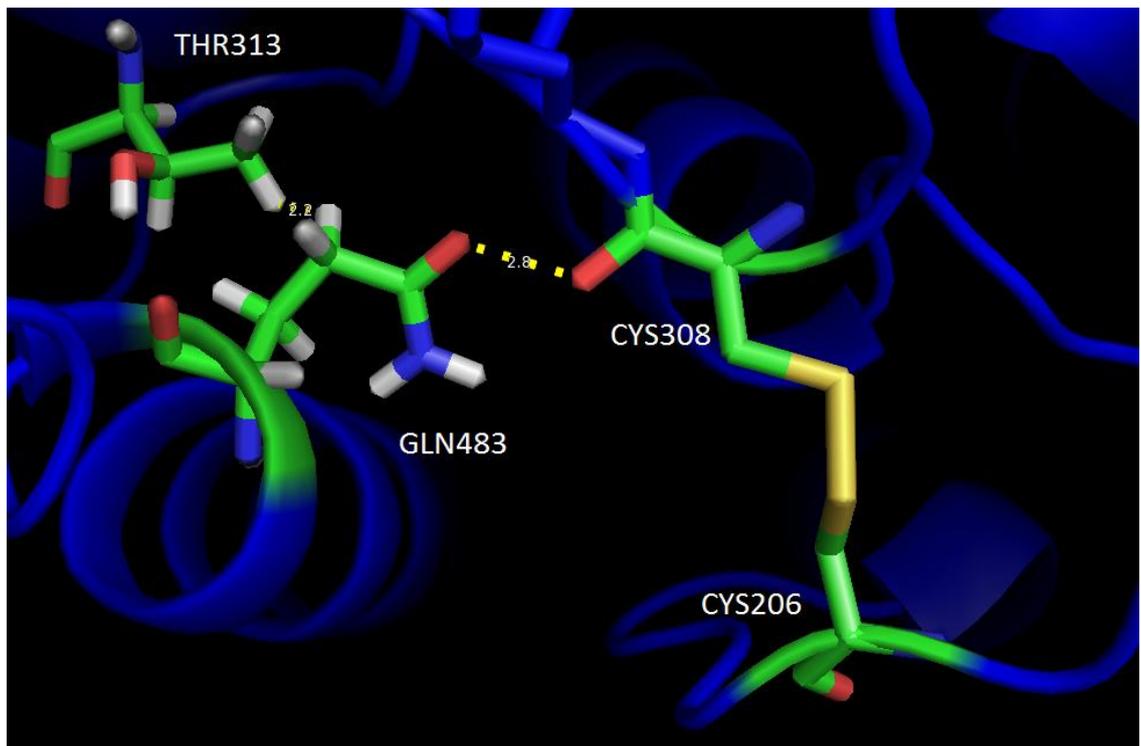


Рис. 8. Маргинальный остаток GLN483 и его окружение в структуре. Серым цветом визуализированы некоторые водороды.

- В белке есть бета-лист на краю структуры, в котором несколько маргинальных остатков. Даже вторичная структура выглядит странно: это криво извернутая 3-10 спираль (рис. 9). На странность вторичной структуры также указывает параметр CaBLAM. Остатки с маргинальными признаками описаны в таблице 2.

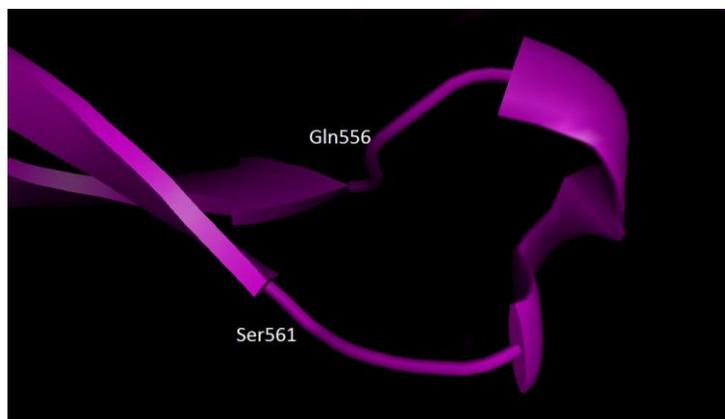


Рис. 9. Изображение вторичной структуры (*cartoons*) участка 556-561.

Табл. 2. Маргинальные остатки участка 556-561. * - разрешенный, но не предпочтительный. Флип – инверсия боковой цепи.

Остаток	Маргинальность до флипа	Маргинальность после флипа (0 – нет, 1 – есть)
Asn556	clash = 0.58 Å (C::Asn556 HD22) запрещенный ротамер: углы хи 71.4,180.1 (0%)	0 0
Gln557	clash = 0.42 Å (O::Asp558 OD1) запрещенный ротамер: углы хи 71.7,131.8,60.8 (0.1%)	1 1
Asp558	clash = 0.42 Å (O:: Gln557 O) запрещенный ротамер: углы хи 100.5,333.8 (0%) CaBLAM неблагоприятный (2.116%) разрешен по Рамачандрану*: 74.9,-15.8 (0.62%)	1 1 1 1
Ser560	серьезное отклонение CaBLAM (0.076%)	1
Ser561	CaBLAM неблагоприятный (2.679%) разрешен по Рамачандрану*: -161.5,-165.3 (1.28%)	1 1

При анализе структуры в MolProbity оказалась предпочтительна инверсия боковых цепей аспарагина 556. Флип поправил пересечение Ван-дер-Ваальсовых радиусов С и HD22 (рис. 10, красная прямая) и конформацию боковой цепи этого аспарагина. Также после флипа не NH₂-группа, а кислород оказалась ориентирована к NH-группам окружающих аминокислот (рис. 10, желтые пунктирные линии). Скорее всего, это выгоднее с точки зрения полярности.

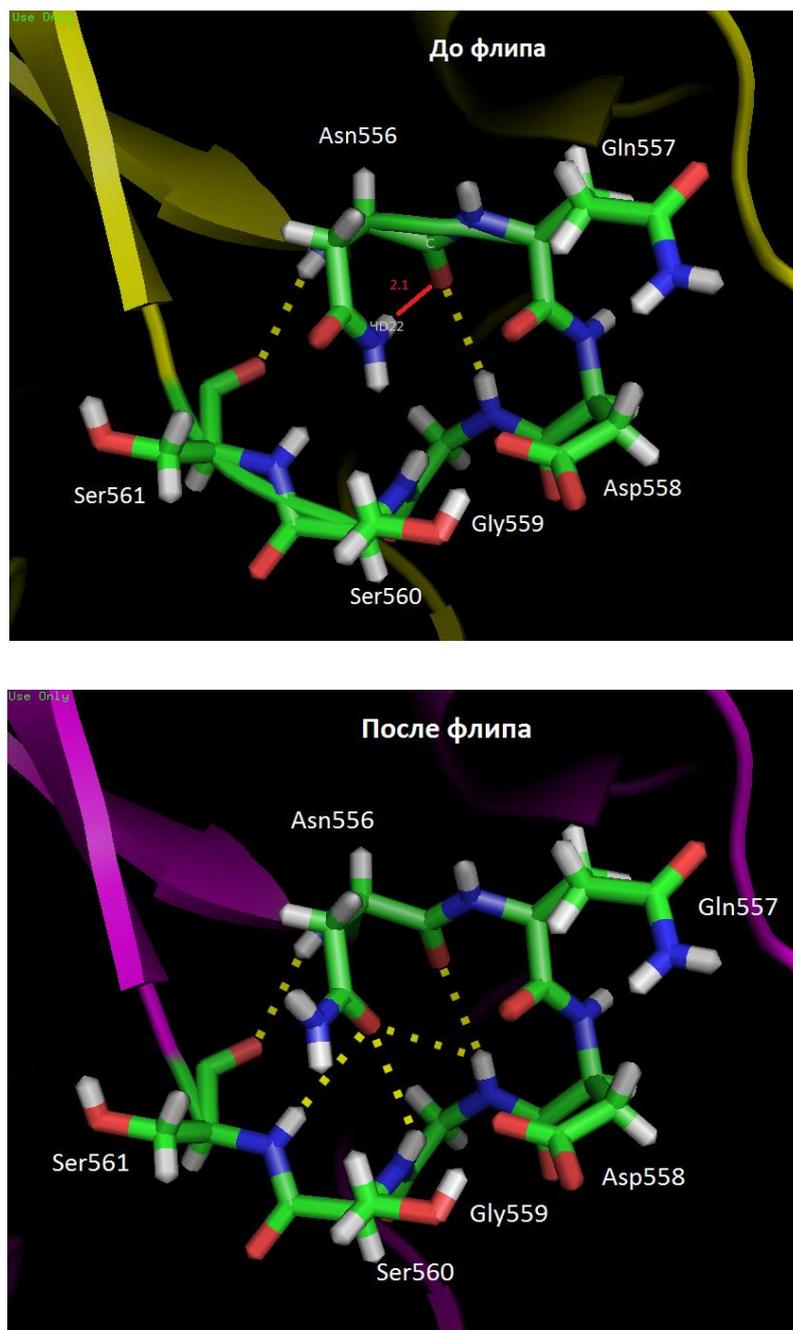


Рис. 10. Положение Asn556 в участке структуры 556-561 до и после флипа. Красная прямая – clash. Желтыми пунктирными линиями обозначены возможные “полярные контакты” (опция в PyMOL, верхний порог 3.2 Å). Числа – расстояния в ангстремах.

Но даже после применения флипа в этом участке структуры осталось много проблем. Конформация боковых цепей Gln557 и Asp558 запрещенная. Причем OD1 аспартата пересекается Ван-дер-Ваальсовым радиусом с радиусом O остова Gln557 (рис. 11, красный пунктир). Кислороды боковой цепи этого аспартата находятся на близком расстоянии от гидроксила Ser560 (2.6 и 2.8 Å). Может быть, имеет место отталкивание отрицательно заряженных кислородов, из-за которого увеличивается угол в –COOH группе Asp558, и OD1 смещается и возникает clash.

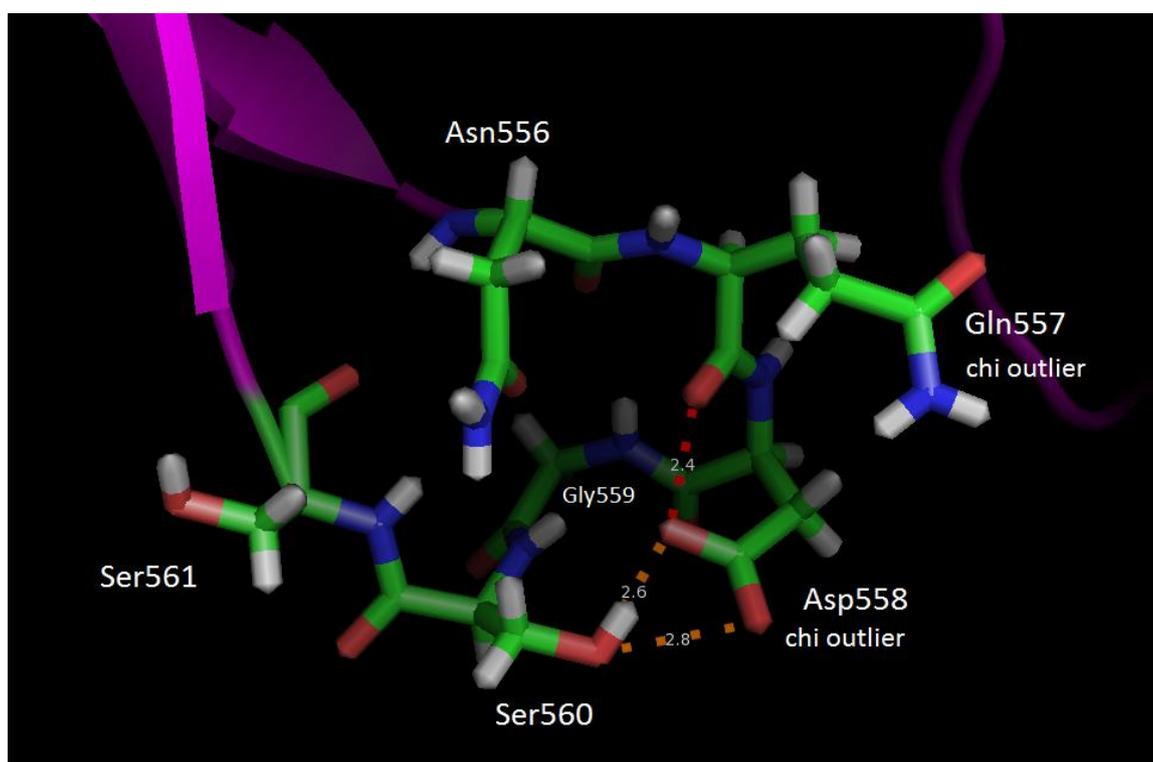


Рис. 11. Участок структуры 556-561. Красным пунктиром обозначен clash O Gln557 и OD1 Asp558. Оранжевым пунктиром обозначено возможное взаимодействие кислородов. Числа – расстояния в ангстремах. Chi outlier – запрещенная конформация боковых цепей.

Вообще говоря, облако электронной плотности на этом участке не дает понять мелкие детали структуры (рис. 12). По нему сложно восстановить конформации аминокислотных остатков. Неудивительно, что в этом участке много маргинальных остатков. Скорее всего, маргинальность аминокислот участка 556-561 возникла из-за ошибок в восстановлении структуры по экспериментальным данным.

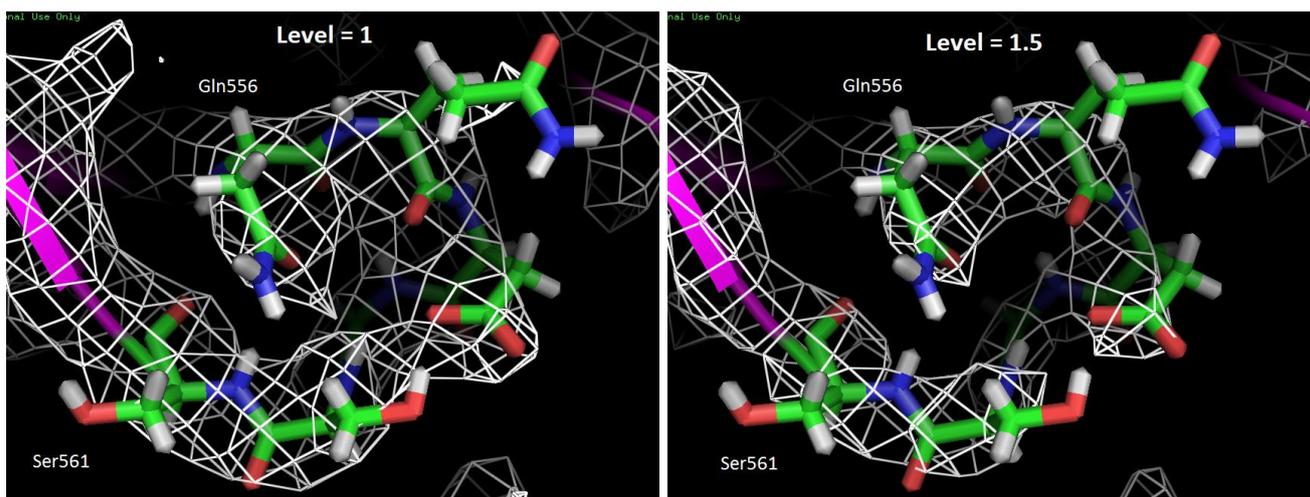


Рис. 12. Визуализация электронной плотности остатков 556-561. Level – уровень восстановления ЭП.

Сравнение модели из PDB с моделью из PDB_redo

Перед тем, как полученную экспериментом структуру помещают в базу данных PDB, она проходит много этапов оптимизации и проверки. Однако это не гарантирует однозначно высокого качества восстановления структуры.

Сервис PDB_redo оптимизирует структуру из PDB стандартной проверенной программой, исходя из представленных авторами PDB файла экспериментальных данных. Этот сервис использует разные алгоритмы, в том числе WHAT_CHECK [4].

WHAT_CHECK выдает параметры изначальной модели (*original PDB entry*), оптимизированной (*re-refined*) и оптимизированной+перестроенной (*final*). Для создания оптимизированной модели программа проверяет координаты атомов и B-факторы (отклонение координат атомов в остатке от средних значений координат). А для создания итоговой модели также перестраиваются боковые цепи, делаются инверсии и др. и модель снова оптимизируется.

В таблице 3 показаны результаты оптимизации. Четыре из семи параметров оценки структуры удалось улучшить. Но упали показатели, оценивающие правильность углов остова и боковой цепи остатков.

Таблица 3. Параметры структуры 5LPC после оптимизации. В последней колонке указано, какие параметры после оптимизации ухудшились или улучшились.

	Original PDB entry	Re-refined	Final	Лучше (+) или хуже (-)
1st generation packing quality	-1.985	-1.977	-1.942	+
2nd generation packing quality	-1.585	-1.375	-1.331	+
Ramachandran plot appearance	-2.319	-2.513	-2.403	-
Chi-1/Chi-2 rotamer normality	-4.440	-4.370	-3.658	-
Backbone conformation	-1.685	-1.834	-1.858	-
Bond length RMS Z-score	0.273	0.401	0.400	<1.000 ok
Bond angle RMS Z-score	0.501	0.641	0.644	<1.000 ok
Total number of bumps (пар атомов со слишком коротким межатомным расстоянием)	42	35	39	+
Unsatisfied H-bond donors/acceptors	85	86	81	+

Все-таки при просмотре итоговой структуры PDB_redo в 3D ясно видны изменения. Например, странный участок 556-561 (который рассматривали при анализе маргинальных остатков) изменил вторичную структуру на более подходящую – петлю, и остатки тоже немного поменяли конформацию (рис.12).

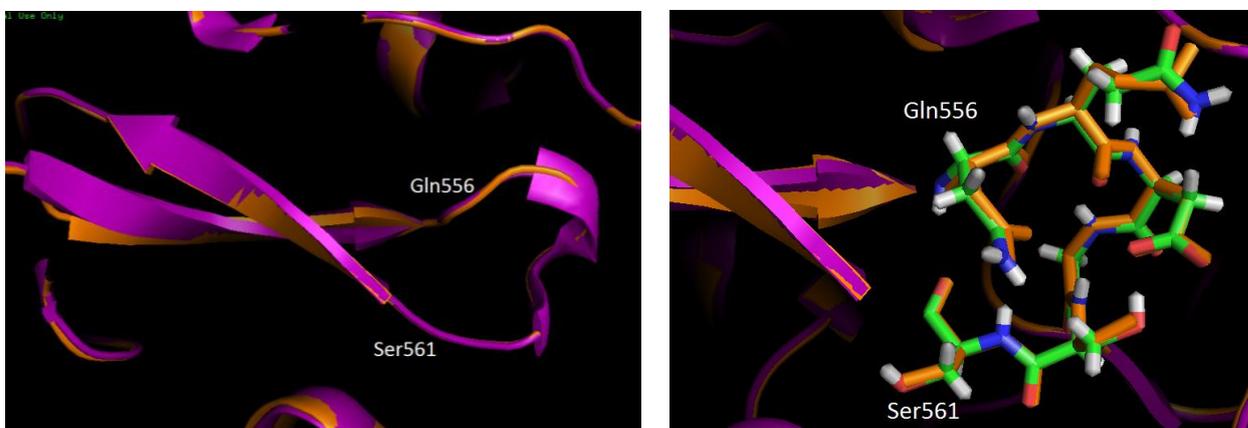


Рис. 12. Сравнение структуры из PDB (фиолетовая и раскраска по типу атомов) с финальной структурой из PDB_redo (оранжевая).

Оптимизацией и перестройкой в PDB_redo была исправлена спорная вторичная структура других участков (рис. 13, бета-листы из двух аминокислот заменены на петли).

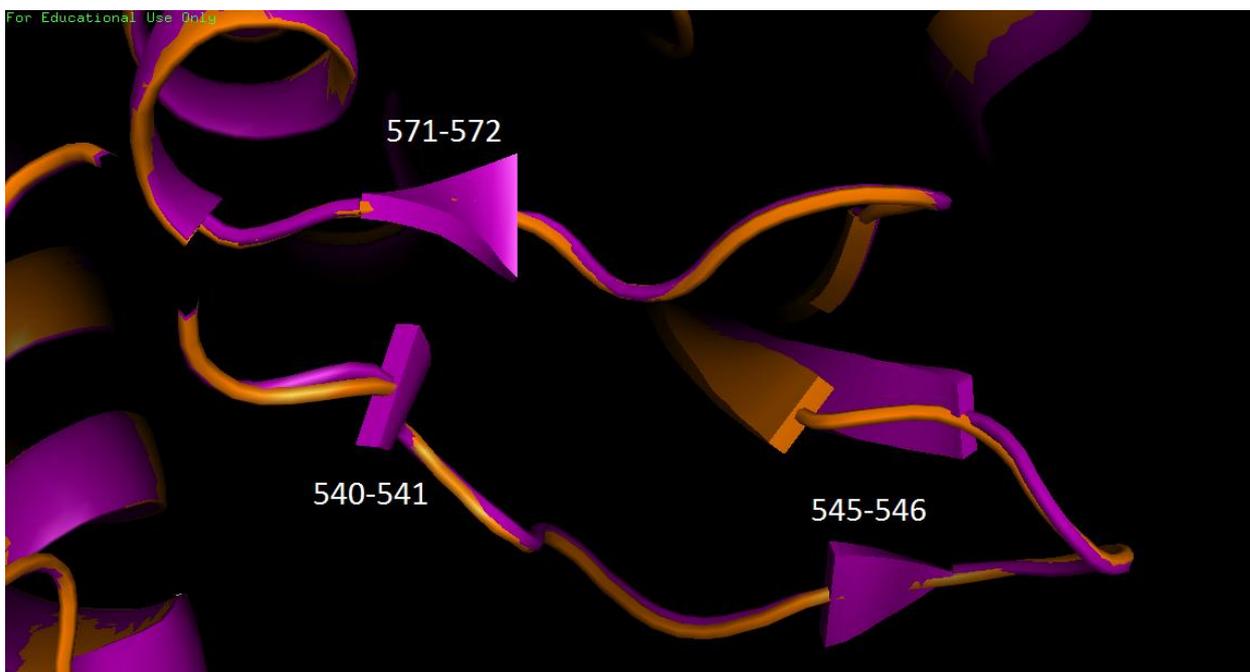


Рис. 13. Сравнение структуры из PDB (фиолетовая) с финальной структурой из PDB_redo (оранжевая).

Заключение

Структура ванадий-зависимой бромопероксидазы 5LPC имеет довольно низкое разрешение: 3.10 Å. Низкое разрешение не позволяет восстановить структуру качественно, возникают ошибки идентификации вторичной структуры и маргинальность аминокислотных остатков, многие из них на краю глобулы. В целом, восстановление структуры не слишком хорошее. Однако удалось восстановить важные остатки активного центра, сайта стабилизации третичной структуры мономера и сайтов димеризации. Возможно, этого достаточно на данном этапе, поскольку 5LPC – только вторая расшифрованная структура среди ванадий-зависимых галопероксидаз прокариот.

Список литературы

1. Fournier J-B, Rebuffet E, Delage L, et al. The Vanadium Iodoperoxidase from the Marine Flavobacteriaceae Species *Zobellia galactanivorans* Reveals Novel Molecular and Evolutionary Features of Halide Specificity in the Vanadium Haloperoxidase Enzyme Family. Lovell CR, ed. Applied and Environmental Microbiology. 2014;80(24):7561-7573. doi:10.1128/AEM.02430-14.
2. Frank A, Seel CJ, Groll M, Gulder T. Characterization of a Cyanobacterial Haloperoxidase and Evaluation of its Biocatalytic Halogenation Potential. ChemBioChem 2016 Nov 3;17(21):2028-2032. doi: 10.1002/cbic.201600417
3. http://molprobity.biochem.duke.edu/help/validation_options/validation_options.html
4. <http://swift.cmbi.ru.nl/gv/whatcheck/>

Сервисы и базы данных

В ходе работы использовали следующие базы данных и сервисы (конкретно указаны в тексте):

1. PDB <http://www.rcsb.org/pdb>
2. EDS <http://eds.bmc.uu.se/eds/>
3. MolProbity <http://molprobity.biochem.duke.edu/>
4. PDB_redo http://www.cmbi.ru.nl/pdb_redo/