

Московский государственный университет им. М.В.Ломоносова  
Факультет биоинженерии и биоинформатики

Отчёт по качеству расшифровки структуры 3E2D,  
полученной методом рентгеноструктурного анализа

Студента 4го курса  
Ефремова Алексея

2017  
Москва

## АННОТАЦИЯ

В данном отчете рассмотрена структура 3e2d щелочной фосфатазы холодоустойчивого штамма *Vibrio* G15-21 (далее - VAP), предложенная в 2009 году авторами Ронни Хелландом, Ренат Лье Ларсеном и Бьярни Асгейрссоном [1].

## ВВЕДЕНИЕ

Щелочные фосфатазы (AP) осуществляют реакцию дефосфорилирования широкого круга субстратов. Щелочные фосфатазы так названы, потому что имеют оптимум pH ~10 у эукариот и ~9 у бактерий. Активность AP позвоночных как правило выше активности AP мезофильных бактерий, и это компенсирует эффект того, что физиологические условия далеки от оптимума. Щелочная фосфатаза из холодоустойчивого штамма *Vibrio* показывает необычно высокую для бактерий активность. Целью статьи [1] являлся поиск структурных причин адаптации к холоду, а также установить, в каком состоянии работает белок: мономерном или димерном.

Известно, что VAP содержит крупную инсерцию (insertion II), уникальную среди её гомологов. Среди семейства щелочных фосфатаз имеются белки, связывающие ионы цинка, магния, железа, кобальта и др. металлов, при этом ионы цинка необходимы для катализа, и аминокислоты, связывающие ионы цинка, являются высоко консервативными. Среди магний-содержащих щелочных фосфатаз варьируется аминокислоты, координирующие магний, и это сильно коррелирует с разницей в активности таких AP.

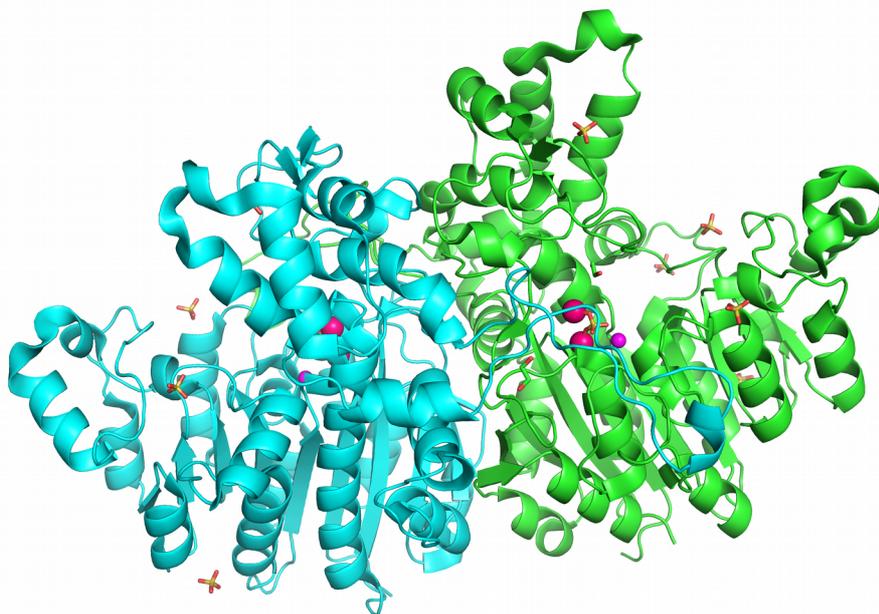


Рис. 1. Структура 3E2D. Видно петлю (insertion II) одного мономера, «обнимающую» другой мономер. Наверху изображения виден так называемый «crown»-домен. Ионы цинка активного центра показаны крупными розовыми сферами, ион магния — мелкой фиолетовой. Сульфаты и этиленгликоль показаны в виде стержневой модели.

## РЕЗУЛЬТАТЫ

### Общая информация о модели.

Модель была получена в 2009 году.

Разрешение модели — 1.40 Å. Число измеренных рефлексов — 195629, полнота набора структурных факторов — 92.6%. Диапазон разрешений — 19.70 - 1.40 Å. Метод решения фазовой проблемы — одноволнового аномального рассеяния (SAD).

R-фактор = 0.155,  $R_{\text{free}}$  = 0.166, при этом доля структурных факторов, использованных для подсчета  $R_{\text{free}}$  составляла 5%. Разница между  $R_{\text{free}}$  и R равна 0.011, что говорит о хорошей обобщающей способности модели и отсутствии переобучения.

Белок представляет собой димер, каждая из цепей содержит в своем составе 502 аминокислотных остатка. Каждый мономер содержит в активном центре два иона цинка и ион магния.

Биологическая ячейка совпадает с асимметрической. Некристаллографических симметрий нет. Тип кристаллографической симметрии —  $P 2_12_12$ . Параметры кристаллографической ячейки —  $a = 118.240 \text{ \AA}$ ,  $b = 165.980 \text{ \AA}$ ,  $c = 57.480 \text{ \AA}$ .

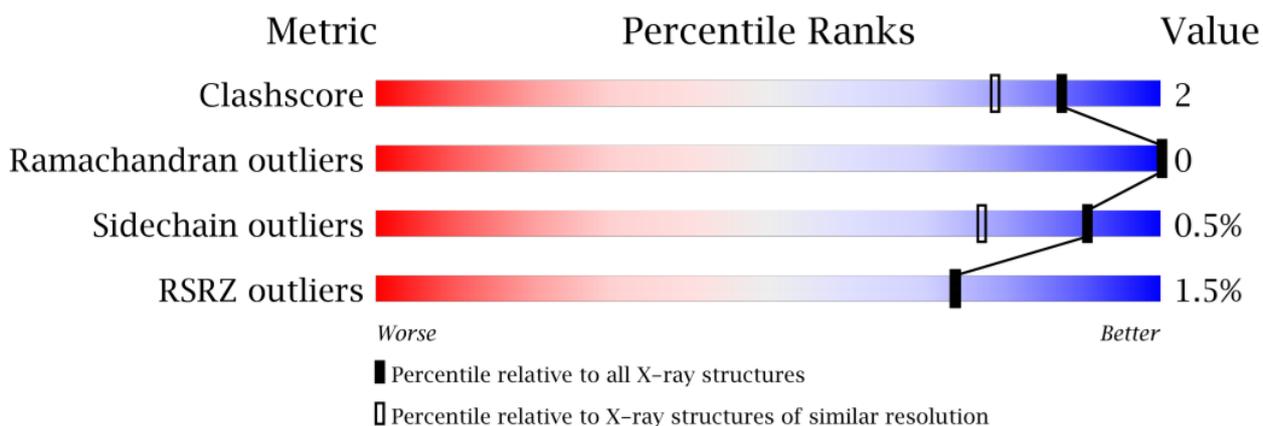


Рис. 2. Основные показатели качества модели по PDB Validation Report [2]. Градиент цвета показывает, насколько параметры данной структуры лучше (синий) или хуже (красный) средних значений (белый). Черные прямоугольники отражают качество относительно всех структур, белые — относительно структур схожего разрешения.

Как видно из рис. 2., модель является достаточно хорошей по основным критериям качества. Структура была загружена в сервис MolProbity [3], в нее были добавлены водороды и произведены инверсии боковых цепей аспарагина, глутамина и гистидина. Из выдачи этого сервиса были получены более подробные данные о качестве модели (Таблица 1).

Таблица 1. Основные показатели качества модели 3e2d, полученные при помощи сервиса MolProbity.

Clashscore, all atoms:	1.31		99 <sup>th</sup> percentile* (N=480, 1.40Å ± 0.25Å)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Poor rotamers	4	0.46%	Goal: <0.3%
Favored rotamers	843	97.57%	Goal: >98%
Ramachandran outliers	0	0.00%	Goal: <0.05%
Ramachandran favored	983	98.30%	Goal: >98%
MolProbity score <sup>^</sup>	0.86		100 <sup>th</sup> percentile* (N=3363, 1.40Å ± 0.25Å)
Cβ deviations >0.25Å	1	0.10%	Goal: 0
Bad bonds:	2 / 8311	0.02%	Goal: 0%
Bad angles:	6 / 11285	0.05%	Goal: <0.1%
Cis Prolines:	0 / 38	0.00%	Expected: ≤1 per chain, or ≤5%
Twisted Peptides:	1 / 1034	0.10%	Goal: 0

В таблице 1 подтверждается информация об отсутствии выбросов в карте Рамачандрана (Рис 3.). В структуре отсутствуют цис-пролины. Все пре-пролиновые остатки, транс-пролины, изолейцины и валины имеют предпочитаемые значения торсионных углов. Часть торсионных углов в остатках глицина слегка выходит за границу предпочитаемых значений.

Параметр Clashscore представляет собой число серьезных перекрываний атомов (Ван-дер-Ваальсовы радиусы соседних атомов пересекаются более чем на 0,4 Å) в расчете на 1000 атомов. Модель 3e2d с достроенными водородами является 99-ым перцентилем среди моделей схожего разрешения по данному параметру качества, что является крайне хорошим результатом.

Структура содержит 4 плохих ротамера, или 0.46% от их общего числа. Это чуть больше порога (<0.3%) для хороших структур. Большинство ротамеров (97.57%) находятся в предпочитаемых конформациях. Это совсем немного меньше порога (>98%).

MolProbity score — общая оценка модели этим сервисом. В данном случае модель является 100-ым перцентилем, то есть лучшей моделью среди моделей сходного разрешения по данному интегральному критерию.

Модель содержит две «плохие» по длине связи, 6 плохих углов и одну цис-пептидную связь без участия пролина.

### Маргинальные остатки.

В таблице 2 указаны некоторые маргинальные остатки по тем или иным критериям.

Таблица 2. Некоторые маргинальные остатки в структуре 3e2d.

№	Аминокислотный остаток	Цепь	Обоснование маргинальности
1	GLN420	B	Cβ deviation = 0.38 Å Выброс по длине связи N-Cα Выброс по углу N-Cα-C Цис-пептидная связь: угол ω = 52.07
2	ALA419	B	Выброс по длине C-O Выброс по углу N-C-Cα
3	HIS490	A	Ротамер: угол χ 264.2, 15.8
4	SER320	A / B	Ротамер: угол χ 218.5 / 219
5	LYS235	A	Clash 0.59 Å NZ-атома с OE2-атомом GLU293 той же цепи

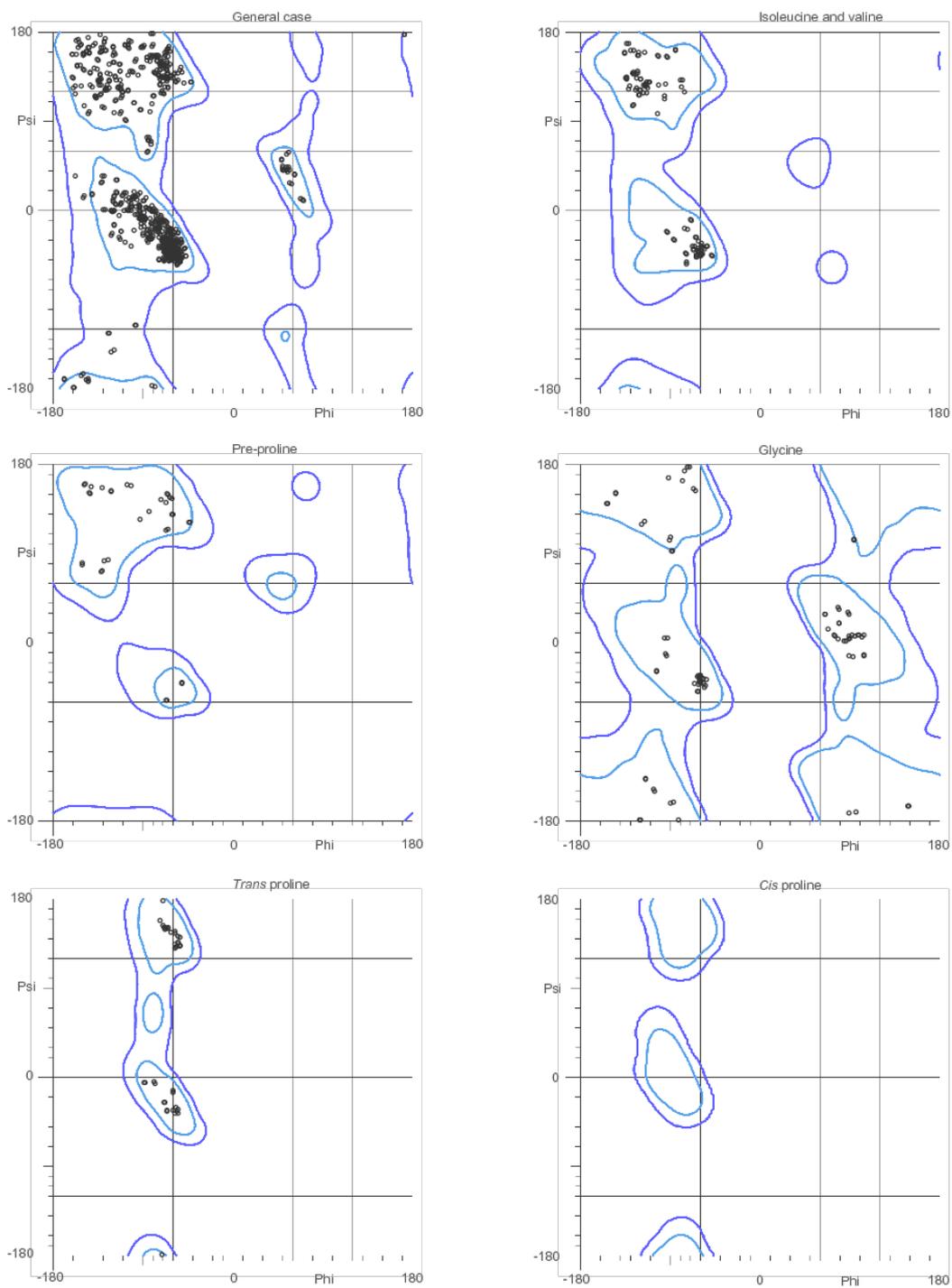


Рис. 3. Карта Рамачандрана для модели 3e2d, построенная сервисом MolProbity. Синими линиями показаны границы разрешенных значений торсионных углов, голубыми — границы предпочитаемых. Точки — конкретные аминокислоты.

## Анализ маргинальных остатков.

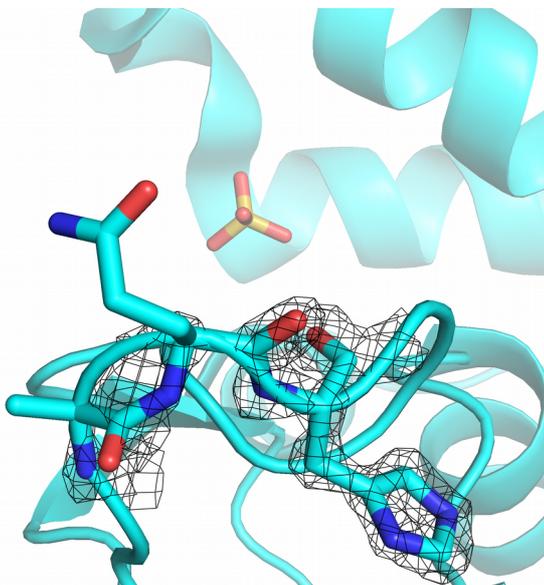


Рис. 4. Цис-пептидная связь между остатками GLN420 и ALA419 в В-цепи. Поверхностями уровней показана электронная плотность вокруг этих остатков, уровень подрезки - 2. Видно, что электронная плотность совсем не покрывает боковую цепь глутамина. Атомы остова относительно хорошо покрыты электронной плотностью.

На рис. 4. показан маргинальный остаток GLN420 и соседние с ним ALA419 и HIS421. Видно, что остов хорошо покрыт электронной плотностью но в нем есть пустые карманы, так что остов мог бы быть вписан и иначе. Боковая цепь глутамина совершенно «голая» в плане электронной плотности. Возможно, это и является причиной неправильной ориентации пептидной связи, а также проблемами с углами и связями в самом глутамине. Скорее всего, данный маргинал — следствие ошибки расшифровки. Рядом расположенный с ним аланин также мог пострадать из-за этого: при вписывании в структуру остатка глутамина и выворачивании пептидной связи некоторым связям в аланине пришлось принять не самые выгодные значения длин и углов.

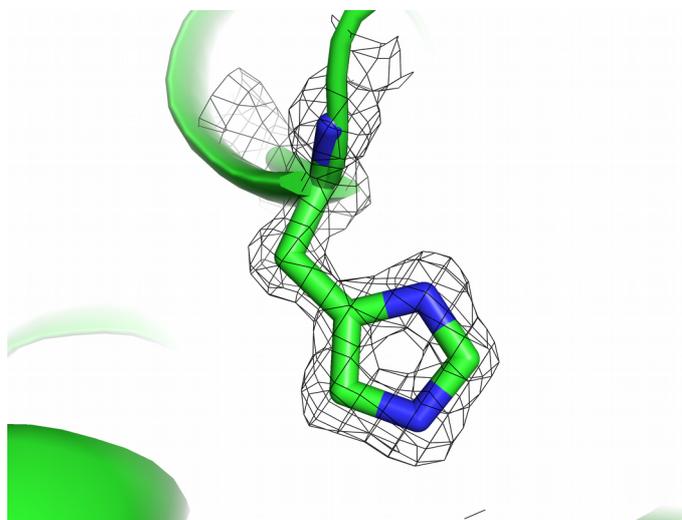


Рис. 5. Боковая цепь HIS490 в А-цепи. Поверхностями уровней показана электронная плотность вокруг остатка, уровень подрезки - 2. Видно, что электронная плотность идеально покрывает боковую цепь гистидина.

На рис. 5. показан ротамер гистидина. Он идеально покрыт электронной плотностью и визуально не выглядит изуродованным или насильно вписанным в модель. Видимо, данные значения углов  $\chi$  являются особенностью данного гистидина, вероятно, из-за стерических затруднений в тесном окружении.

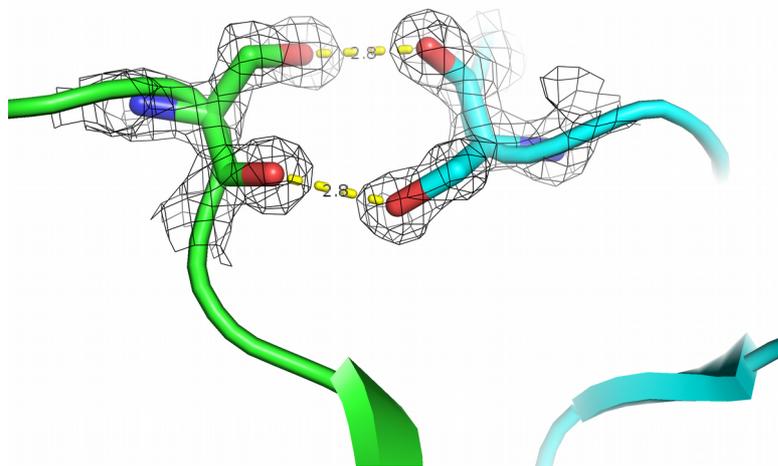


Рис. 6. Боковые цепи SER320 в обеих цепях. Поверхностями уровней показана электронная плотность вокруг остатка, уровень подрезки - 2. Видно, что электронная плотность идеально покрывает боковые цепи серинов.

На рис. 6 показаны два ротамера серинов. Они хорошо покрыты электронной плотностью. Эти остатки могут образовывать водородные связи друг с другом. Они находятся на поверхности контакта двух субъединиц VAP, возможно, этим и объясняется особенность их углов в боковой цепи.

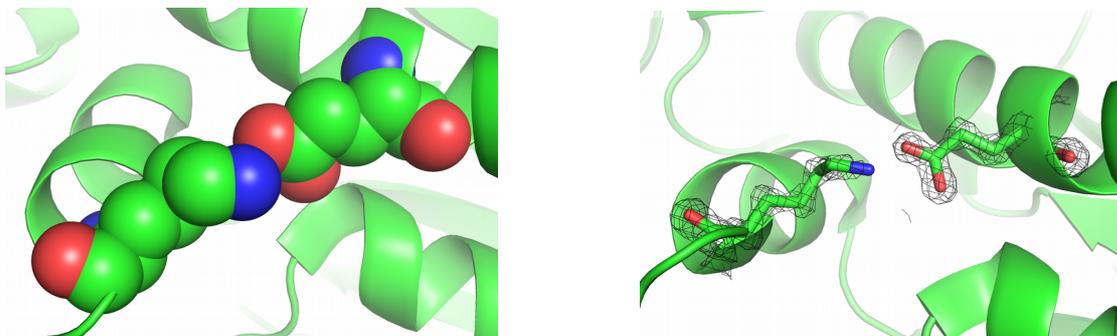


Рис. 7. Аминокислотные остатки LYS235 и GLU293. Слева остатки показаны сферами. Справа остатки показаны стержневой моделью, поверхностями уровней показана электронная плотность вокруг остатков, уровень подрезки - 2.

На рис. 7 показаны перекрывающиеся ван-дер-ваальсовыми радиусами остатки аминокислот. Эти две аминокислоты должны образовать ионную связь внутри белка, но из-за ошибки расшифровки их вообще частично вписало друг в друга, хотя электронная плотность покрывает не всю боковую цепь лизина.

## PDB\_redo.

Из базы данных PDB\_redo [4] была взята исправленная структура 3e2d, и с ней были проведены те же манипуляции в сервисе MolProbity, что и с исходной структурой.

Таблица 2. Основные показатели качества исправленной модели 3e2d, полученные при помощи сервиса MolProbity.

Clashscore, all atoms:	1.37		99 <sup>th</sup> percentile* (N=480, 1.40Å ± 0.25Å)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Poor rotamers	4	0.46%	Goal: <0.3%
Favored rotamers	843	97.57%	Goal: >98%
Ramachandran outliers	0	0.00%	Goal: <0.05%
Ramachandran favored	985	98.50%	Goal: >98%
MolProbity score <sup>^</sup>	0.87		100 <sup>th</sup> percentile* (N=3363, 1.40Å ± 0.25Å)
Cβ deviations >0.25Å	2	0.21%	Goal: 0
Bad bonds:	30 / 8311	0.36%	Goal: 0%
Bad angles:	20 / 11285	0.18%	Goal: <0.1%
Cis Prolines:	0 / 38	0.00%	Expected: ≤1 per chain, or ≤5%
Cis nonProlines:	1 / 996	0.10%	Goal: <0.05%

В таблице 2 указаны показатели качества, выданные сервисом MolProbity. Видно, что незначительно улучшились только количество остатков в предпочитаемой области карты Рамачандрана и интегральная оценка MolProbity. Остальные параметры либо не изменились, либо серьезно ухудшились: так, появилось целых 28 «плохих» связей и 18 «плохих» углов.

Дальнейший анализ этой структуры считаю бесполезным. Возможно, PDB\_redo может улучшить картину для изначально не очень хороших структур, но в данном случае исходная структура уже достаточно высокого качества, поэтому PDB\_redo сумел лишь подогнать несколько аминокислот в предпочитаемую область карты Рамачандрана, что затронуло качество других частей структуры.

## ВЫВОДЫ

Структура 3e2d имеет высокое разрешение и хорошее качество. Пользоваться исправленной при помощи PDB\_redo структурой в данном случае не имеет смысла.

## ИСТОЧНИКИ

1. Ronny Helland, Renate Lie Larsen, Bjarni Ásgeirsson (2009) The 1.4 Å crystal structure of the large and cold-active *Vibrio* sp. alkaline phosphatase. *Biochimica et Biophysica Acta* 1794 297–308
2. [https://files.rcsb.org/pub/pdb/validation\\_reports/e2/3e2d/3e2d\\_full\\_validation.pdf](https://files.rcsb.org/pub/pdb/validation_reports/e2/3e2d/3e2d_full_validation.pdf)
3. <http://molprobity.biochem.duke.edu/>
4. <https://pdb-redo.eu/db/3e2d>