

Отчёт о качестве расшифровки структуры  
галектина LEC-6 *Caenorhabditis elegans*  
методом рентгеноструктурного анализа  
(PDB код 3VV1)

студента четвёртого курса  
факультета биоинженерии и биоинформатики  
МГУ им. М. В. Ломоносова  
Трофимова Дениса

## Аннотация

В данном отчете высказано частное мнение насчет качества структуры 3VV1, сформировавшееся при интерпретации данных сервисов для оценки качества структур белковых молекул *MolProbity* и *EDS*. В ходе работы над отчетом также удалось познакомиться с некоторыми способами улучшения параметров качества структур.

## Введение

Галектины – семейство лектинов – белков, связывающихся с углеводными остатками – связывающихся с углеводными цепочками, содержащие  $\beta$ -галактозидные остатки; галектины играют важную роль в развитии организма и регуляции иммунитета у животных и грибов. Среди позвоночных основным эпитопом галектинов считается дисахарид Gal $\beta$ 1-4GlcNAc; для беспозвоночных основной эпитоп неясен: углеводные цепочки из множественных повторов Gal $\beta$ 1-4GlcNAc, которые являются лигандами галектинов позвоночных, редко встречаются у таких видов как *C.elegans* и *D.melanogaster*. Однако, как недавно стало известно, эндогенными лигандами для LEC-6 (основной галектин организма *C. elegans*) являются N-гликаны, содержащие дисахарид Gal $\beta$ 1-4Fuc. В статье [1] авторы выясняют молекулярный механизм взаимодействия между LEC-6 у *C.elegans* и Gal $\beta$ 1-4Fuc, который отличается от такового у галектинов позвоночных: наличие остатка глутаминовой кислоты (Glu67) в LEC-важно для избирательного связывания остатка фукозы в дисахариде Gal $\beta$ 1-4Fuc.

## Результаты и обсуждение

Общая информация о структуре и показатели её качества

Структура 3VV1 представляет собой димер двух цепей, А и В, массой ~16 кДа каждая и длиной 160 аминокислотных остатков. Также в структуре есть ионы Mg и остатки  $\alpha$ -L-фукозы (FUC) и  $\beta$ -D-галактозы (GAL)[2].

Модель структуры 3VV1 была получена в 2013 году [1], разрешение структуры – 1.5 Å. В эксперименте было измерено 52195 рефлексов, из них удовлетворяло критерию 51804 (99,25%). В самом файле структуры указано число рефлексов как 51775. Разрешение гармоник - 60.065-1.50 Å.

Для решения структуры использовался метод молекулярного замещения, в качестве модели была взята структура галектина-7 человека (PDB ID: 1BKZ). Идентичность по аминокислотной последовательности между моделью и решаемым объектом 28%.

После молекулярного замещения для улучшения фаз была применена программа Parrot из пакета CCP4. При этом, значения R-фактора и  $R_{free}$  изменились с 0.25 до 0.20 и с 0.27 до 0.23 соответственно. Затем, для уточнения модели использовалась программа REFMAC (v5.6), после чего значения R-фактора и  $R_{free}$  стали равны 0.147 и 0.179. При этом контрольная выборка рефлексов составляет 5.1% от общего числа рефлексов.

### Анализ маргинальных остатков

По данным сервиса *MolProbity* [3] структура 3VV1 считается одной из лучших (на основе *MolProbity score*) среди моделей с равным разрешением (1.50Å ± 0.25Å). (Табл. 1)

Таблица 1. Общая информация о модели 3VV1. Красным выделены параметры, не удовлетворяющие пороговому значению для структур подобного разрешения.

Параметр	Значение
Clashscore	0.46 (100 % * (N=598, 1.50Å ± 0.25Å))
Недопустимые ротамеры	1 (0.42%)
Выбросы на карте Рамачандрана	0 (0%)
Остатки в предпочтительной области на карте Рамачандрана	274 (98.92%)
<i>MolProbity</i> score	0.66 (100%* (N=4836, 1.50Å ± 0.25Å))
Отклонения Cβ-атомов > чем на 0.25Å	0 (0%)
Отклонения длин связей остова	22 (0.95%)
Отклонения валентных углов остова	7 (0.22%)

\* – указан процент элементов из выборки структур близкого разрешения, для которых значение данного параметра «хуже» указанного для рассматриваемой структуры

Также, хочется отметить отсутствие в структуре инверсий боковых цепей у остатков Asn/Gln/His.

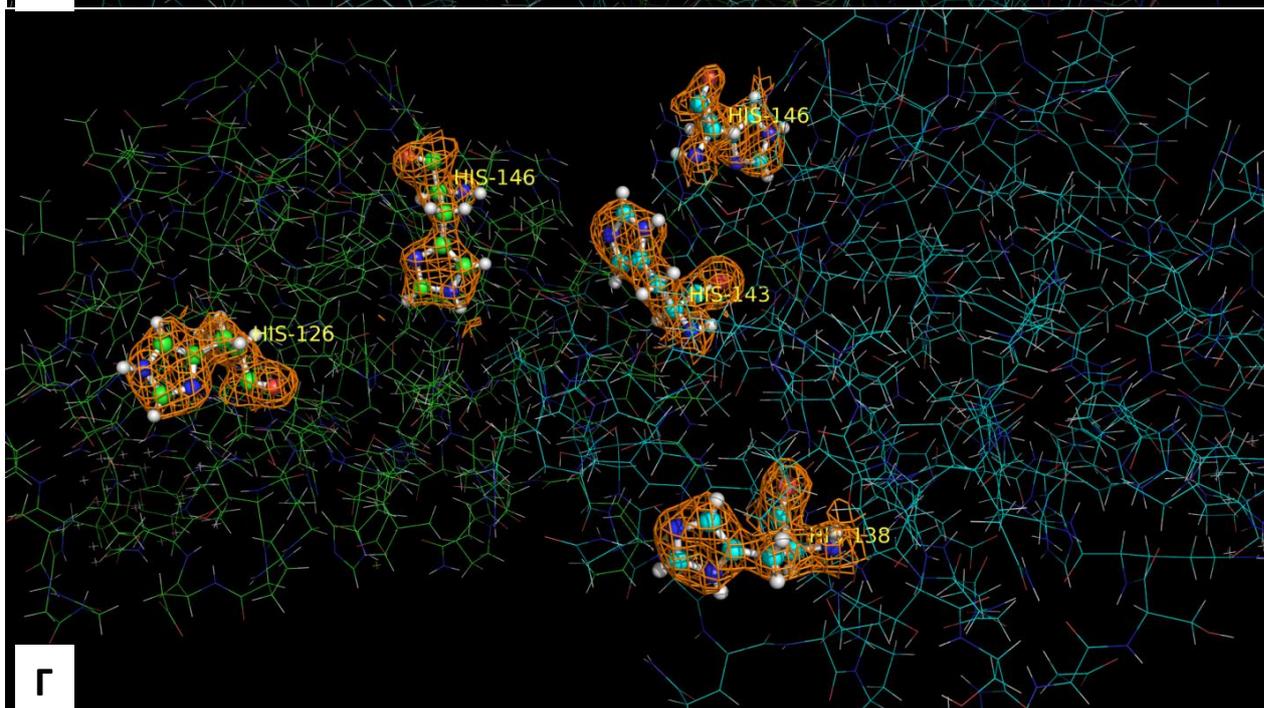
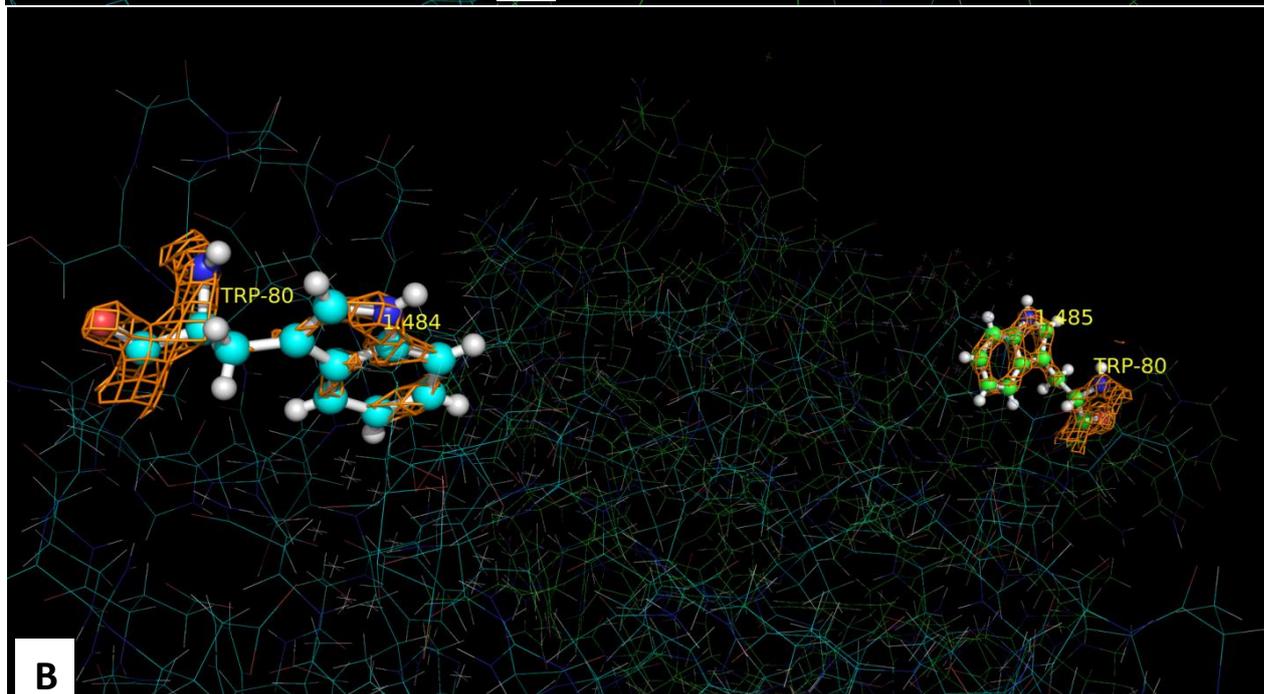
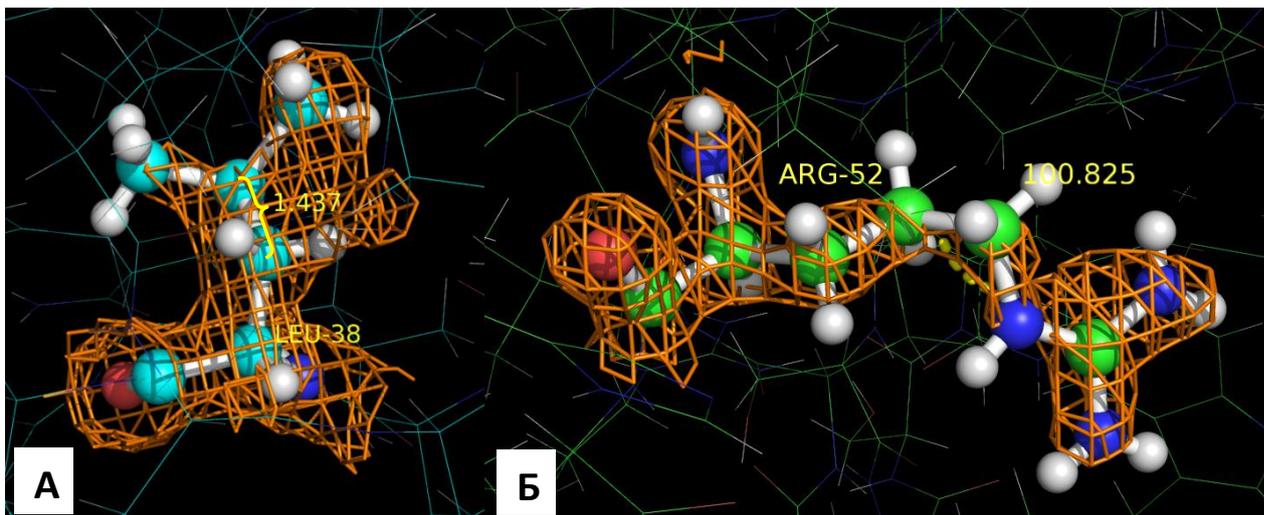
Значения некоторых параметров *MolProbity* неудовлетворительны: для данного разрешения количество недопустимых (отличаются от стандартных значений более чем на 3σ) длин ковалентных связей в структуре должен быть равно 0% от общего числа связей, а количество недопустимых валентных углов - не превышать 0.1% от общего числа (Табл. 2).

Таблица 2. Остатки, геометрия которых ухудшают оценку *MolProbity*

Параметр <i>MolProbity</i>	Остатки, не подходящие под критерий
Длины ковалентных связей	Цепь A: His60, Trp80, His86, His126, His138, His140, His146 Цепь B: Leu38, His48, His60, Trp80, Gln81, His86, Asn93, His115, His129, His138, His140, His143, His146
Валентные углы	Цепь A: Arg52, His129, His143 Цепь B: His146

Т.к. почти все атомы вышеперечисленных остатков, несмотря на геометрию, довольно хорошо разрешены в электронной плотности, то можно говорить об их (вероятно) маргинальном характере (необоснованное отклонение в геометрических параметрах может служить признаком) (Рис. 1)

Рис. 1. Некоторые из вероятно маргинальных остатков. (А) Остаток Leu38 и электронная плотность вокруг него (уровень подрезки 2σ). Нестандартно длинная связь между атомами CB и CG. (Б) Неподходящий угол CG-CD-NE остатка Arg52 довольно хорошо вписывается в электронную плотность (уровень подрезки 2σ). (В) Остатки Trp80 (обеих цепей); несмотря на большое отклонение от средней длины, связям CD2-CE2 в обоих остатках соответствует хорошая электронная плотность (уровень подрезки 3σ). (Г) Остатки His (их слишком много, чтобы перечислять) с отклонениями в длинах связей в боковых радикалах; электронная плотность (1σ) ароматических колец боковых групп не имеет заметных дефектов.



На карте Рамачандрана, отображающей соответствие пар торсионных углов  $\phi/\psi$  остатков цепи, все точки лежат в пределах допустимой области (рисунок 2)

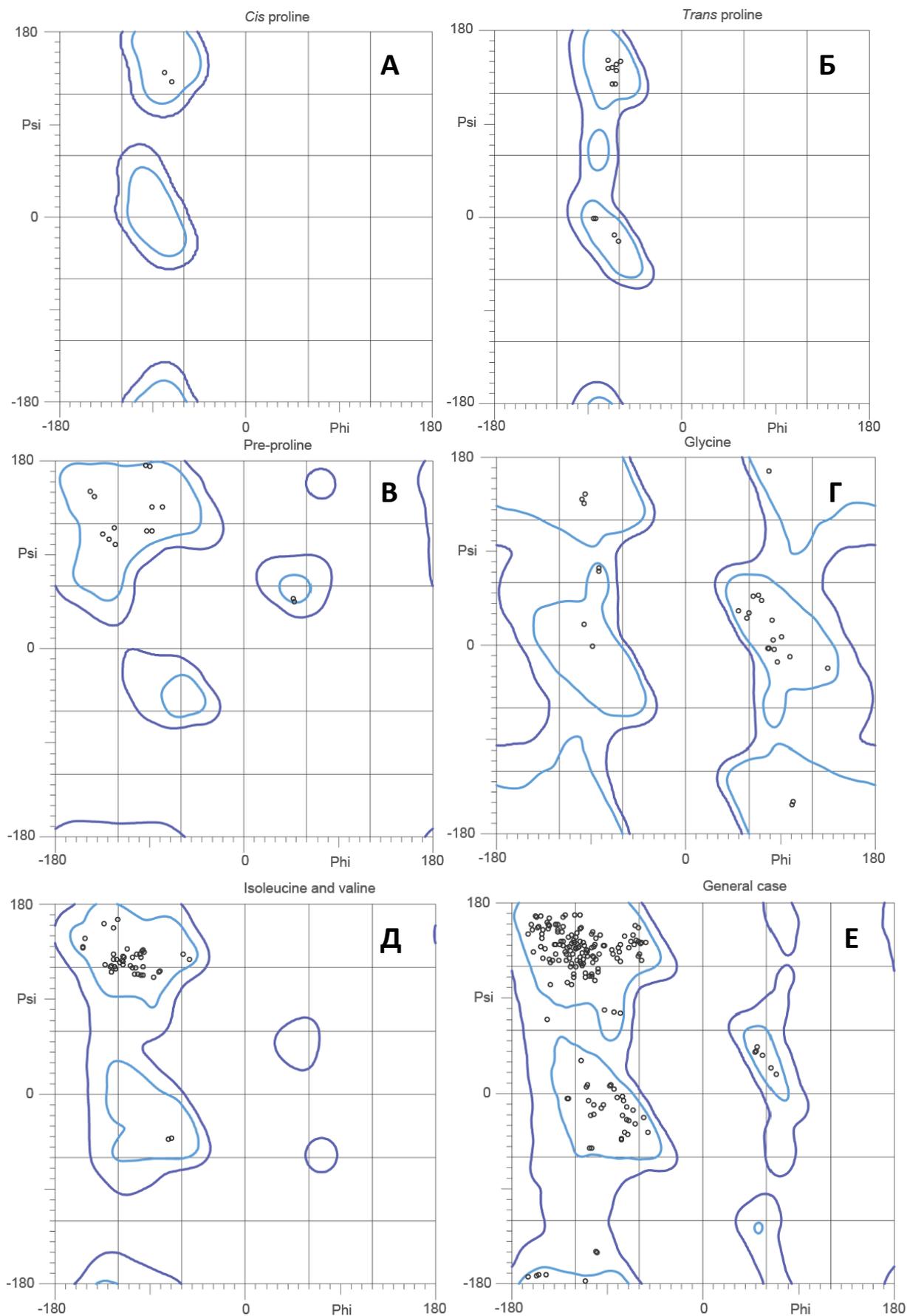
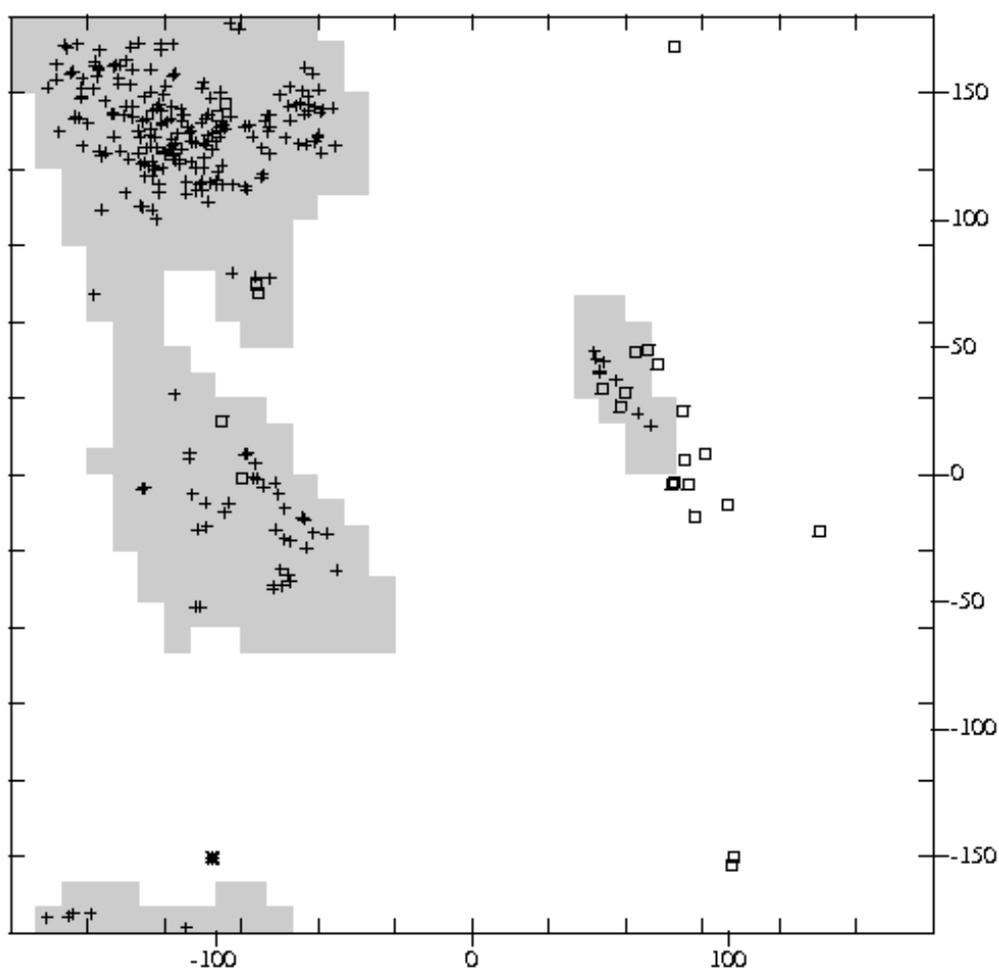


Рис.2 .Карты Рамачандрана, полученные средствами сервиса MolProbity [3], для разных типов остатков модели 3VV1 (слева направо, сверху вниз): цис-пролина (А); транс-пролина (Б); остатков перед пролином (В); глицина (Г); изолейцина и валина (Д); всех остальных остатков (Е).

Хотя по карте Рамачандрана, построенной с помощью сервиса MolProbity (рис. 2), маргинальные остатки отсутствуют, на карте, построенной средствами Electron Density Server (EDS) [4] (рис. 3), по одному остатку в каждой цепи ((A/B)Gln81) отмечены вне разрешённой области (угол  $\phi$ : -101.0/-101.2, угол  $\psi$ : -151.0/-150.6)



X = PHI mapped to [-180,180>

Y = PSI mapped to [-180,180>

Рис. 3. Карта Рамачандрана, полученная средствами EDS, для модели 3VV1. Квадратами отмечены остатки глицина, крестиками – остатки в разрешенной области, звездочками – остатки вне её.

Однако, нельзя назвать данные остатки маргинальными: помимо того, что они находятся на границе разрешенной области, другие параметры также свидетельствуют в пользу «немаргинальности»: усредненный по всем атомам остатка B-фактор в обоих случаях не превышает 20; усредненный RSR не превышает 9%; отсутствует перекрытие ван-дер-ваальсовых радиусов атомов остатков с остальными атомами структуры.

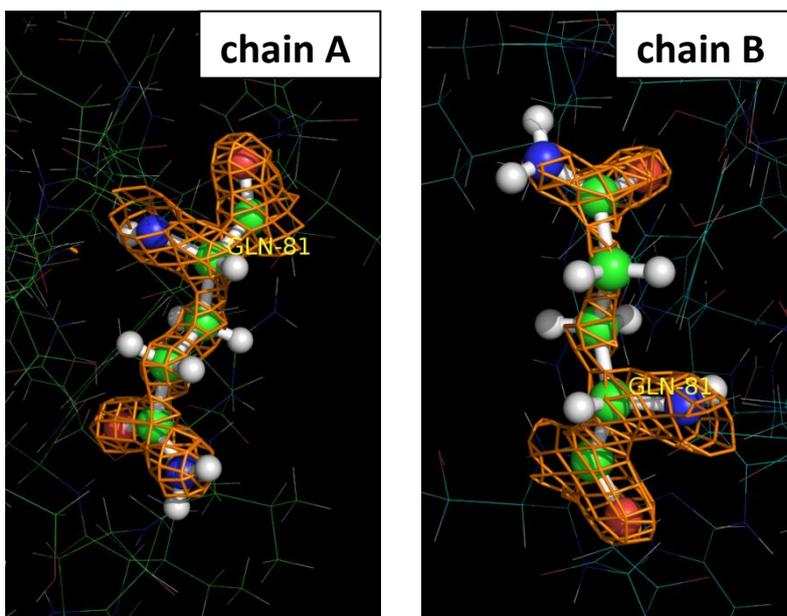


Рис. 4. Остатки Gln81 в модели 3VV1 (соответственно, А и В цепи) и изображение электронной плотности на уровне подрезки 3σ.

Помимо карты Рамачандрана, средствами EDS можно получать и другую информацию. Например, гистограммы пространственного R-фактора (RSR) для обеих цепей:

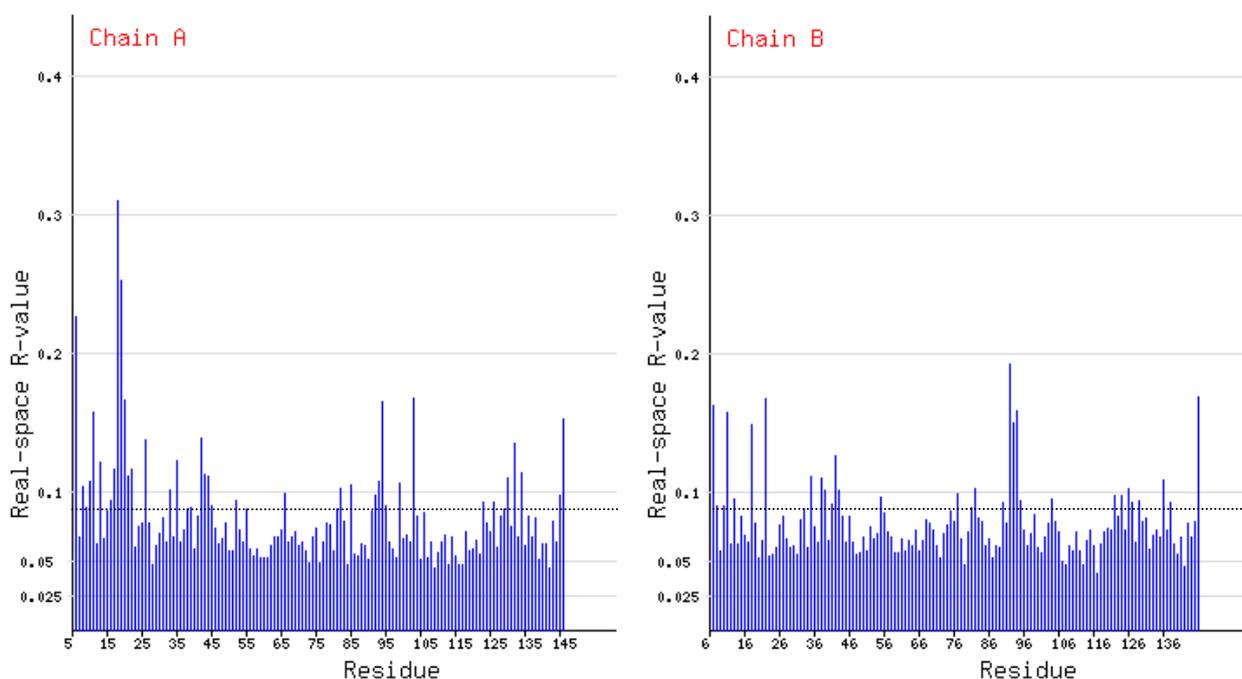


Рис. 5. Значения RSR-фактора для аминокислотных остатков цепей А и В в модели 3VV1.

На фоне усредненного для всех атомов (кроме атомов молекул растворителя) значения RSR, равного 0,080 (ст. откл. 0.033), заметными выбросами кажутся значения RSR для следующих аминокислотных остатков: цепь А – Ile6(0.226), Arg11(0.158), Gln18(0.310), Thr19(0.253), Pro20(0.166), Gln26(0.138), Ser42(0.139), Lys94(0.165), Asn103(0.167), Glu132(0.135), His146(0.152); цепь В –

Gly7(0.162), Arg11(0.158), Gln18(0.149), Asn22(0.167), Gln92(0.193), Asn93(0.150), Lys94(0.159), His146(0.169).

Однако сами по себе выбросы RSR-фактора мало что говорят (фактически, только о плохой подстановке атомов в электронную плотность, что может быть вызвано, например, ручной подгонкой в пользу геометрии). Гораздо лучше рассматривать их с точки зрения *Z-score*:  $Z\text{-score} = (RSR - \langle RSR \rangle) / \sigma$ , т.е. сравнить электронную плотность со средним значением функции электронной плотности для остатков того же типа в структурах с таким же разрешением (1.40-1.60Å). Если остатку соответствует положительный пик длиной больше  $3\sigma$ , то можно считать, что координаты данного остатка не соответствуют электронной плотности, т.е. остаток маргинальный.

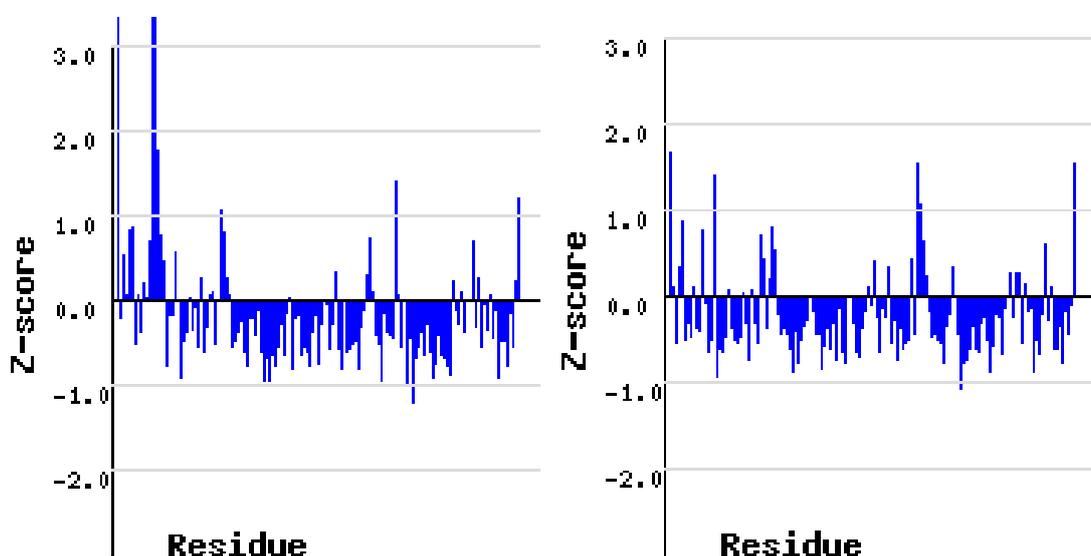


Рис. 6. Гистограмма *Z-score* остатков для цепей А (слева) и В (справа) в модели 3VV1.

Из всех вышеперечисленных остатков, положительные отклонения больше  $3\sigma$  характерны только для остатков Ile6 (4.179), Gln18 (3.529) и Thr19 (3.899) цепи А. Стоит рассмотреть эти остатки подробнее.

Если обратимся к другому параметру, получаемому средствами EDS – температурному фактору (В-фактор), то можем заметить следующие моменты:

- усредненный по атомам всей структуры (за исключением атомов молекул растворителя) температурный фактор равен  $16.5 \text{ \AA}^2$  (ст. откл 6.1).
- есть три остатка со значением В-фактора  $>30 \text{ \AA}^2$ : Ile6 (35.820), Gln18(38.870) из цепи А, Gln92(34.540) из цепи В
- для остатка Thr19 из цепи А В-фактор равен  $26.230 \text{ \AA}^2$

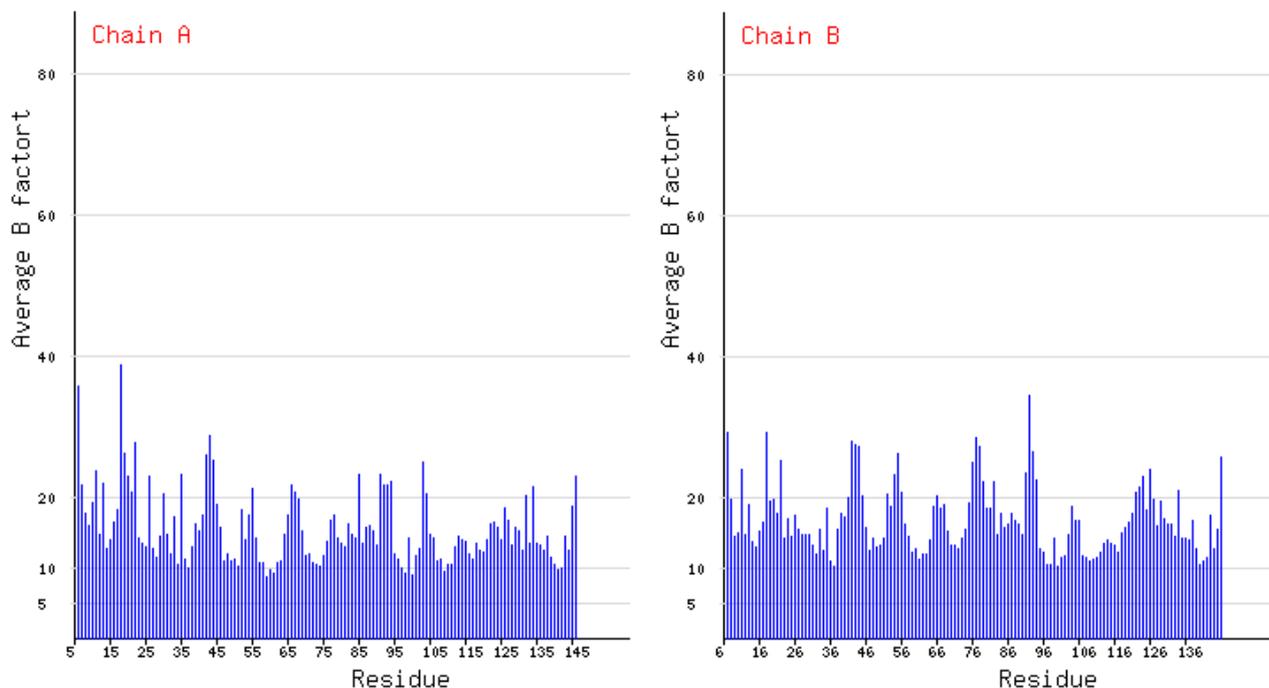


Рис. 7. Распределение значений температурного (B) фактора для атомов модели 3V1

Такое на первый взгляд странное положение вещей можно объяснить, посмотрев на электронную плотность этих остатков:

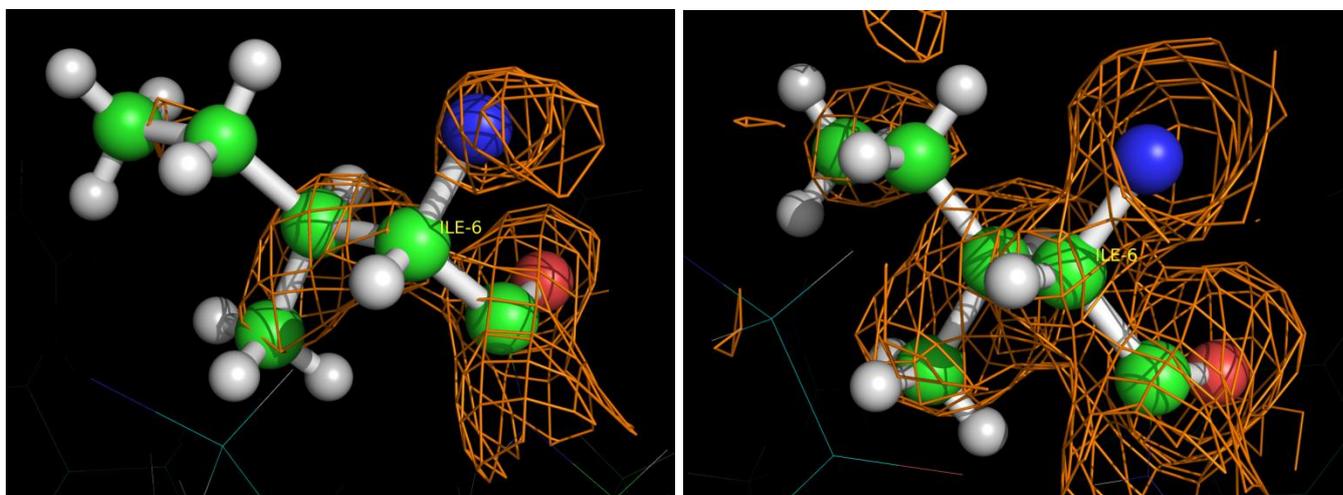
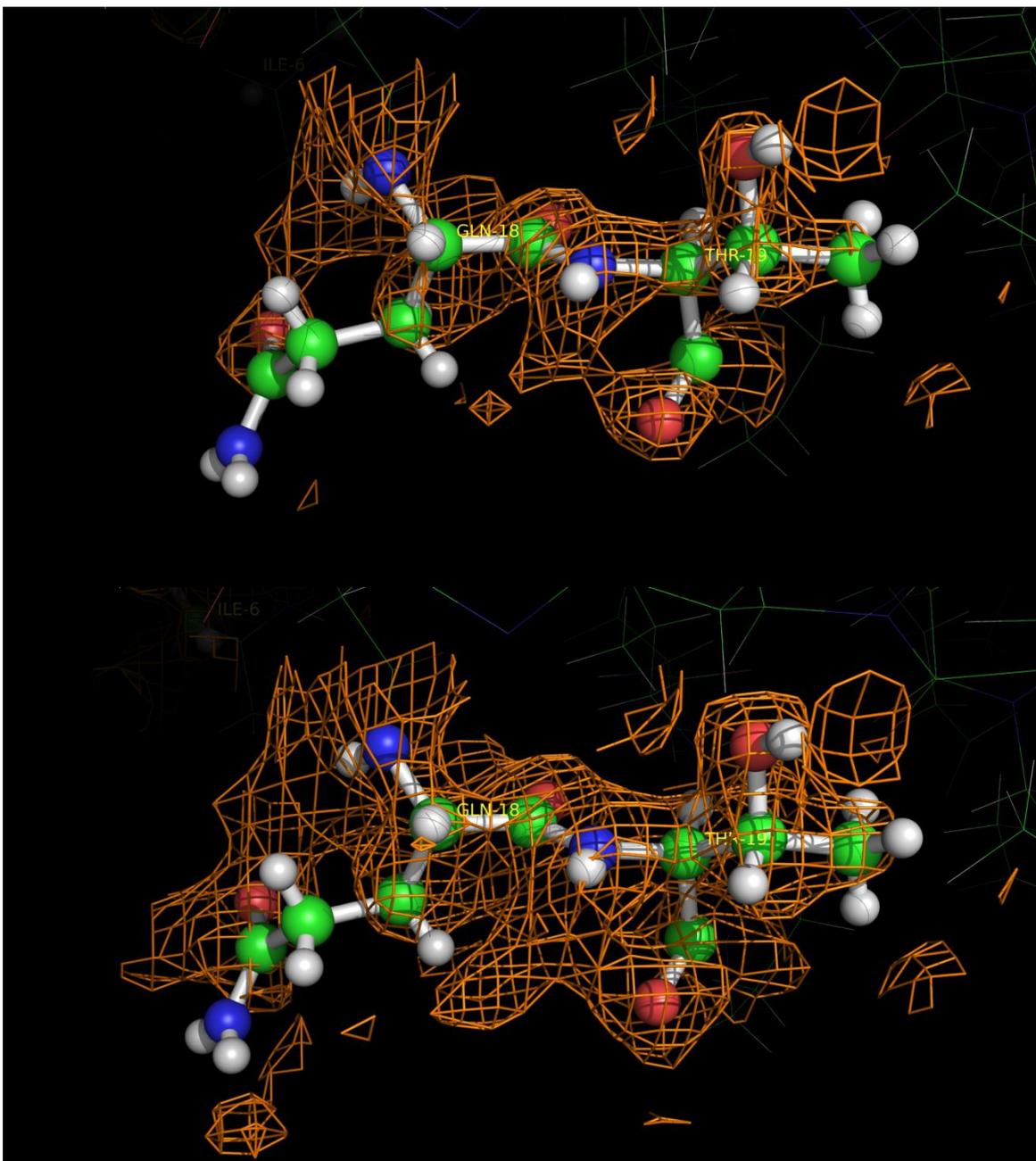


Рис. 8. Электронная плотность вокруг остатка Ile6 цепи A (слева – уровень подрезки  $1\sigma$ , справа –  $0.5\sigma$ ). Следует отметить, что один из атомов (CG1) боковой цепи не покрыт электронной плотностью даже столь низкого уровня подрезки, как  $0.5\sigma$ , вследствие чего возникает такой заметный температурный фактор ( $35.820 \text{ \AA}^2$ )

Рис.9. (на след. странице) Электронная плотность вокруг остатков Gln18 и Thr19 цепи A. (сверху – уровень подрезки  $1\sigma$ , снизу –  $0.5\sigma$ ). Можно заметить странное распределение плотности вокруг боковой цепи Gln18, избегающее, по всей видимости, атом CG. В случае же Thr19 наложение атомов бокового радикала на электронную плотность уровня подрезки  $1\sigma$  как среднее (плохо накладывается атом CG2), а в случае уровня подрезки  $0.5\sigma$  – как отличное (все атомы хорошо наложены). Значение температурных факторов равны  $38.870$  и  $26.230 \text{ \AA}^2$  соответственно.



Таким образом, довольно низкое по сравнению с остальными маргинальными остатками значение температурного фактора остатка Thr19 можно объяснить лучшим наложением атомов на электронную плотность (температурный фактор атома, не покрытого электронной плотностью, объективно выше, чем у остальных). Тем не менее, это не меняет тот факт, что атомы остатка Thr19, по сравнению с атомами остальных аминокислотных остатков, существенно хуже вписаны в электронную плотность, т.е. подтверждает маргинальный характер [остатка].

## Сравнение модели из PDB с моделью из PDB\_REDO

Интересно сравнить структуру 1UWZ из PDB с оптимизированной моделью из PDB\_REDO [5].

Параметр	PDB	PDB_REDO
$R_{\text{free}}$	0.1790	0.1888
R-фактор	0.1470	0.1709
$R_{\text{free}}-R$	0.032	0.0179
Число остатков с:		
Отклонениями длин связей остова	22 (0.95%)	0
Отклонениями валентных углов остова	7 (0.22%)	0
Число маргиналов по карте Рамачандрана [3]	0	0
Число остатков в предпочитаемой области карты Рамачандрана [3]	274 (98.92%)	273 (98.56%)
MolProbity score	0.66	0.87

Таблица 3. Некоторые параметры качества построения структур 3VV1 из PDB и из PDB\_REDO

В целом, можно сказать, что оптимизация структуры средствами PDB\_REDO позволила устранить часть аномалий в модели 1UWZ (например, отклонения длин связей и валентных углов остова), а также улучшить общее качество структуры. Но следует также отметить, что это несколько противоречиво, так как вместе с уменьшением разности  $R_{\text{free}} - R$  заметно возросли  $R_{\text{free}}$  и R-фактор.

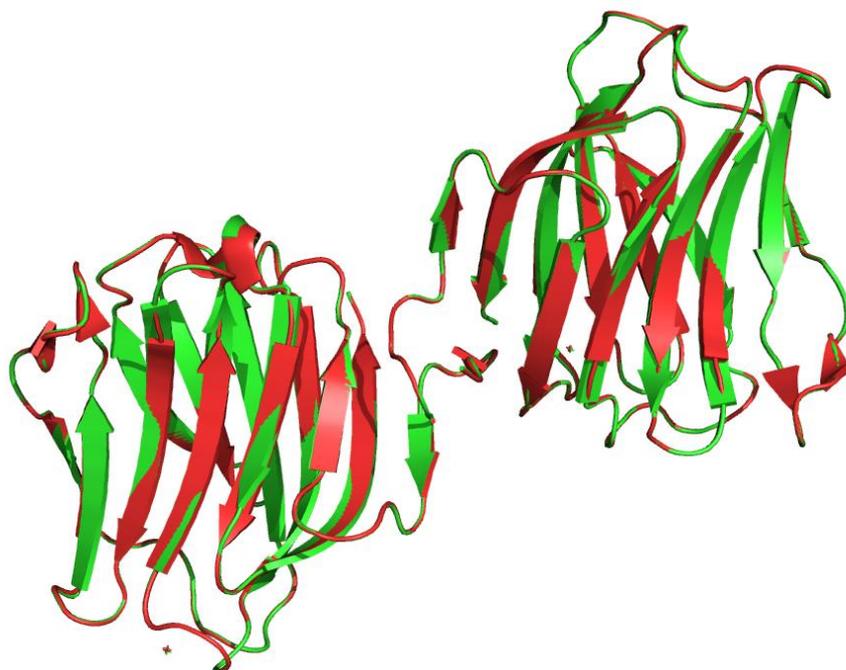


Рисунок 10. Совмещение элементов вторичной структуры моделей 3VV1 из PDB (красным цветом) и PDB\_REDO (зеленым цветом).

## Заключение

3VV1 представляет собой структуру довольно высокого качества, отличного от качества заметного большинства структур столь же высокого разрешения. Из наблюдаемых в 3VV1 аномалий можно отметить остатки с высокими значениями пространственного (*real-space R-factor*) и температурного (*B-factor*) параметра, по всей видимости, плохо разрешенными при построении модели. В структуре были также выявлены остатки со слишком сильными отклонениями длин связей и углов в остове (рис. 1), допущенные, по-видимому, вследствие ошибки расшифровки.

## Список литературы

1. Hisayoshi Makyio et al. «Structural basis of preferential binding of fucose-containing saccharide by the *Caenorhabditis elegans* galectin LEC-6» *Glycobiology*. 23.7 (2013):797-805
2. <http://www.rcsb.org/pdb/explore.do?structureId=3vv1>
3. <http://molprobity.biochem.duke.edu/>
4. <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=3VV1>
5. [http://www.cmbi.ru.nl/pdb\\_redo/uw/3vv1/](http://www.cmbi.ru.nl/pdb_redo/uw/3vv1/)