

Отчёт по качеству расшифровки структуры комплекса  
RanGDP-Nup153.

Работа была выполнена Безсудновой Ольгой (4 курс ФББ)  
2016 год

## Введение

Для данной работы был выбран комплекс транспортный Ran-белок человека с мышинным белком ядерной поры Nup-153 (PDB ID 3gj8).

В цитоплазме белок Ran- GDT не активен. Комплекс импортина и груза связывается с Ran-GDP и случайным образом проникает через пору в ядро.

Далее комплекс диссоциирует. Ran- GDT конвертируется в Ran-GTP с помощью фактора RCC1. Связавшись с пустым импортином, Ran-GTP выходит обратно в цитоплазму. Для выхода комплекс должен затратить энергию, поэтому Ran-GDT превращается в Ran- GDT. Примерно аналогичное действие происходит с белком экспортином.

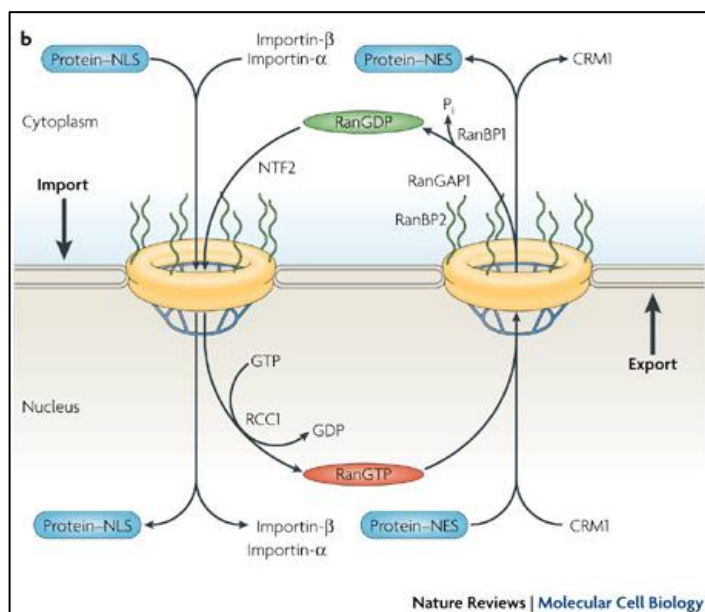


Рис.1 Схема цикла Ran-GTP – Ran-GDT[1]

Белок Nup-153 входит в состав ядерной поры. Данный белок находится на поверхности и является первым белком с которым связывается входящая молекула.

## Общая информация о модели

Представленный комплекс состоит из человеческого транспортного Ran-белка (длина полипептидной цепи – 221) и белка ядерной поры Nup153 (длина полипептидной цепи – 92). Белки были синтезированы в бактерии *Escherichia Coli*, далее комплекс из белков был закристаллизован в виде гетеродимера. Также в данном кристалле присутствуют и другие небольшие молекулы, а именно ионы металлов -  $Mg^{2+}$ ,  $Zn^{2+}$  и молекула GDP. Структура комплекса 3gj8 была расшифрована в 2009 году группой ученых под руководством Partridge, J.R., Schwartz.

Для решения фазовой проблемы был использован метод молекулярного замещения. В результате эксперимента, разрешение структуры, заявленное авторами, получилось равным 1.82 Å, а диапазон разрешений структурных факторов: 33.75 - 1.82 Å.

По данным, представленным в БД EDS (electron density server), в эксперименте было получено 53233 рефлексов. Но для дальнейшего анализа было использовано 50109(94%) структурных факторов. Таким образом, полнота данных составляет 99.2 %.

Для расшифровки структуры также важными характеристиками являются параметры ячейки. По данным БД PDB ячейка имеет следующие параметры – длины сторон:  $a=74.26$  Å,  $b=61.71$  Å,  $c=70.64$  Å; размер углов:  $\alpha=90.00$ ,  $\beta=112.31$ ,  $\gamma=90.00$ . Пространственная группа симметрии -  $P 1 2_1 1$ .

## Значения индикаторов качества модели в целом

Для того, что понять насколько хорошо была расшифрована структура, научное сообщество придумала множество индикаторов, таких как R-фактор, R\_free, ротамеры и т.д. Рассмотрим более подробно некоторые из этих характеристик структуры.

R-фактор показывает соответствие модели экспериментальным данным. При оптимизации модели данный фактор минимизируется. Чем меньше значение фактора, тем лучше. Для выбранной нами модели R-фактор составляет 0,170 (17% < 25%). Такое значение является хорошим значением.

Параметр R\_free является контролем для R-фактора. R\_free структуры равен 0.203(20,3%). Данное значение является тоже хорошим, так как меньше 25%, и меньше чем на 10% отличается от R-фактора.

Следующим важным фактором является число маргинальных остатков, определяющееся с помощью карты Рамачандрана. Для выбранного белка было найдено 9 маргинальных остатков (2.1%) по сервису EDS(Рис.2). Далее представлено распределение маргинальных остатков по цепям.

- Chain A - 4 outliers, 2.1%
- Chain B - 1 outliers, 4.0%
- Chain C - 3 outliers, 1.6%
- Chain D - 1 outliers, 4.2%

Удивительно, что, анализируя данную структура программой MolProbity, не было найдено ни одного маргинала. (Рис.3 ниже)

Поскольку процент маргинальных остатков мал, то

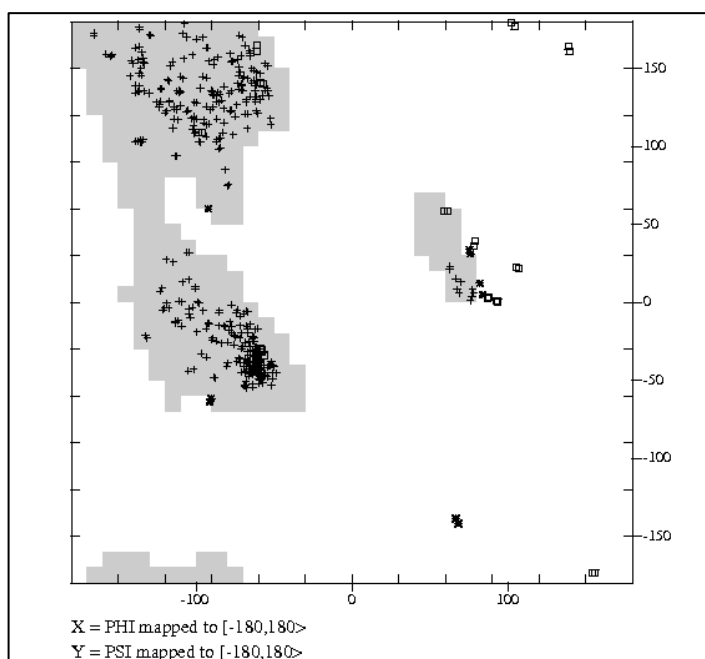


Рис.2 Карта Рамачандрана (сервис EDS)

данную структуру можно считать правильной.

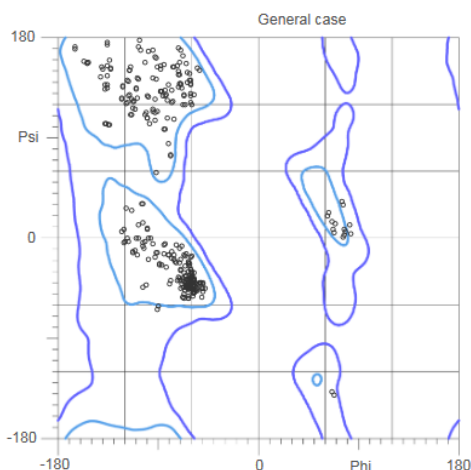


Рис.3 Карта Рамачандрана для углов psi и psi (сервис MolProbity)

Также с помощью программы MolProbity были найдены ротамеры. Ротамерами называются аминокислотные остатки, чья боковые цепи находится в типичных для данного типа остатка конформациях. Данная структура имеет 382 ротамеров (95.98 %). А остальные аминокислоты (4,02%) находятся в нестандартном расположении или на краю допустимых границ. Но только 6 (1,51%) аминокислот являются маргинальными.

Данные представлены в Таблице 1.

Таб.1 Распределение ротамеров по цепям белка (PDBeReport)

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	176/188 (94%)	176 (100%)	0	100	100
1	C	176/188 (94%)	172 (98%)	4 (2%)	58	42
2	B	23/78 (30%)	23 (100%)	0	100	100
2	D	23/78 (30%)	23 (100%)	0	100	100
All	All	398/532 (75%)	394 (99%)	4 (1%)	82	77

Другим индикатором разрешения структуры является пространственный R-фактор. Данный параметр позволяет найти маргинальных аминокислотных остатка по соответствию боковой цепи “экспериментальной” электронной плотности. Для данной модели этот фактор имеет значение 0.109. Данное значение находится в границах допустимых значений. Для детального рассмотрения, как же вписана аминокислотная последовательность в электронную плотность, ниже представлены Рис.4 распределения значений пространственной R-фактора по всей цепи.

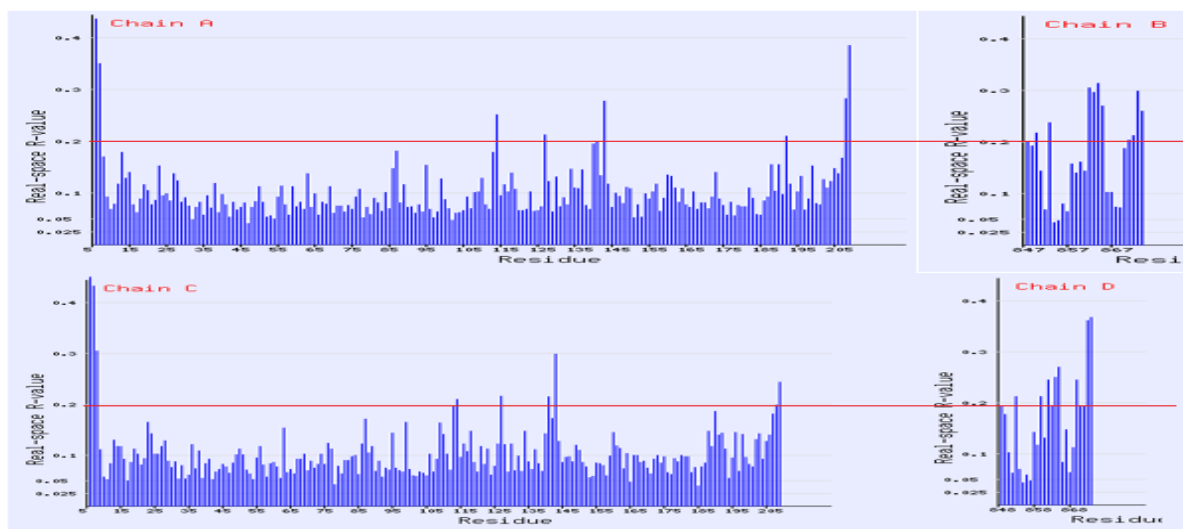


Рис 4. Распределение значений пространственного R-фактора (Красная линия - 20%)

Как видно из рисунка, цепи А и С вписаны очень хорошо в электронную плотность, а цепи В и D наоборот не очень хорошо вписаны, скорее всего это связано с тем, что данные цепи маленькие и находятся на поверхности белка.

Существует относительная оценка RSR, называемая RSR – Z-score. Высокие положительные значения  $Z > 2$  свидетельствуют о том, что остаток плохо вписан в электронную плотность по сравнению с другими структурами с тем же разрешением. Данная структура имеет значение равное -0.11. Распределение по полипептидной цепи представлено на Рис.5.

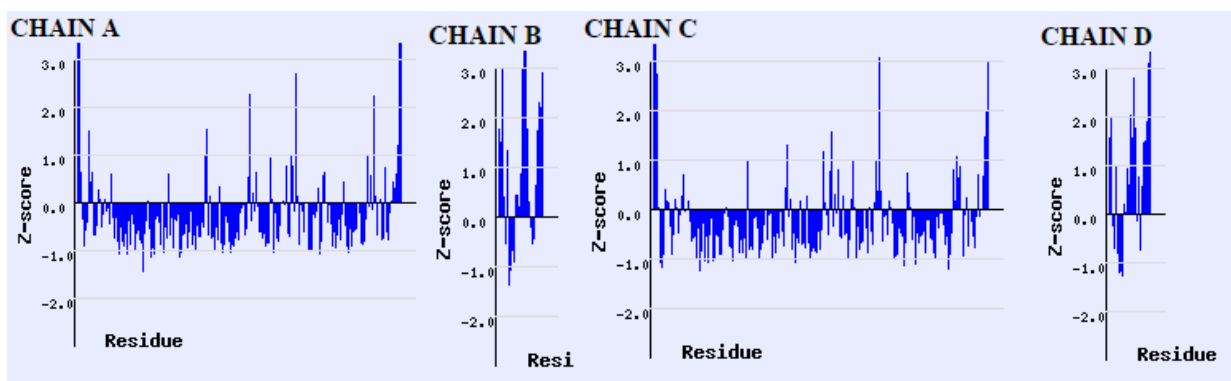


Рис.5 Распределение RSR-Z по аминокислотной цепи (сервис EDS)

Следующим важным фактором является комфортность окружения каждого аминокислотного остатка. В программе WhatCheck рассчитывается Z-score для комфортности окружения каждой боковой цепи. Результаты представлены ниже. Из этих данных можно сделать вывод, что в целом все аминокислотные остатки находятся в комфортном для себя окружении.

All contacts : Average = -0.296 Z-score = -1.96

Интересно заметить, что некоторые аминокислоты нуждаются в инверсии, чтобы лучше вписаться в электронную плотность или уменьшить напряжение в цепи. В таблице 2 представлен список таких аминокислот.

Таб.2 Список аминокислот, нуждающихся в инверсии

Mol	Chain	Res	Type
1	A	48	HIS
1	A	100	ASN
1	A	103	ASN
1	A	139	HIS
1	C	100	ASN
1	C	103	ASN
1	C	139	HIS

Последний фактор, о котором будет упомянуто называется B-фактор или температурный фактор. В гармоническом приближении тепловой фактор учитывает среднеквадратичные отклонения каждого атома от своего положения равновесия. Диапазон значения B-фактора от 10 до 30.

На Рис.6 представлено изображение комплекса, раскрашенному по распределению B-фактора. Данное распределение является типичным для белков.

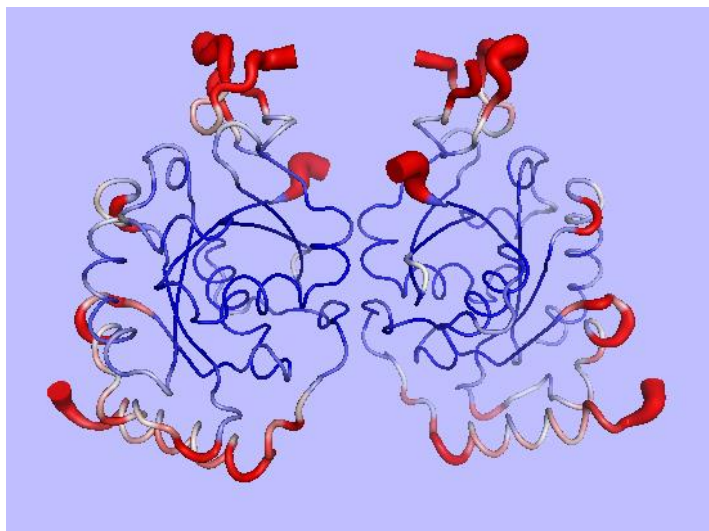


Рис.6 Изображение комплекса раскрашенного по B-factor.

## Маргинальные остатки

Для наглядного представления маргинальности аминокислотных остатков представлен Рис.7 из PDB-Report. На рисунке зеленым представлены остатки, подходящие под все геометрические критерии, желтыми – остатки, не подходящие под один критерий, оранжевыми – под два критерия, красными – три и более. Красные точки показывают, что данные остатки плохо вписаны в электронную плотность, серым – аминокислоты, представленные только в структуре. Ниже представлена таблица, в которой представлена информация о некоторых маргинальных остатках.

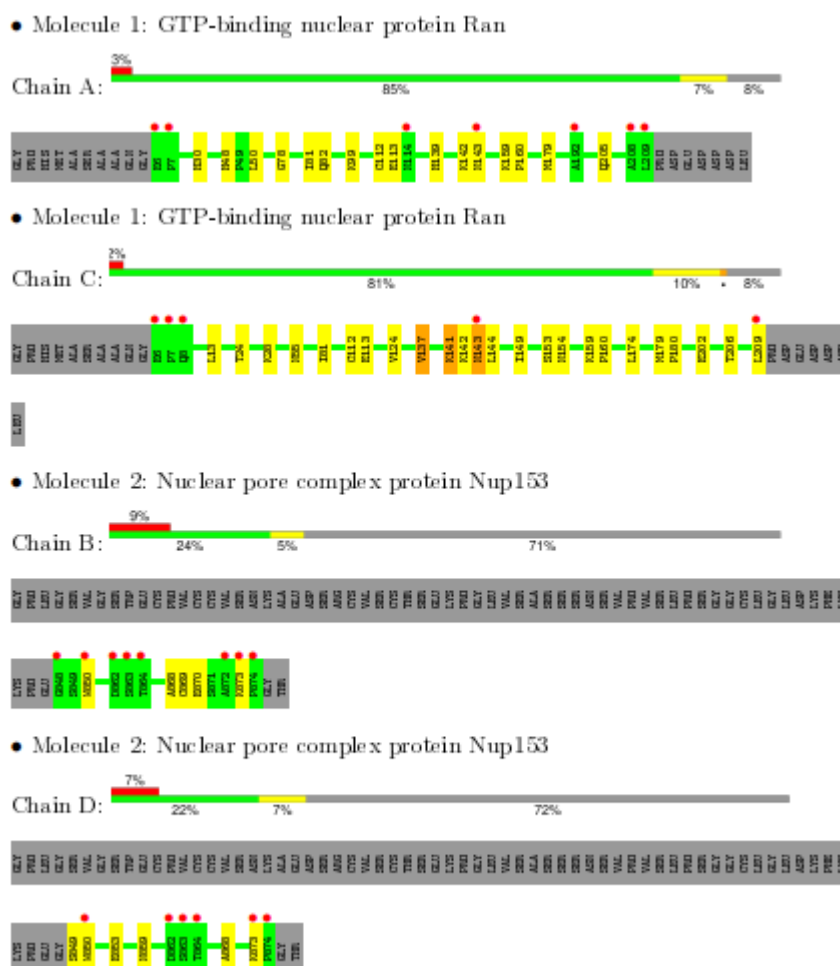


Рис.7 Аминокислотная последовательность белка, раскрашенная по маргинальности



№	Остаток	Цепь	Причина маргинальности
1	Val 137	C	Не является ротамером; и т.д.
2	Lys 141	C	Не является ротамером; Участвует в неблагоприятных перекрываниях атомов;
3	Asp 143	C	Не является ротамером; Маргинальный по значению RSRZ; Неблагоприятных перекрываниях атомов;
4	His 48	A	Нуждается в инверсии
5	Leu 209	A	Маргинальный по значению RSRZ
6	Ser 863	B	Плохо вписан в электронную плотность
7	Trp 850	B	Участвует в неблагоприятных перекрываниях атомов;
8	Ala 872	B	Плохо вписан в электронную плотность
9	Ala 868	D	Участвует в неблагоприятных перекрываниях атомов
10	Pro 874	D	Маргинальный по значению RSRZ

**Анализ 5 маргинальных остатков и/или гетеромолекул (воды, ионов, лигандов).**

Из данных представленных выше, в данной структуре есть 3 остатка маргинальных по нескольким параметрам: V(137), K(141), N(143)

## Asp 143

Остаток (143) также является маргиналом по нескольким параметрам. Данный остаток находится в конце право-закрученной альфа-спирали, прям на изгибе полипептидной цепи (Рис.8). Данная аминокислота является маргиналом по нескольким признакам. Во-первых, данная аминокислота плохо вписана в электронную плотность, точнее для нее нет электронной плотности, что видно на Рис.9Б. Во-вторых, остаток не является ротамером, так как  $\chi_1$ -угол равен - 220.6,258, что не является нетипичным для Asp. Так же данный остаток имеет атом  $C\beta$ , который имеет неблагоприятное перекрытие с соседним атомом O из аминокислоты Lys (142). Данное несоответствие показано на Рис.9А. Из выше представленных данных можно сделать вывод, что данный остаток является ошибкой расшифровки.

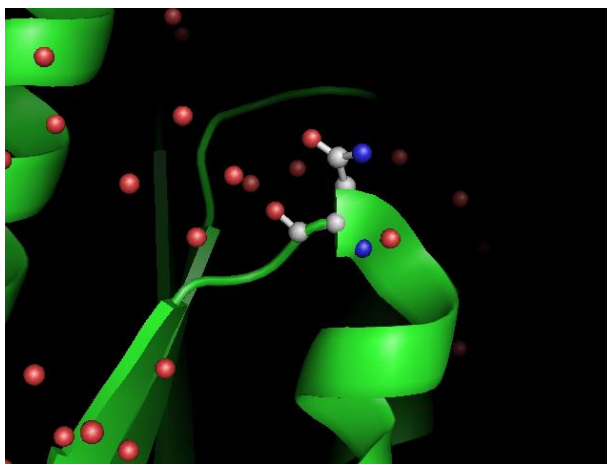


Рис.8 Изображение Asp 143

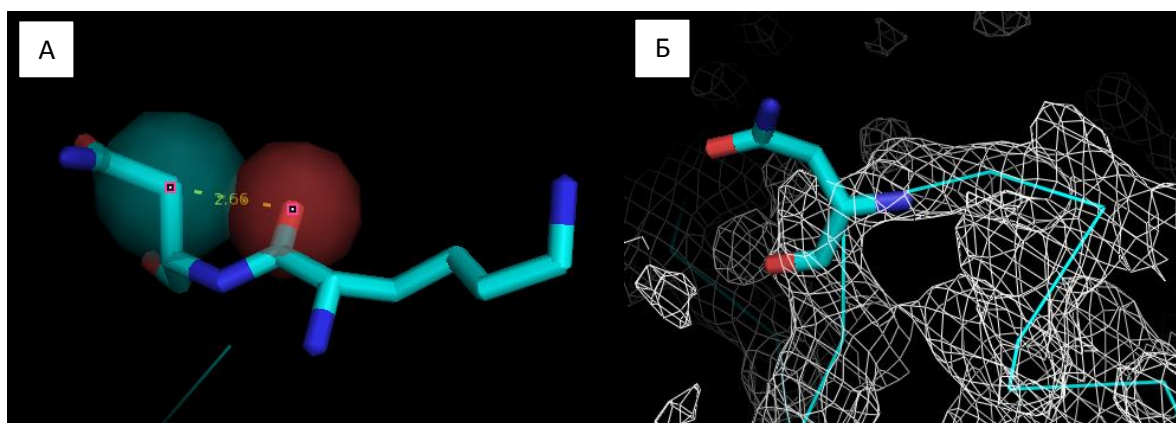


Рис.9 А) Неблагоприятное перекрытие атомов Б) Изображение электронной плотности

## Val 137

Несмотря на то, что остаток Val(137) является гидрофобным, он находится на поверхности молекулы белка и соприкасается с водой (Рис.10). Аминокислота находится в альфа-спирале. С помощью окрашивания по B-factor (Рис.11А) можно предположить, что данный остаток является мало подвижным.

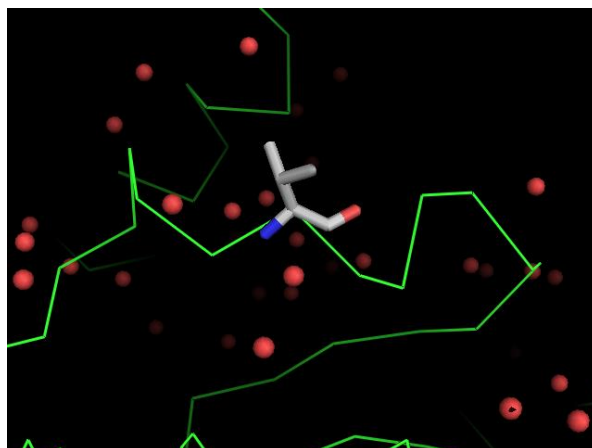


Рис.10 Изображение остатка Val. Красными кругами показана вода.

Аминокислота не является ротамером, так как имеет  $\chi_1$  угол - 133.3, не типичный для валина. Но данный остаток хорошо вписан в электронную плотность. (Рис.11Б) Также остаток имеет неблагоприятное перекрытие с соседним атомом Н из аминокислоты Lys (141) и атомом N из аминокислоты Lys (141). Данный остаток не является маргинальным по RSRZ. С моей точки зрения, это правильно вписанный остаток, являющийся особенностью.

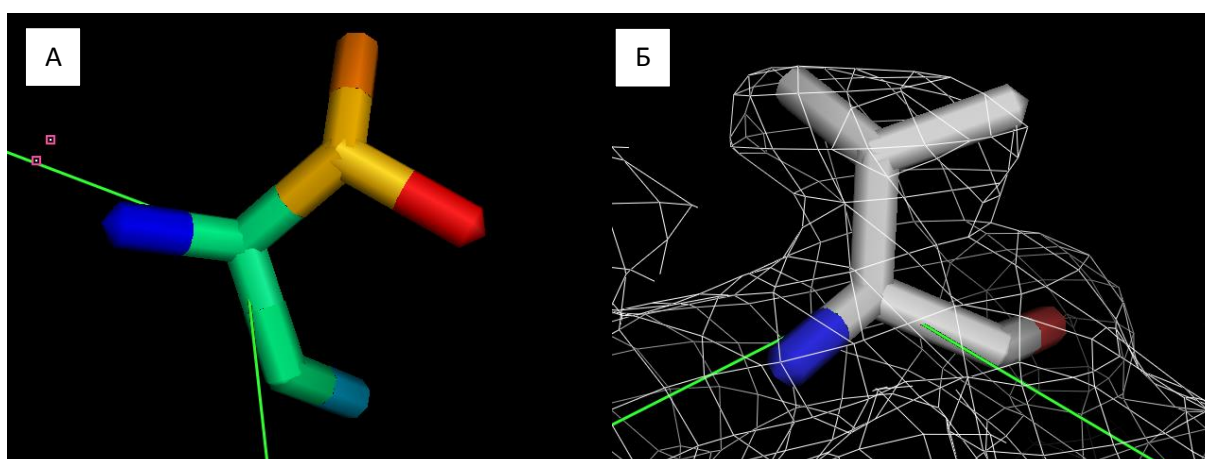


Рис.11 А) Раскраска Val по температурному фактору Б) Изображение электронной плотности

## Lys 141

Данный остаток хорошо вписан в электронную плотность. (Рис.12 ниже), но из-за этого возникает множественное перекрытие атомных радиусов, а также задевает атомы соседних аминокислот. (См таблицу 3)

Таб.3 Данные о неблагоприятных перекрытиях аминокислотного остатка Lys 141

Атом-1	Атом-2	Расстояние	Перекрытие
Lys-141-HB2	Lys-141-NZ	1.99	0.77
Lys-141-HZ1	Val-137-HG13	1.51	0.75
Lys-141-HZ	Val-137-HG13	2.12	0.64
Lys-141-HB2	Lys-141-HZ2	1.67	0.59
Lys-141-HZ2	Lys-141-CB	2.28	0.46
Lys-141-CB	Lys-141-NZ	2.76	0.45

Чтобы подтвердить предположение, что данная аминокислота является особенностью данного белка, было просмотрено НММ logo семейства Ran-белков. Оказалось, что в гомологичных белках лизин не находится в этом месте.(Рис.13)

```
#HWI K1vlvGdsavGKtslllrfvkdKfkesyebTigwdfjtkelavdgkewkleinDTAGaeefksrlrelvrdagillvyditsresfenkkwveelrvaaenpivlvGnkVDleekravsteegeelakelglkflletSAktneveeafeelrsei
#MATCH K1vlvGd+++GKt+++G+ ++++++T+gv+ ++++++k+++DTAGae+f lr+yy ++++++d+tsr ++++++rvae n+pivlvGnkVD+++ g+vv+++ ++++++1++++SAK+n+nve+ f 1+r+++
#PP 90*****
#SEQ KLVLVGdGGTgKlTFvKRHLTGESEKkyWATLVGEVPLVPHHTNRGPIKFMVMDTAGQEKFGGLRDGYIQAQCALIHFVDTSRVTVKINPNNRDLVRVCE-HPVILCGNKVDIID-IRVKAASIV-FHRKKNLQYDISAKSNVHFEKPLWLRKIL
```

Рис.13 Сравнение НММ и последовательности белка

Данный остаток не является маргинальным по RSRZ, но имеет не типичных торсионных углов ( $\chi_1$  углы: 218.3,76.5,56.7,289) . Подводя итоги, для данной аминокислота скорее всего была плохо получена электронная плотность, поэтому ее так и изображали.

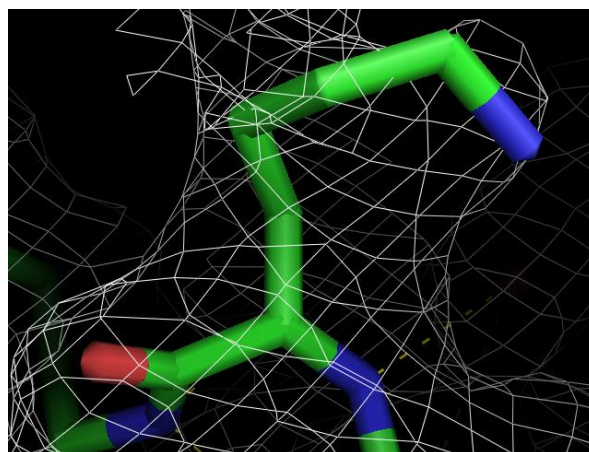


Рис.12 Скрученный Lys, вписанный в электронную плотность

## Участок цепи С

Рассмотрим участок из аминокислотных остатков 137-143(VFHRKKNL) цепочки С. На Рис.13 изображены водородные связи, которые являются нетипичными для альфа-спирали. Можно заметить, что 5 остаток из 7 не являются ротамерами. Также часто происходит неблагоприятное перекрытие атомов, а остаток His должен быть подвержен инверсии. Скорее всего, данная альфа-спираль является супер-скрученной.

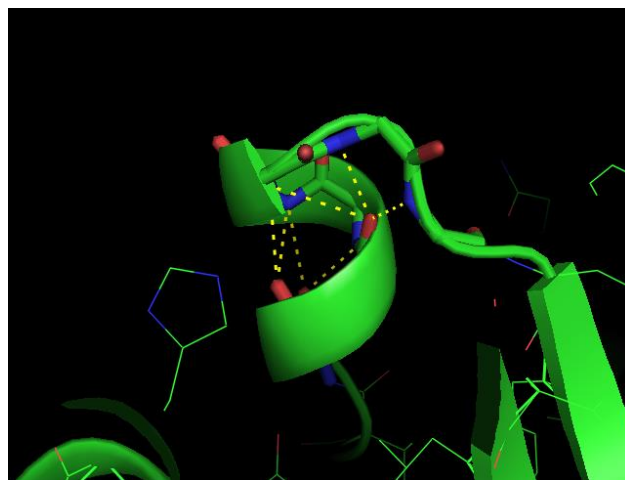


Рис. 13 Альфа спираль из остатков 137-143

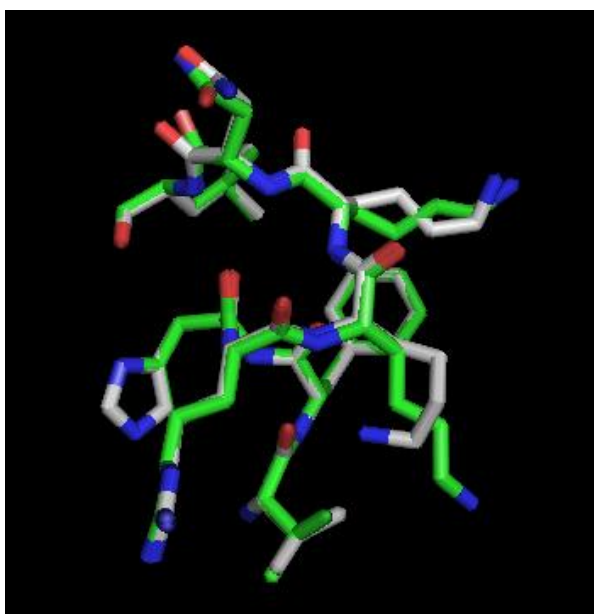


Рис. 14 Альфа спираль из остатков 137-143. Оптимизированная структура изображена зеленым, белым – исходная.

С помощью программы PDB\_redo, была получена оптимизация. Результат оптимизации можно увидеть на Рис.14. После оптимизации аминокислоты приняли комфортные для себя положения, но не вписывающиеся в электронную плотность. Скорее всего данная спираль является особенностью данного комплекса, возможно возникает при соприкосновении двух белков.

## GPD

Выбранный мною белок является GDP связываемым белком. Расшифрованная с помощью PCA электронная плотность дает четкое представление куда именно будет вписан молекула GDP. (Рис. 15)

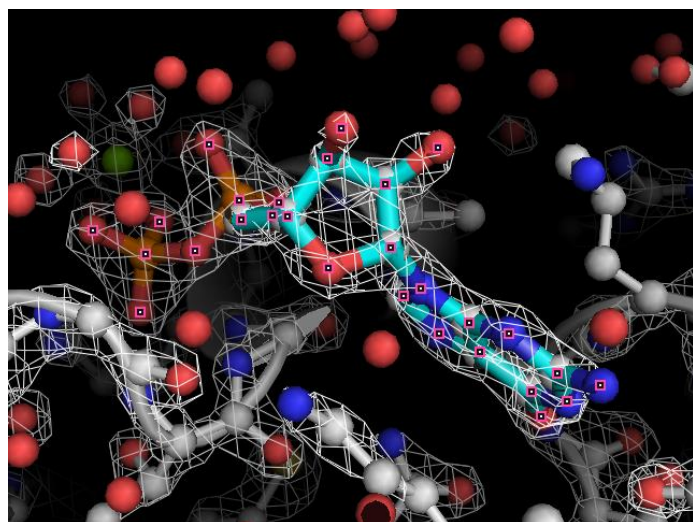


Рис. 15 Молекула GDP, вписанная в электронную плотность.

Видно из Рис.16 данная молекула крепко связывается с белком, а также с металлом магния. Также GDP связывается в месте скопления полярных частиц, что говорит о прочном связывании молекулы и белку.

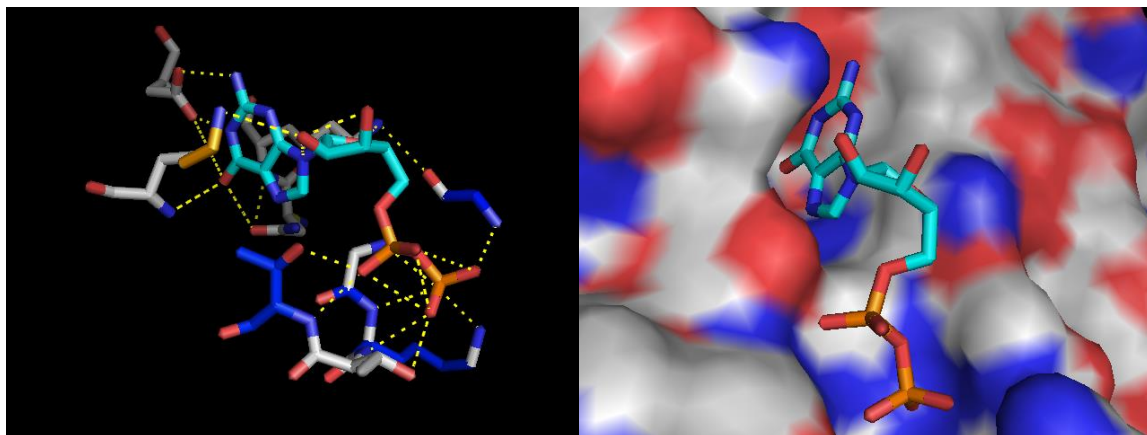


Рис. 16 Молекула GDP, связанная с белком

Для молекулы GPD показан Z-score для длины связей между атомами плохой только 1(4%) случаях, а для углов в 6(30%) случаях. Скорее всего такие изменения происходят из-за возникших связей между белком и молекулой. Подробные данные представлены в таблице 3.

Таб. 3 Данные о маргинальных значениях в молекуле GDP

Atoms	Z	Observed(A)	Ideal(A)
C5-C4	3.10	1.47	1.40
C6-C5	3.64	1.48	1.41
Atoms	Z	Observed()	Ideal(A)
C5-C6-N1	-4.14	117.93	123.59
C5-C6-C4	-3.36	116.88	120.90
PA-O3A-PB	-2.58	124.02	132.67
C6-N1-C2	4.89	122.73	115.94
N3-C2-N1	-3.09	122.74	127.4
C4-C5-N7	-3.13	106.60	109.48

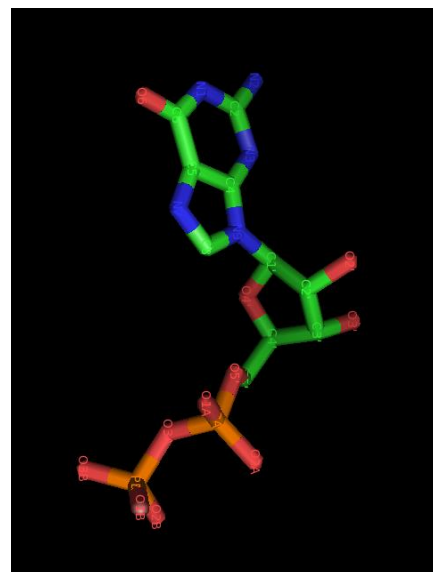


Рис.17 Молекула GDP с обозначенными атомами

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

Далее был запущен сервис PDB\_redo для оптимизации модели. В результате, индикаторы модели получились следующие. (они представлены в таблице 4). Таким образом, произошла оптимизация, но не значительная.

Таб.4 Индикаторы модели до и после оптимизации

	From PDB header	Calculated from data	After re-refinement	After re-refinement and rebuilding
R	0.1700	0.1684	0.1653	0.1651
R-free	0.2030	0.1997	0.1946	0.1916
$\sigma$ R-free		0.0039	0.0038	0.0037
Z(R-free)		-0.03	0.37	1.11

На Рис.18 показано пространственное совмещение двух моделей в двух видах визуализации. Видно, что модели в целом сходны и повторяют одна другую, а отличия заключаются в мелких поворотах боковых, но было произведено 25 изменений в конформациях аминокислот.

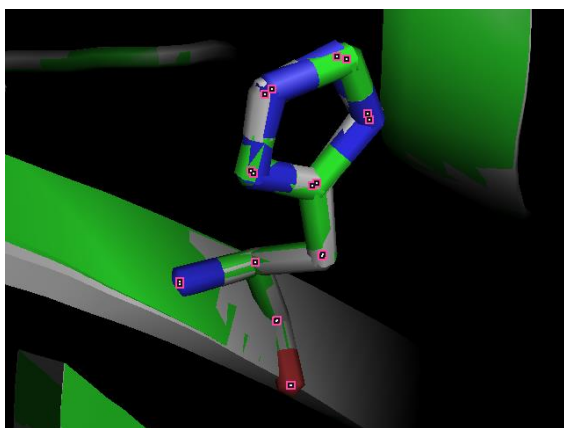


Рис. 19 His 48. Оптимизированная структура изображена зеленым, белым – исходная.

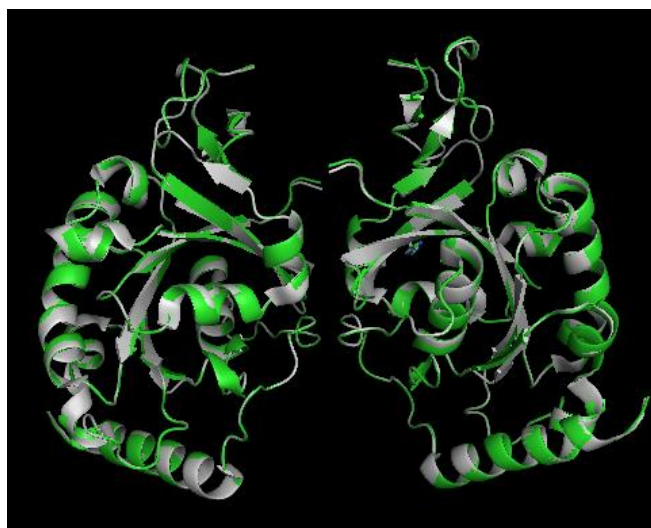


Рис. 18 Совмещение оптимизированной и исходной структуры. Оптимизированная структура изображена зеленым, белым – исходная.

Значительное изменение пептидной последовательности представлено на Рис. 13. Данная программа также произвела нужные инверсии, один из примеров приведен на рис.19.



## **Вывод**

Из выше представленной информации можно сделать вывод о качестве расшифровки структуры, а именно предсказано данная структура хорошо. Основные параметры находятся в диапазоне допустимых значение или даже и имеет хорошие значения. Маргинальных остатков немного, скорее всего они являются особенностями данного комплекса. При оптимизации не произошло качественно улучшения параметров структуры. Подводя итоги, расшифровка данной структуры хорошая.