

Отчет о расшифровке структуры человеческого
белка 1 гистидиновой триады (PDB код 3TW2)
методом рентгеноструктурного анализа

Студентки четвертого курса факультета биоинженерии и
биоинформатики МГУ им. М. В. Ломоносова
Тишиной Софии

Москва, 2015

Аннотация

В ходе выполнения отчета более близко познакомились с серверами, которые предоставляют возможность оценить качество белковой структуры.

Введение

Белок 1 гистидиновой триады или HINT1 играет важную роль в различных биологических процессах и есть у многих живых существ, у животных, растений, грибов и бактерий. HINT1 принадлежит к семейству гидролаз (ЕС 3.-.-.-) [1]. Сначала белок 1 был найден как белок, связывающийся с протеин киназой С, тем самым ингибируя ее. Позже было выяснено, что он может связывать и производные аминокислот, например, АМФ, и нуклеотидную последовательность, тем самым участвуя в регуляции транскрипции. Причем, по данным Uniprot у HINT1 есть несколько нуклеотид связывающих регионов – в один в начале последовательности и два в конце [2].

Таким образом, у HINT1 есть как минимум три молекулярные функции: гидролазная активность, нуклеотид-связывающая и регуляция работы киназы С.

Белок 1 является гомодимером, каждая цепь в котором состоит из 5 антипараллельных бета-листов и 2 альфа-спиралей (рис. 1.1). Как видно на рисунке, более короткая альфа-спираль (1 спираль) локализована ближе к N-концу, в то время как большая (2 спираль) находится между 3 и 5 бета-листом. Нуклеотид-связывающий карман сформирован 1, 3 и 6 петлями спирали1 и фрагментом бета-листа (рис. 1.2).



Рис. 1. Структура димера человеческого HINT1.

Первая структура белка 1 была получена в 1996 году [3], позже получали еще структуры белка из человека, кролика и рекомбинантный из E.Coli. В итоге на сегодняшний момент в Uniprot есть 10 структур с разрешением от 1.38 Å до 2 Å, причем это наибольшее разрешение принадлежит рассматриваемой в данной работе структуре 3TW2 (до нее была получена структура с разрешением 1.50 Å, но только одной цепью). 3TW2 – это структура белка вместе со связанным АМФ (рис. 1).

Результаты и обсуждение

Общая информация о модели и значения индикаторов качества модели в целом

Как говорилось ранее, модель белка 1 гистидиновой триады 3TW2 представляет собой гомодимер из двух цепей (А и В), которые являются биологически активной единицей и каждая цепь состоит из 126 аминокислотных остатков. Эта модель структуры была получена относительно недавно, в 2011 году группой исследователей из Франции [4].

Авторы для построения модели воспользовались методом молекулярного замещения (MR) с использованием полученной ранее модели 1КРФ. Разрешение структуры - 1.38 Å (при полноте 99 %). Такое разрешение является почти атомарным (рис. 2). В ходе эксперимента были измерены 46352 рефлекса, у 92% из которых значение отклонения превышает стандартное больше или в 3 раза (см файл ref.xlsx). Общая информация о модели представлена в табл. 1.

Табл.1. Общая информация о модели 3TW2. Данные взяты из PDB файла.

Разрешение	1.38 Å
Минимальное и максимальное разрешение	25.58 Å и 1.38 Å
Полнота	99.0%
Число померенных рефлексов	46352
Число рефлексов с силой сигнала $\geq 3\sigma$	42605
Число рефлексов используемых для построения модели	43580

В кристаллографической ячейке лежит одна глобула белка (состоящая из двух цепей). Параметры ячейки указаны в табл. 2. Данная ячейка

ассиметрична, и в данных не указано никаких других симметрий, кроме кристаллографической, матрица симметрии указана в табл. 3. В файле 3TW2 лежит 2199 атома, 1767 из которых принадлежат белку.

Табл. 2. Параметры кристаллической ячейки. Данные взяты из cif файла.

a	77.4213 Å
b	46.4489 Å
c	64.0341 Å
alpha	90.0000°
beta	94.4227°
gamma	90.0000°

Табл.3. Матрица кристаллографической симметрии. Данные взяты из PDB файла.

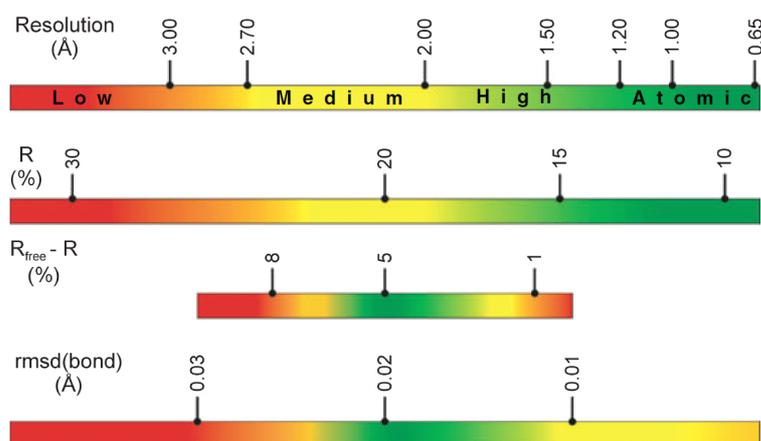
```

REMARK 290  SMTRY1  1  1.000000  0.000000  0.000000  0.000000
REMARK 290  SMTRY2  1  0.000000  1.000000  0.000000  0.000000
REMARK 290  SMTRY3  1  0.000000  0.000000  1.000000  0.000000
REMARK 290  SMTRY1  2 -1.000000  0.000000  0.000000  0.000000
REMARK 290  SMTRY2  2  0.000000  1.000000  0.000000  0.000000
REMARK 290  SMTRY3  2  0.000000  0.000000 -1.000000  0.000000
REMARK 290  SMTRY1  3  1.000000  0.000000  0.000000  38.71050
REMARK 290  SMTRY2  3  0.000000  1.000000  0.000000  23.22450
REMARK 290  SMTRY3  3  0.000000  0.000000  1.000000  0.000000

```

REMARK 290 SMTRY1 4 -1.000000 0.000000 0.000000 38.71050
 REMARK 290 SMTRY2 4 0.000000 1.000000 0.000000 23.22450
 REMARK 290 SMTRY3 4 0.000000 0.000000 -1.000000 0.00000

Качество модели в целом можно оценивать и по отличным от разрешения при данной полноте характеристикам. Так, например, значение R-фактора равняется 0.1207, а R_{free} равно 0.1615, следовательно разность между ними составляет $0.0408=4.08\%$, что лежит в области хороших значений (рис. 2).



From: Wlodawer *et al*, FEBS Journal 275 (2008) 1–21

Рис. 2. Ориентир при оценке качества структуры [5].

Другим показателем качества рентгеноструктурной модели является карта Рамачандра. Она показывает предпочитаемые, допустимые и недопустимые положения пар торсионных углов ϕ/ψ для остатков полипептидной цепи. Запустив сервер MolProbity получили карту Рамачандра (рис. 3) и сводную таблицу с анализом геометрии остатков структуры (табл. 4) [6].

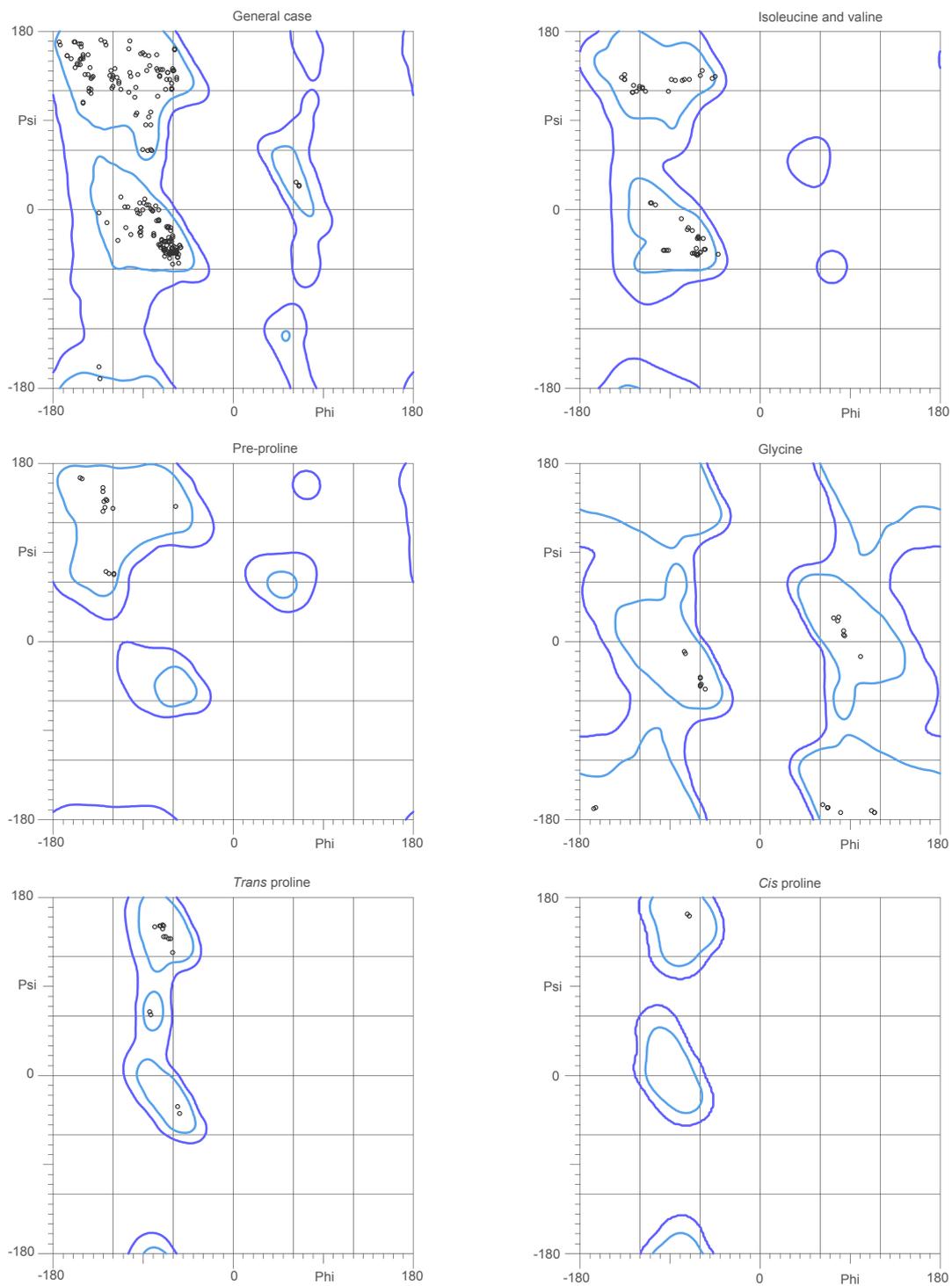


Рис. 3. Карта Рамачандра полученная с помощью сервера MolProbity для различных остатков [6]. На этой карте 99.2% остатков лежат в предпочтительных областях и 100% в разрешенной области.

Таблица 4. Сводную таблицу с анализом геометрии остатков структуры

Protein Geometry	Poor rotamers	2	0.88%
	Favored rotamers	219	96.48%
	Ramachandran outliers	0	0.00%
	Ramachandran favored	259	99.23%
	C β deviations >0.25Å	0	0.00%
	Bad bonds:	0 / 2148	0.00%
	Bad angles:	0 / 2950	0.00%
Peptide Omegas	Cis Prolines:	2 / 13	15.38%

Из карты Рамачандра (рис. 3) и таблицы 4 видно, что 99.2% остатков лежат в предпочтительных областях и 100% в разрешенных. Другими интересными параметрами качества структуры являются отклонения от теоретических значений положения бета атомов углерода, ковалентных связей и углов. Отклонений бета атомов углерода больше 0.25 ангстрема от допустимого нет, как нет и отклоняющихся от теоретических ковалентных связей и углов [7]. Но есть 2 маргинальных остатка, ими являются 2 цис пролина. Цис пролины редко встречаются в белковых структурах, но вполне возможно, что они действительно такие, и это не ошибка расшифровки структуры, потому что пролин часто встречается в поворотах и иногда просто необходимо иметь цис конформацию (рис. 4). Проанализируем их чуть позже.

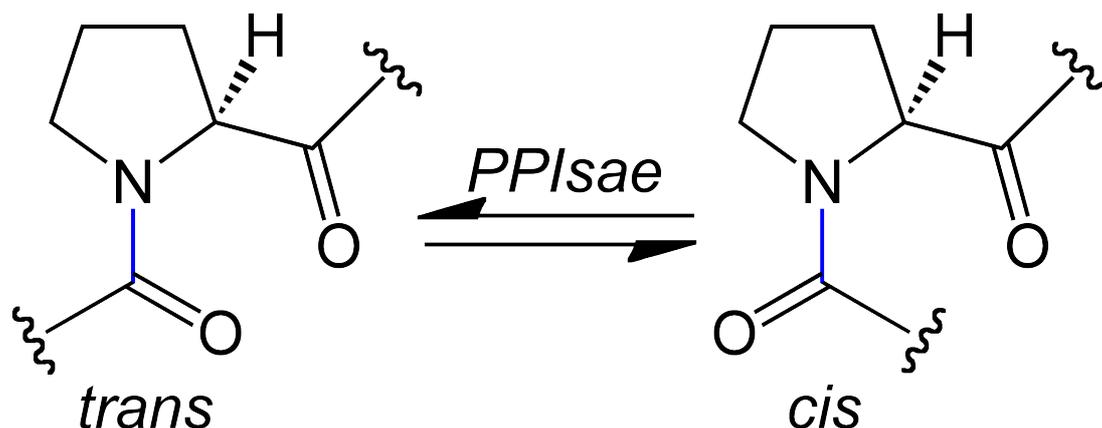


Рис.4. *Trans* и *cis* конформации пролина [8].

База данных PDBREPORT позволяет искать еще больше информации о аномалиях в структуре [9]. С ее помощью нашли некоторые остатки, у

которых неблагоприятное окружение: Arg2, His112, Gln37 и некоторые другие. Также нашлась молекула воды не фиксированная водородными связями (232). И некоторые другие в основном мелкие несуразности.

Маргинальные остатки

По карте Рамачадрана, построенной с помощью сервиса MolProbity (рис. 3) нет маргинальных остатков, но в таблице 4 мы видим, что есть 2 пролина имеющие cis конформацию. База данных позволяет нам определить какие это остатки, это 2*Pro125. Так же находит еще в дополнение к ним два глицина (Gln62), которые тоже имеют необычные значения торсионных углов. Данная структура является очень хорошей исходя из описанных выше параметров, поэтому у нее немного маргинальных остатков.

Используя данные полученные сервером PDBREPORT сделаем таблицу с маргинальными остатками (табл. 5).

Таблица 5. Маргинальные остатки [9].

Маргинальные остатки, цепь	В чем маргинальность	Возможная причина этому
Pro125, A и B	cis конформация	Пролины находятся на краю в области петель, предположительно активный центр (рис. 4)
Gln62, A и B	Редкий торсионный угол: z-score=-2.1 (что меньше порога в -2.0)	Глицин находится в повороте и на поверхности, возможно причина в этом. Но редкий торсионный угол для глицина? Сомнительно. Что-то тут не так. Может PDBREPORT слишком жесткий порог Z-score поставил? (рис. 5)

His122, A и B	БД считает, что если «повернуть кольцо» можно больше образовать Н-связей	Теперь еще больше похоже на активный центр (рис. 6)
His, B	БД считает, что если «повернуть кольцо» можно больше образовать Н-связей	Почти наверняка ошибка в структуре (рис. 7)

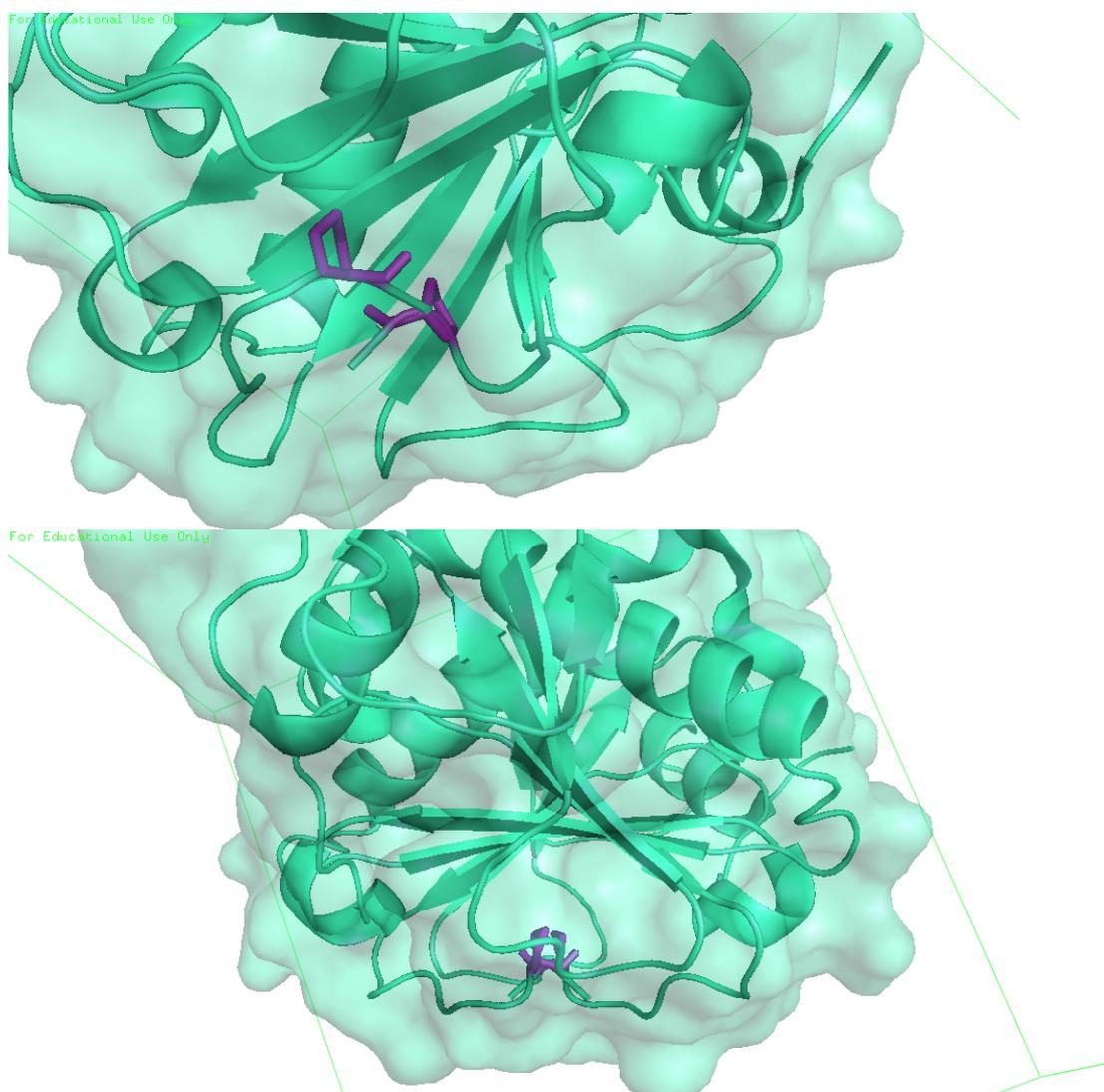


Рис. 4. Маргинальные пролины (фиолетовые), находящиеся в

неупорядоченном участке на краю цепи. На нижнем рисунке видно, что участок, где находятся эти пролины похож на активный центр (с одной стороны, вероятно способные к движению, петли, с другой поддерживающие структуру белка бета-листы, к тому же часто встречается активный центр в месте контакта двух субъединиц белка, если белок работает в димерном состоянии), и тогда вполне возможна нестандартная конформация остатков, чтобы обеспечить функцию белка.

For Educational Use Only

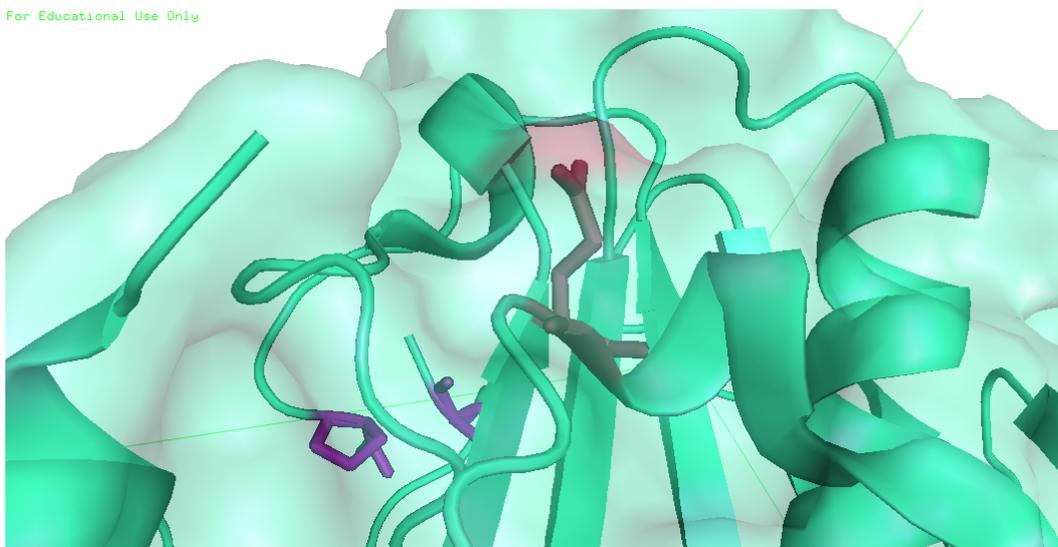


Рис. 5. Один из маргинальных глицинов (красный; второй аналогично в другой цепи). Глицин находится в повороте и его поверхность контактов выступает в растворитель.

For Educational Use Only

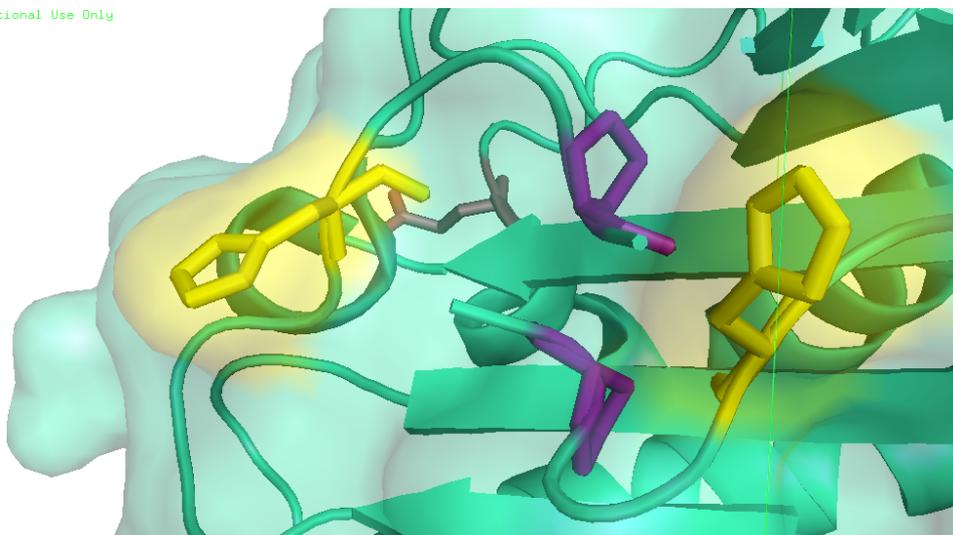


Рис. 6. Маргинальные гистидины рядом с пролинами (желтые). Очень похоже на активный центр.

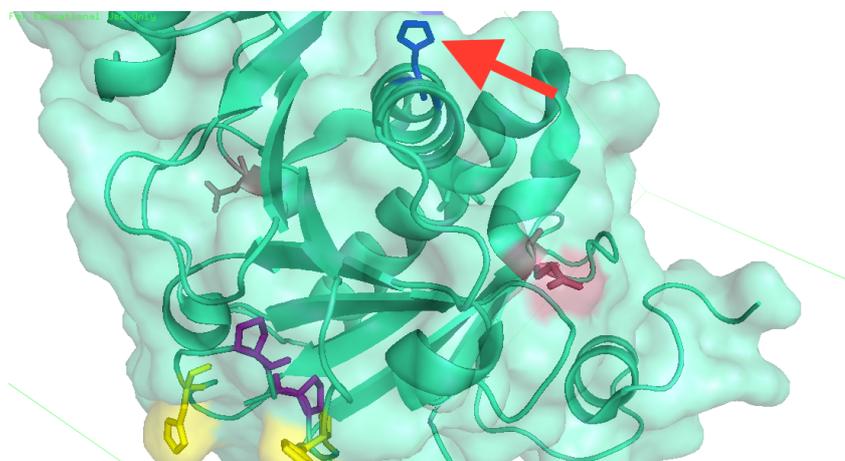


Рис. 7. Маргинальный 76 гистидин (синий). Не похоже, что он играет какую-то особенную роль в белковой цепи (кроме участия в формировании альфа-спирали), поэтому скорее всего просто погрешность в структуре и его надо «чуть повернуть» чтобы было больше выгодных контактов.

Заключение

Из всего вышесказанного следует, что структура расшифрована качественно. Об этом свидетельствует и почти атомарное разрешение (1.38 Å) при отличной полноте (99%) и хорошие показатели R фактора, R-free, а также малое число маргинальных остатков, причем те, что нашлись похоже являются не ошибками расшифровки структуры, а выполняют важную роль в функционировании белка.

Список литературы и источников информации

1. http://www.genome.jp/dbget-bin/www_bget?pon:100171549

2. <http://www.uniprot.org/uniprot/P49773>
3. Lima, C. D., Klein, M. G., Weinstein, I. B. & Hendrickson, W. A. (1996). Proc. Natl Acad. Sci. USA, 93, 5357–5362
4. Dolot, Rafał et al. “A New Crystal Form of Human Histidine Triad Nucleotide-Binding Protein 1 (hHINT1) in Complex with Adenosine 5'-Monophosphate at 1.38 Å Resolution.” Acta Crystallographica Section F: Structural Biology and Crystallization Communications 68.Pt 8 (2012): 883–888.
5. Лекция 3. В.Ю. Лунин
6. <http://molprobity.biochem.duke.edu/data/cves5lrg672006jf4f2u3ikm92/charts/3tw2-rama.pdf>
7. http://molprobity.biochem.duke.edu/viewtable.php?MolProbSID=cves5lrg672006jf4f2u3ikm92&file=/home/rlab/Desktop/MolProbities/MolProbity150915_42refresh/public_html/data/cves5lrg672006jf4f2u3ikm92/raw_data/3tw2-multi.table
8. https://commons.wikimedia.org/wiki/File:Prolyl_isomerisation_General_Formulae_V.2.png
9. <http://www.cmbi.ru.nl/pdbreport/cgi-bin/nonotes>