Критический анализ модели белка изоцитрат дегидрогеназы *Bacillus subtilis*, представленной в банке PDB, код 1HQS.

*Маслова А.О., 4 курс ФББ МГУ*

В отчете приведены результаты анализа качества структуры изоцитрат дегидрогеназы *Bacillus subtilis*, расшифрованной методом РСА разрешением 1,55 Å в 2001 году Сатиндером Сингхом с коллегами и содержащимся в PDB под кодом 1HQS.

**Введение**

Изоцитрат дегидрогеназа *Bacillus subtilis* (BcIDH) входит в семейство металлзависимых декарбоксилирующих дегидрогеназ. Ее кристаллическа структура была расшифрована при разрешении 1,55 Å с помощью РСА и детально сравнена с аналогичной дегидрогеназой из *Escherichia coli* (EcIDH). Хотя эти белки очень похожи, у них есть три важных отличия: неполярные β-структуры и 2 связывающих петли в малом домене EcIDH замещены 2-мя полярными α-спиралями в BcIDH; у BcIDH есть так называемая фосфорилирующая петля, которая сужает вход в активный центр с 12 до 4 Å; хотя BcIDH и гомодимер, но субъединицы не являются абсолютно идентичными. Структурные различие между BcIDH и ЕсIDH могут объяснить различия работы белков как субстратов для IDH киназы фосфорилазы (IDH-K/P). Есть несколько гипотез. Доступ к сайту фосфориляции у BcIDH гораздо белее затруднен, чем у ЕсIDH. Второе возможное объяснение пониженной афинности IDH-K/P к BcIDH состоит в том, что внутренний участок BcIDH менее подвижен. В активном сайте много неполярных аминокислот, которые неспособны образовать водородные связи. В третьих домен ЕсIDH очень хорошо подходит нароль сайта докинга для IDH-K/P в отличие от внутреннего региона BcIDH.

Ссылка на статью:

<http://www.jbc.org/content/276/28/26154.full.pdf+html>

**Результаты**

* Комплекс состоит из цепей А и В длиной 423 а.о. В составе имеются следующие лиганды: лимонная кислота, S-1,2-папандиол и R-1,2-пропандиол. Также есть модифицированный аминокислотный остаток СМЕ (S,S-2-гидроксиэтилтиоцистеин);
* Структура опубликована 25.07.2001, J.Biol.Chem.;
* Авторы: Singh, S.K., Matsuno, K., LaPorte, D.C., Banaszak, L.J.;
* Метод решения фазовой проблемы: молекулярное замещение;
* Число измеренных рефлексов: 114794;
* Разрешение: 1.55 Å, 19.72 - минимальное и 1.50 - максимальное разрешение для использованных рефлексов;
* Процент использованных рефлексов относительно всех возможных рефлексов с разрешением ниже разрешения структуры в целом: 114793 (99,99%) имеют силу сигнала более 3,0,
* Наличие некристаллографических симметрий в асимметрической ячейке: таковые симметрии есть.

CRYST1 73.690 73.290 80.900 90.00 109.48 90.00 P 1 21 1 4

В ячейке 4 копии одного и того же.

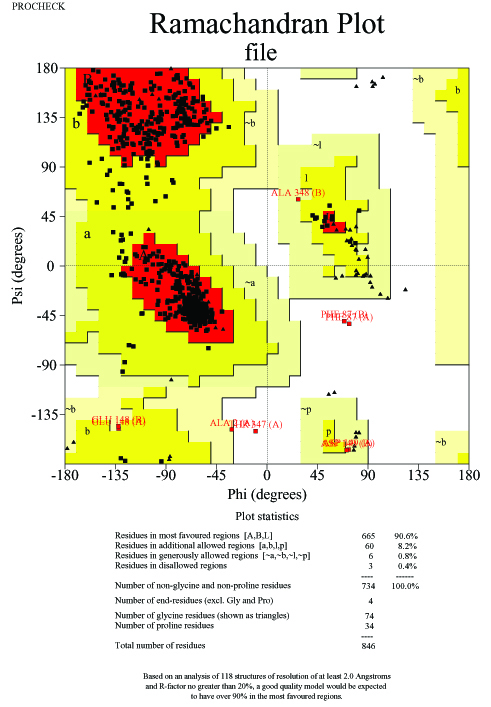
* значения R-фактора и свободного R-фактора и свои выводы:

В статье: R-фактор (R-cryst) = 22,7%; R-free фактор = 28,1%. R-фактор показывает насколько хорошо расчетная модель подходит экспериментальным данным. Поскольку полностью случайный набор атомов дает значение R-фактора = 0,63, идеальное соответствие – 0, а средние значения около 0,20, то данная структура является более-менее средней по качеству. R-free фактор для идеальной модели совпадает с R-фактором, обычно он чуть выше, около 0,26. Для данной модели он немного >25%, поэтому о ней можно говорить как о средней.

R-free - R-фактор = 5,4% опять же, чуть больше порогового значения в 5% для хорошей модели. Но разница, на мой взгляд, не настолько велика, чтобы предполагать overfitting.

* анализ выдачи программы PROCHECK, в том числе копию карты Рамачандрана и процент остатков (отличных от Gly и Pro), попавших в предпочитаемые области на карте, а также свои выводы:

В наиболее предпочитаемые области карты попало 665 а.о. (90,6%). Обычно считается, что для хорошей модели процент а.о., попавших в лучшие области карты должен составлять более 90%, так что можно заключить, что качество структуры чуть выше среднего. В плохие области попало всего 3 остатка.

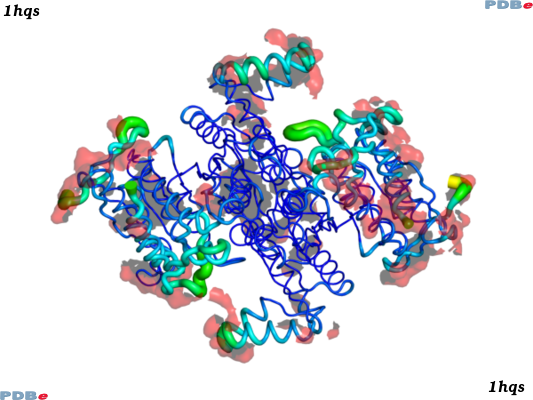


*Рис. 1. Карта Рамачандрана.*

* значения других глобальных индикаторов, смысл которых вам ясен:

MEAN B VALUE (OVERALL, A\*\*2) : 23.00

В-фактор, рассчитываемый для каждого атома, показывает насколько его электронная плотность шире , чем у идеальной модели. В среднем для данной модели <40 (при хорошем разрешении в 1,55 Å), соответсвенно, координатам атомов можно доверять.



*Рис. 2. Остов 1HQS. B-факторы показаны различной толщиной цепи (тонкие = низкие, толстые = высокие).Цвет, варьирующий от синего к красному соответствует величине В-фактора (от 10 до 100 Å2). Красные поверхности показывают регионы, где находятся контакты молекул в кристаллической ячейке (5Å cut-off).*

На сервере EDS можно посмотреть значения RSR. Значение Map R-value = 0.206. Оно находится в пределах нормы.

* список нескольких маргинальных остатков (или гетеромолекул). Каждый из маргинальных остатков может быть выделен по одному или нескольким из следующих признаков:

|  |  |
| --- | --- |
| Остаток | RSR |
| Thr347 | 0,452 |
| Asp423 | 0,394 |
| Met1 | 0,376 |
| Ala2 | 0,344 |
| Lys350 | 0,316 |

Плохое окружение у Thr347 (выдача WhatIf показала, что углерод треонина аномально близок к азоту аланина):

347 THR А CB <-> 348 ALA A N 0.15 2.55 INTRA BF

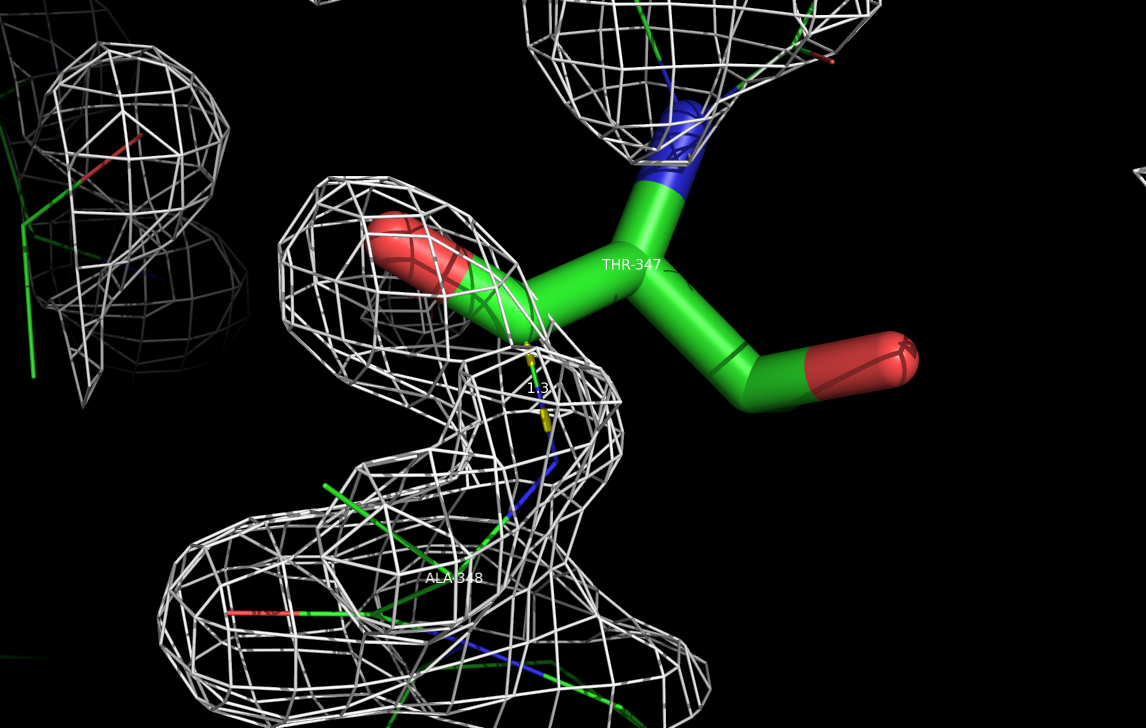
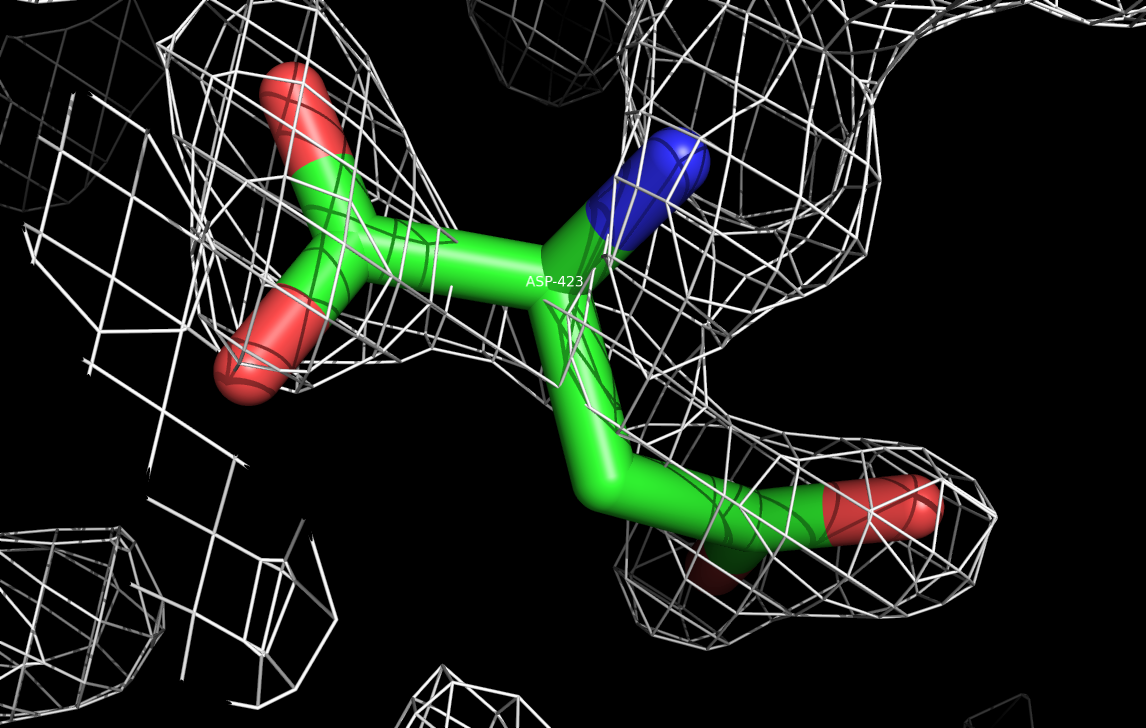


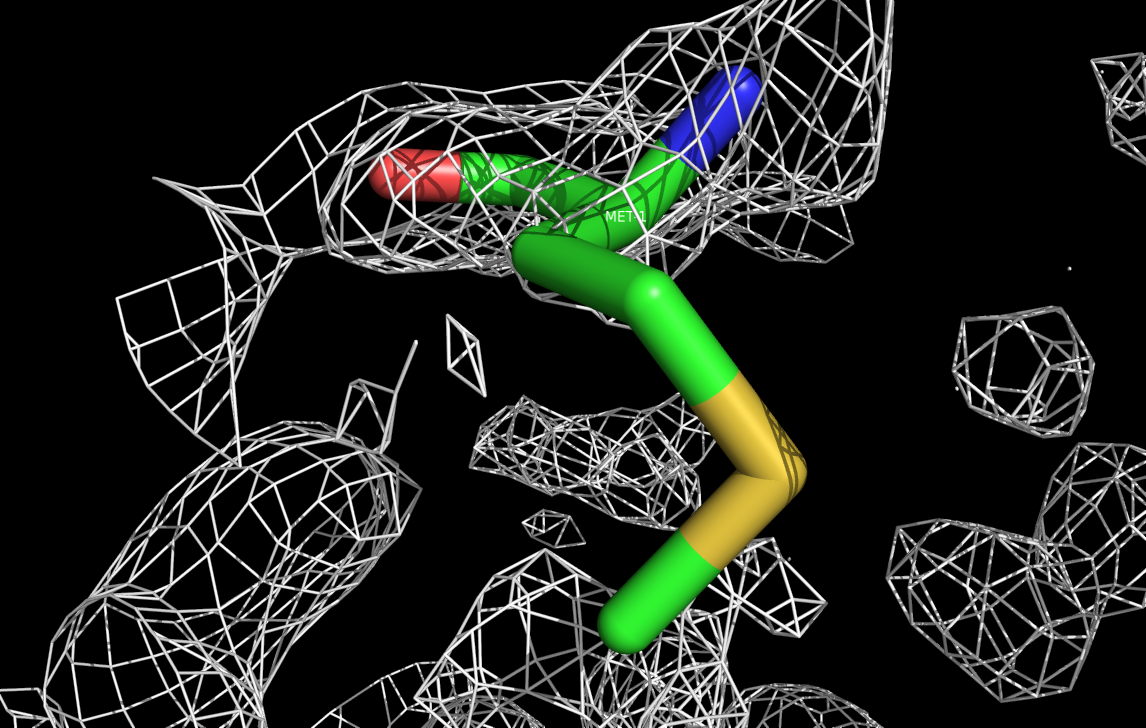
Рис. 3. *После команды* isomesh map1, map, 1.0, chain a and thr347, 2*. Расстояние мужду CB Thr и N Ala 1,3 Å.*

У Asp423 угол χ2 не попадает в интервал [-90; 90].



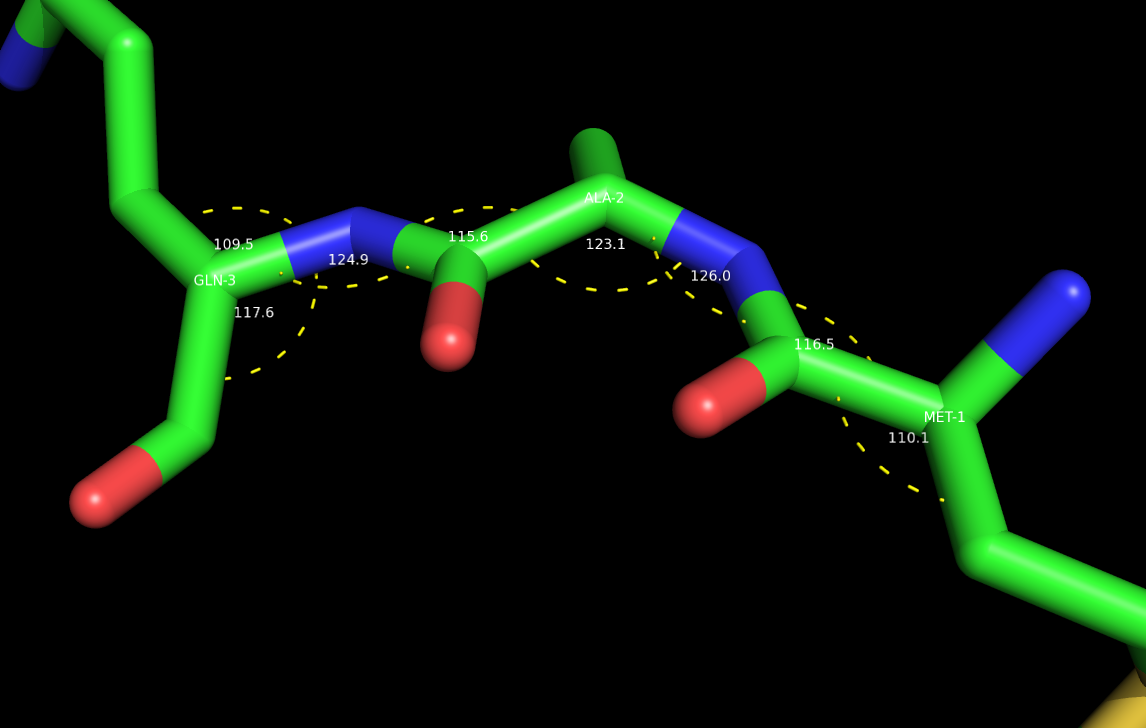
*Рис. 3. Asp423, цепь А.*

Атомы CG, SE, CE у Met1 имеют коэффициент заполнения = 0.



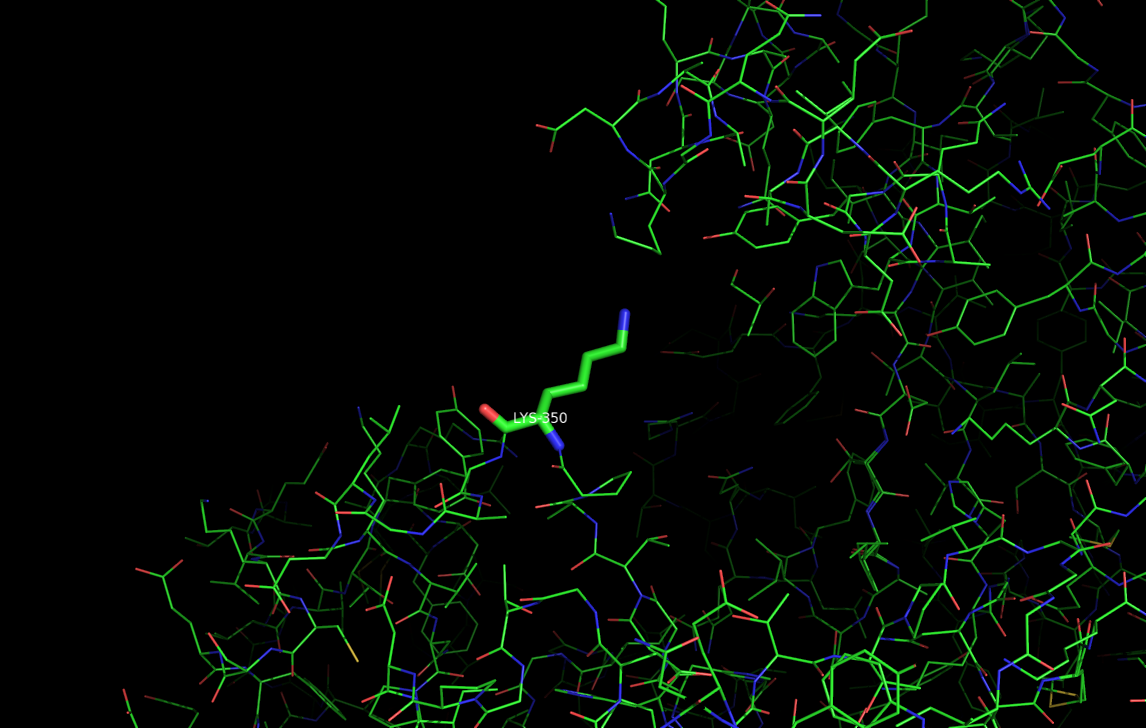
*Рис. 4.Met 1 после команды* isomesh map1, map, 0.5, met1, 1

Угол связи Ala2 больше чем на 4 сигмы отличается от нормального. Я не совсем поняла, что здесь может быть неправильного, если посмотреть на рисунок:



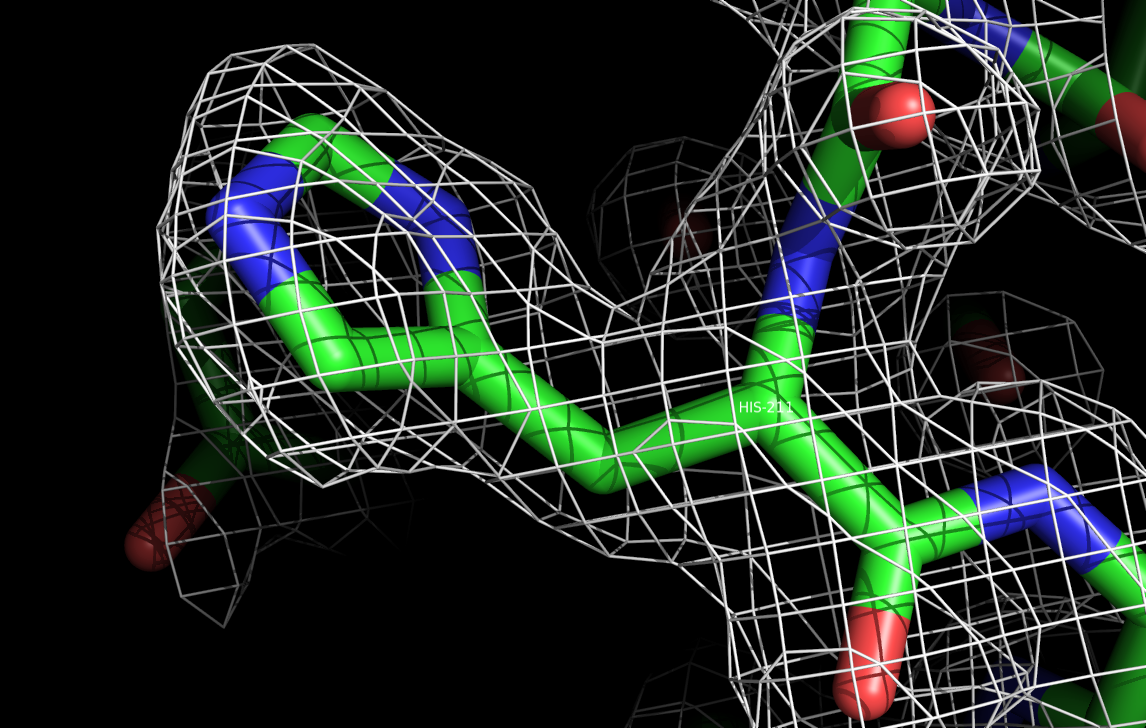
*Рис. 5. Углы между связями Met1-Ala2-Gln3.*

Азот Lys350 образует неверные Н-связи.



*Рис. 6. Lys350. Как видно из рисунка, лизин вообще не может образовать водородные связи, т.к. на расстоянии до 3Å вокруг нет акцепторов электронов.*

В протоколе WHATIF говорится, что у His211 плохое окружение, однако:



*Рис.7. His 211, цепь А.*

Электронная плотность в порядке. Возможно, т.к. он не находится в активном центре фермента, это говорит об особенностях молекулы.

**Заключение.**

 Общее качество модели: глобальные индикаторы качества свидетельствуют, что данная модель хорошо расшифрована, в пределах допустимых значений, число маргиналов невелико, сильных отклонений нет. “Переоптимизация” (overfitting) могла быть, номаловероятно, поскольку разница между соответствующими R-факторами 0,4%.

 Тем не менее, протокол WHAT\_CHECК, на мой взгляд, содержит достаточно много предупреждений о плохих углах и окружении атомов, однако в структуре очень мало атомов и остатков, которые были бы «плохими» по 2-3 и более позициям.

 Наибольшее подозрение вызывают боковые участки полипептидной цепи, там достаточно плохо накладывается электронная плотность на координаты остатков, представленный в PDB-банке.

 Все-таки часть маргинальных остатков – особые. Например, первый остаток метионин – всем известно, что с метионина начинается подавляющее число эукариотических белков, однако его положение плохо коррелирует с полученной электронной плотностью; некоторые а.о., находящиеся в боковых цепях, имеют плохое окружение, хотя их конформация можнт быть достаточно свободной из-за отсутствия большого количества а.о. вокруг.

 Очень подозрительные в отношении правильности расшифровки данные для обоснования биологического вывода авторами статьи использованы не были.

**Литература**

1. <http://www.ncbi.nlm.nih.gov/pubmed/11290745?dopt=Abstract>

2. <http://www.wwpdb.org/docs.html>