# Введение:

В этом отчёте речь будет идти о структуре 1DGK из Protein Data Bank (PDB). Это мутант человеческой гексокиназы в комплексе с АДФ и глюкозой.

Гексокиназа — фермент, катализирующий первую реакцию гликолиза — фосфорилирование глюкозы. Гликолиз — универсальный метаболический путь, используемый для получения энергии.

**Табл.1 — Общие сведения о структуре (из файла 1DGK.pdb)**

|  |  |
| --- | --- |
| Год добавления в PDB | 1999 |
| Организм | Homo Sapiens |
| Метод | РСА |
| Авторы | Aleshin, A.E., Liu, X., Kirby, C., Bourenkov, G.P., Bartunik, H.D., Fromm, H.J., Honzatko, R.B. |
| Resolution (high - low) | 2.8Å – 8.0Å |
| Полнота | 78% |
| R-value (free) | 0.308 |
| R-value (work) | 0.258 |
| ΔR | 0.05 |
| Число рефлексов | 23’882 |
| R-free рефлексов | 1’241 |
| Пропущенные АА | 1-15, 914-917 |
| Остатков-выбросов по углу ковалентной связи | 103 |
| Число мутаций | 6 |
| Параметры ячейки | 45.78 Å / 146.83Å / 58.59 Å  90 / 90 / 90 |
| B-фактор (теоретический) | 54 Å2 |

# Результаты и обсуждение

## Файл 1DGK.pdb

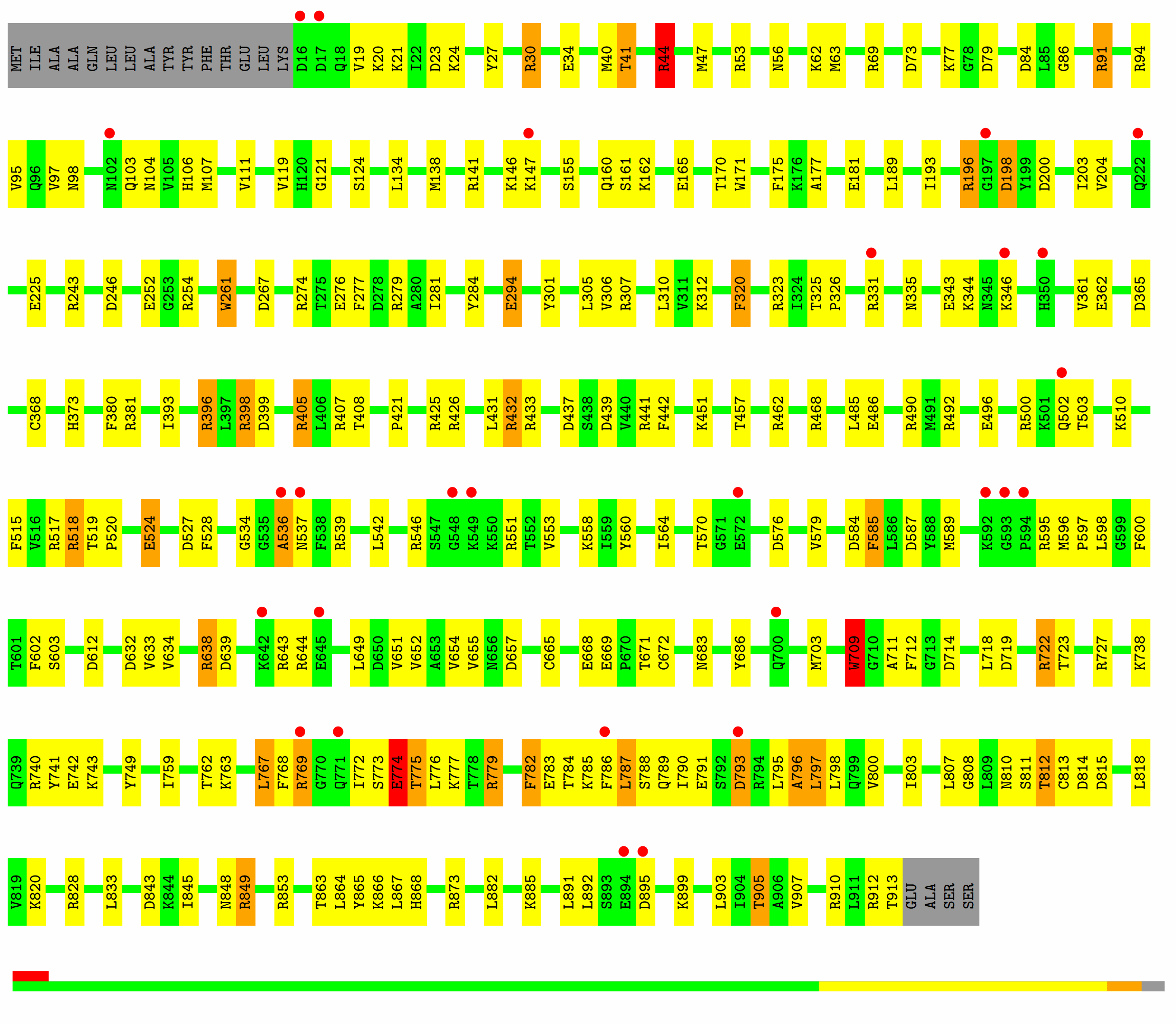
Некоторые данные о качестве пространственной модели фермента содержатся в самом файле 1DGK.pdb (см.табл.1), но на сайте PDB для каждой записи также имеется и отчёт о качестве (“Validation report”). Этот отчёт во многом дублирует данные из самого файла, но предоставляет информацию в более удобном виде.

Например, в отчёте есть график, на котором показано, сколько стерических ошибок содержит каждый остаток в модели (рис.1).

Что интересно, в некоторых местах отчёта цифры не совпадают с указанными в PDB-файле. Например, 1dgk.pdb содержит данные только о теоретически рассчитанном изотропном В-факторе («FROM WILSON PLOT ... (A\*\*2) : 54.00»). В отчёте тоже есть вилсоновский В-фактор — 64.7 Å2, при этом средний В-фактор составляет 50.0Å2. Такие значения кажутся подозрительными: почему на однои сайте для одной структуры были посчитаны два разных В-фактора? Почему средний В-фактор настолько меньше теоретического?

R-value в отчёте и файле тоже разнятся: 0.219 и 0.258, соответственно. Стоит отметить, что в файле структурных факторов не отмечены рефлексы, использованные для подсчёта R-free, что поднимает сомнения насчёт значения этого параметра, указанного в PDB-файле.

Отчёт также содержит информацию об остатках-выбросах по какому-либо параметру. В отчёте указаны 156 выбросов по углам ковалентной связи (против 103, названных в PDB-файле). Возможно, это связано с разными определениями выброса. Тем не менее в отчёте выбросами считаются углы с Z-score >5, а выбросов в нём всё равно больше.



**Рис.1** — графики из “PDB Validation Report”; в верхней части изображена полипептидная цепь гексокиназы, окрашенная в соответствии с числом стерических ошибок в остатке: зелёный — 0 ошибок, жёлтый — 1 ошибка, оранжевый — 2 ошибки, красны ≥3 ошибок, серый — эти остатки в структуре пропущены. Несколько безошибочных остатков объединяются в зелёную линию. На диаграмме в низу рисунка показаны доли остатков с и без ошибок.

В отчёте указано, что структура не содержит выбросов по хиральности или по длине связи и содержит один выброс по планарности остатка — аланин-796 — что не кажется серьёзной ошибкой.

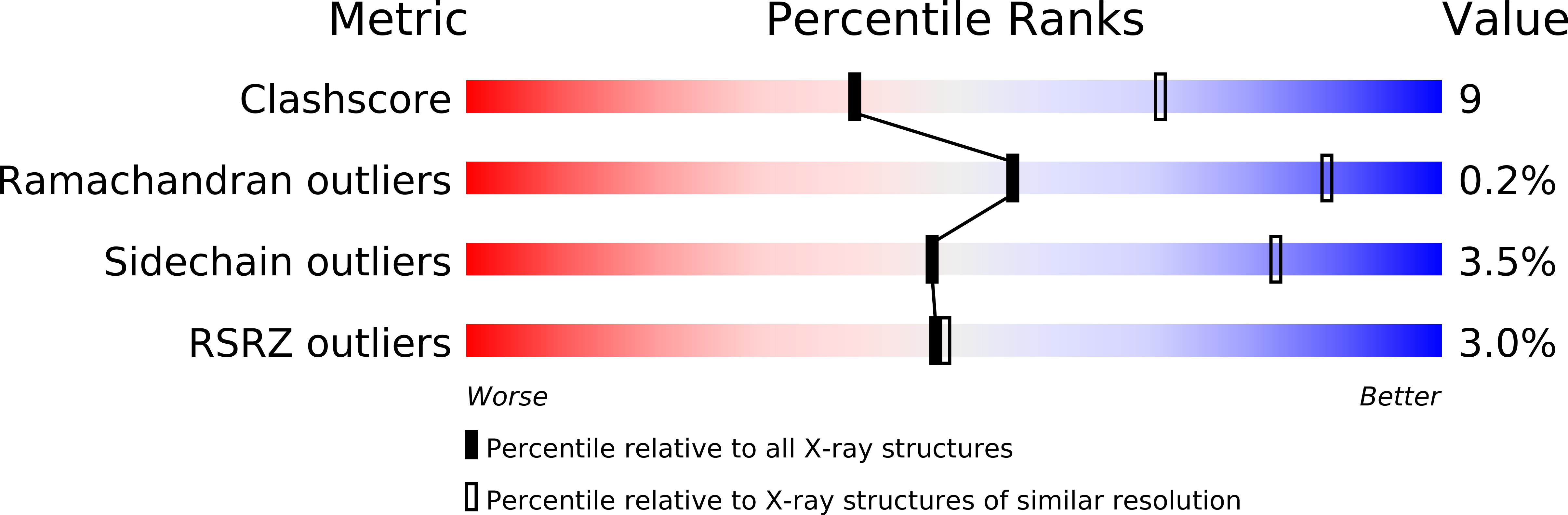
Таким образом, структура 1DGK содержит много стерических ошибок, а PDB файл производит такое впечатление, будто бы его авторы пытались искуствееер изменить показатели качества модели.

Некоторые индикаторы качества пространственной модели были введены позже 1999 года, когда была составлена 1DGK, поэтому о них нет информации в PDB-файле, но есть в отчёте.

Clashscore — промилле нетипично близко расположенных пар атомов, используется для оценки структур с 2010. Для 1DGK клашскор составляет 9, что гораздо больше, чем в среднем для структур с таким разрешением (см.рис.2). В 2010 также стали оценивать число остатков, не соответствующих распределению Рамачандрана углова фи и пси; в 1DGK таких остатков два. Также у 28 остатков нетипичные ротамеры, или конформации радикалов.

Стоит отметить и чрезмерное число выбросов (Z-score > 2) в молекулах АДФ (13% выбросов по длине связи и 17-24% по углу) и глюкозы (23-41% по углу связи).

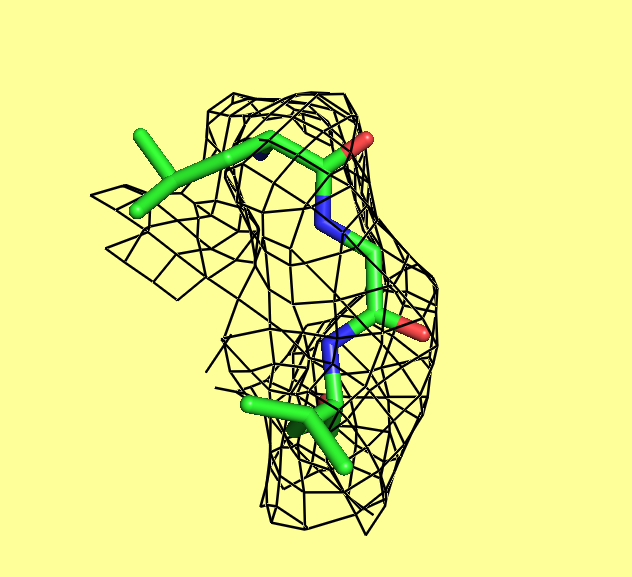
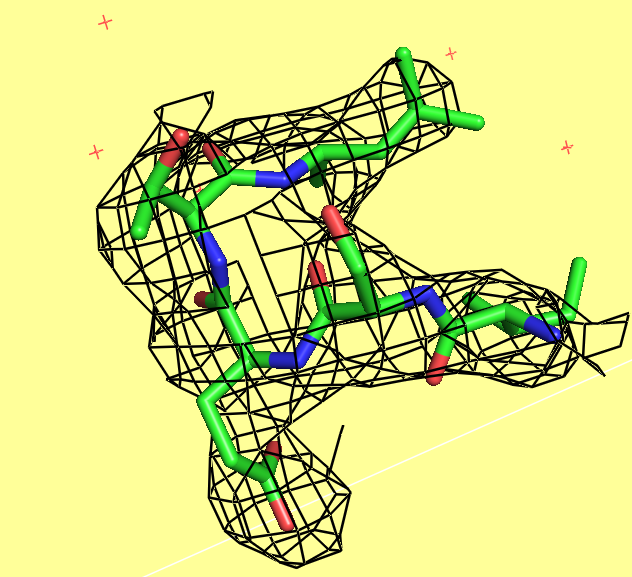
В целом структура 1DGK вызывает много вопросов, и согласно индикаторам качества, не соответсвует указанному разрешению (см.рис.2).



**Рис.2** — Некоторые индикаторы качества структуры и их значения (правая колонка) для структуры 1DGK. Во второй колонке изображены диаграммы с индикаторами 1DGK в сравнении со всеми структурами, полученными методом РСА, (чёрные палочки) и со структурами такого же разрешения (прозрачные палочки). Видно, что качество 1DGK гораздо хуже, чем стоило бы ожидать от структуры с разрешением 2.8 Å.

## Примеры маргинальных остатков в структуре 1DGK:

Первым делом я проверил два выброса по картам Рамачандрана: Glu774 и Gly808. В случае с Glu774 и его окружением, ЭП и вписанные в неё остатки выглядят правдоподобно, но в случае Gly808 окружающие остатки (Leu807, Leu809, Gln806, Asn810) не вписываются в ЭП.

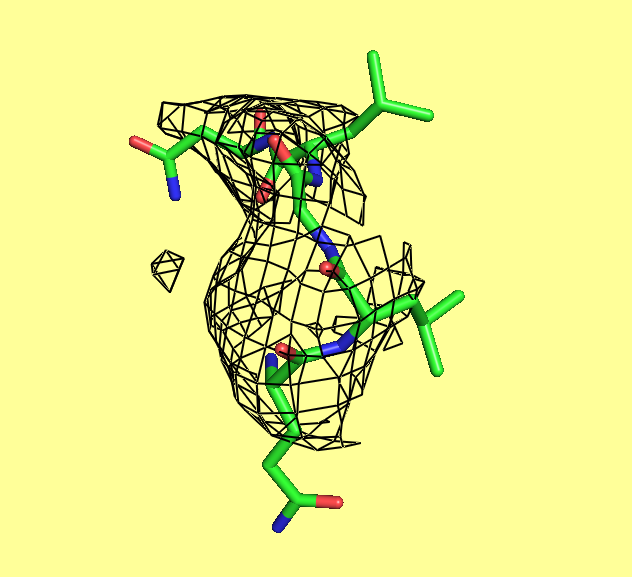


Glu774

Gly808

Б

А

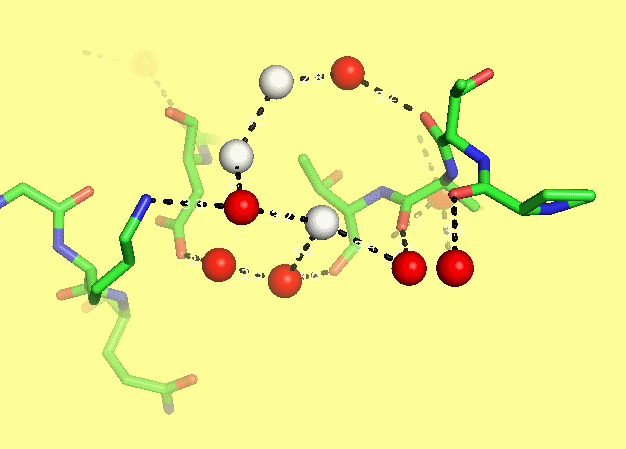
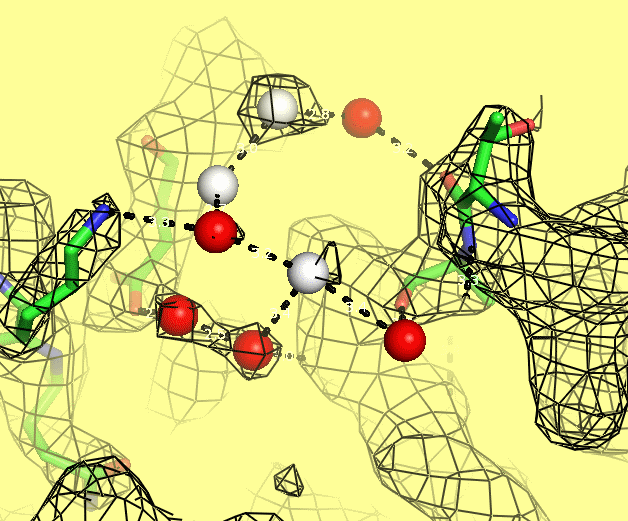


В

**Рис.3** — А: остатки 772-776, уровень срезки 1; расположение карбоксильной группы Glu774 cовпадает с зоной повышенной электронной плотности. Б: остатки 807-809, уровень срезки 0.7; Gly808 неоднозначно вписывается в ЭП; стоит также обратить внимание на радикалы соседних остатков, выпирающих за обозначенные границы ЭП (В).

Упомянутые Leu807, Leu809 и Asn810 также являются выбросами по структуре боковой цепи. То есть данный участок модели в целом является проблемным.

Далее были проверены остатки воды в модели. Из 94 молекул воды 90 образуют водородные связи с белком, а 4 — с другими молекулами воды. То есть вода в этой структуре вряд ли часто использовалась как простой способ заполнить электронную плотность. Тем не менее в одной области структуры заметно крупное скопление молекул воды (рис.4).

Б

A

**Lys290**

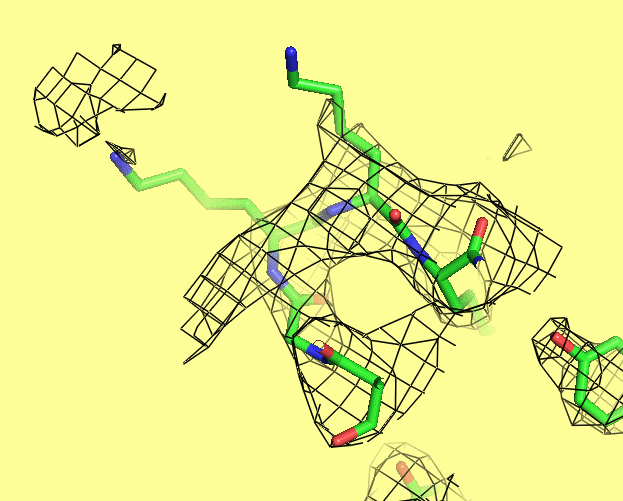
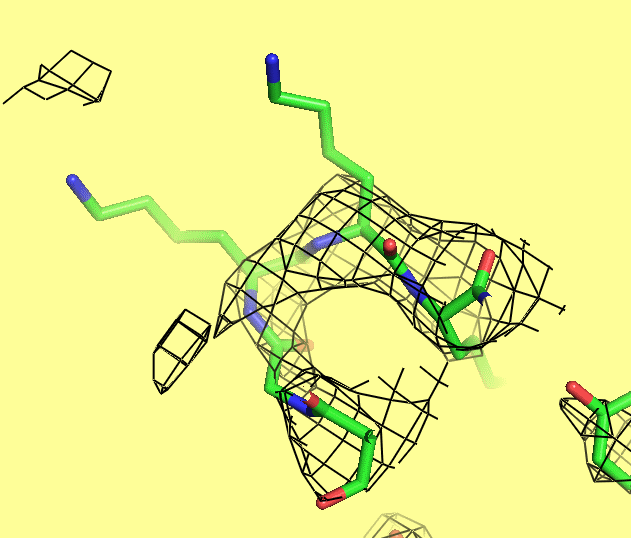
Gln291

Asp266

Thr58

**Рис.4** — А: скопление молекул воды в структуре 1DGK; белым отмечена вода, не связанная напрямую с ферментом. Thr58 и Lys290 являются выбросами по углам связей. Б: электронная плотность на уровне срезки 1 в указанной области. ЭП воды лежит обособленно от ЭП белка. Это говорит о том, что в данном случае молекулы воды не использовались для заполнения ЭП, в которую могли вписаться аминоксилоты. То есть, вероятней всего, кристаллы гексокиназы действительно содержат эти молекулы.

Самая частая и очевидная ошибка в этой модели — неправильные. Во многих участках структуры ЭП позволяет определить только положение остова фермента, но не радикалов. На рис.5 приведён пример такого участка.

Lys550

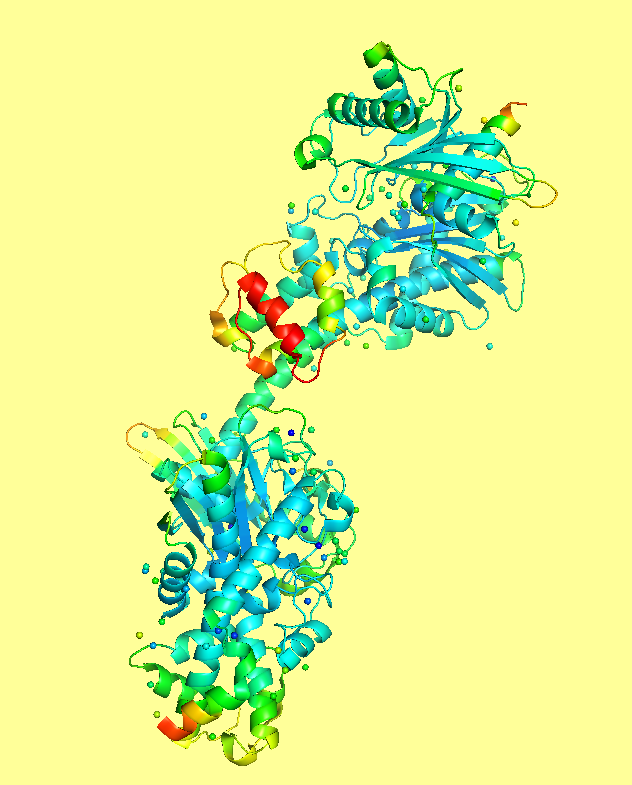
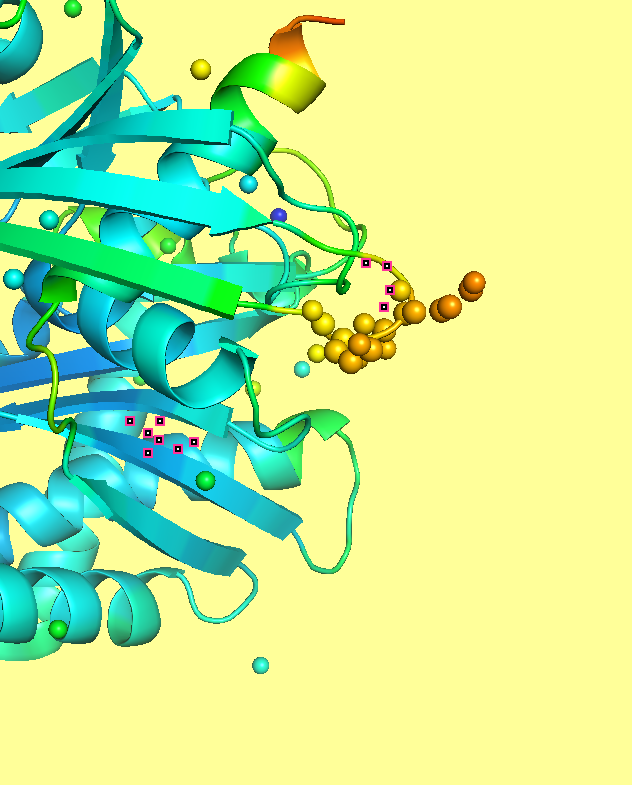
Lys549

Б

A

**Рис. 5** — Участок структуры 1DGK с неопределённым положением ротамеров (А: уровень срезки 0.7; Б: уровень срезки 1.0). Видно, что радикалы не окружены ЭП. Более того, у Lys549 видна область с высокой ЭП (отмечена стрелкой), не заполненная ничем. Возможно, в ней и должен располагаться радикал этой аминокислоты. В структуре 1DGK много подобных участков, что позволяет усомниться в адекватности составленной модели.

Подобные ошибки чаще всего встречаются в петлях структуры, обладающих более высоким В-фактором, чем прочие участки (см.рис.6).

**Рис. 6** — А: структура 1DGK, раскрашенная по значениям В-фактора из файла 1DGK.pdb. Б: Lys549 и Lys550 (отмечены стрелкой) с рис.5 находятся в области структуры с высоким В-фактором.

## Сравнение модели с моделью из PDB\_redo:

PDB\_redo ­— база данных обновлённых и улучшенных структурных моделей, построенным по результатам РСА.

В этой БД имеются пересчитанные значения R и R-free: 0.2241 и 0.2292, соответственно, что значительно отличается от указанных в оригинальном файле 0.258 и 0.308.

Справка об 1DGK в PDB\_redo сообщает, что положение 28 остатков было значительно улучшено, а 1 — значительно ухудшено.

Элементы вторичной структуры при этом не изменились, но в петлях структуры поменялось расположение некоторых радикалов (см.рис.7). И хотя общее качество по индикаторам, приведённым на странице 1DGK, стало лучше, мне кажется, что полученная в ходе РСА ЭП недостаточно точна, чтобы однозначно составить структуру гексокиназы. Поэтому к любым модификациям 1DGK необходимо относиться по-прежнему осторожно.

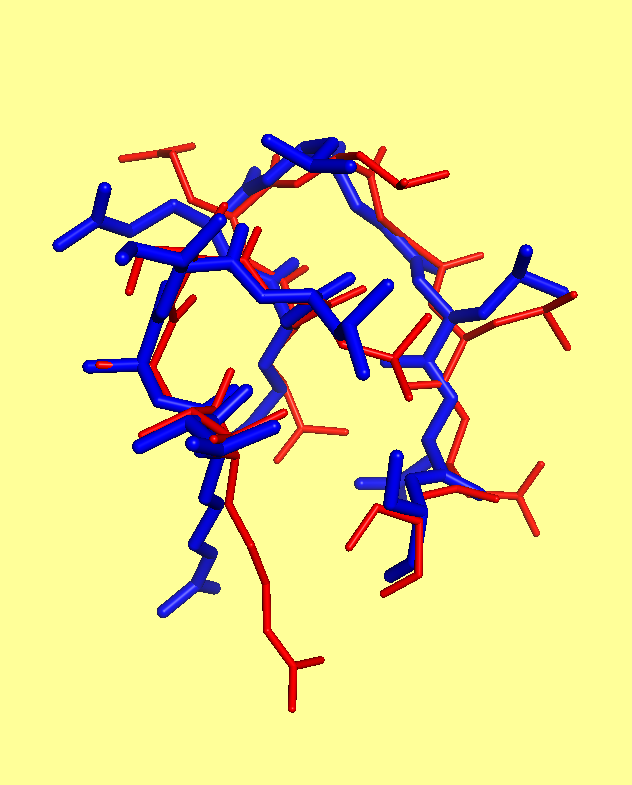


Рис.7 — фрагмент 800-811 структуры 1DGK. В PDB\_redo указано, что структура 3 аминокислот из этого промежутка была значительно улучшена. Красным обозначена модель из PDB, синим — из PDB\_redo. Видно, что положения радикалов аминокислот рознятся. Тем не менее, ЭП в этом участке (не показано), как и во многих петлях этой структуры, позволяет на взгляд определить только положение пептидного остова гексокиназы, из-за чего сложно сказать, в какой БД модель более правильная.

# Заключение:

Исходная ЭП страдает высокими В-факторами в участках, которые при составлении модели были определены как петли.

В подобных участках нельзя однозначно определить ориентацию радикалов, поэтому к их положению часто стоит относиться с осторожность. В PDB\_redo утверждается, что модель была значительно улучшена, хотя при ЭП, которая на довольно низких уровнях срезки охватывает только остов белка, сравнение качества двух моделей — задача странная.

Описание 1DGK в PDB-файле составлено небрежно. Неясно, как и по каким рефлексам считались R-free. Средний В-фактор не указан. Занижено число остатков со стерическими затруднениями.

В итоге, 1DGK — структура с гораздо более низким качеством, чем ожидается от РСА с указанным разрешением (2.8А).

# Ccылки:

[1DGK из PDB](http://www.rcsb.org/pdb/explore/explore.do?structureId=1dgk)

[1DGK из PDB\_redo](http://www.cmbi.ru.nl/pdb_redo/dg/1dgk/index.html)

[MOLprobity, сервис производящий оценку качества PDB-структур](http://molprobity.biochem.duke.edu/)

[Папка с файлами, отчётами о качестве, картинками](https://www.dropbox.com/s/2m6r5f02t5g4p2w/1dgk.7z?dl=0)