

Отчёт по качеству расшифровки
структуры белка глицеринкиназы
Thermococcus kodakarensis
методом РСА

Выполнила студентка
4 курса ФББ МГУ
Заира Сефербекова

Аннотация

В отчёте рассмотрены некоторые индикаторы качества модели структуры 2ZF5, а также проведён анализ индикаторов локального качества структуры для ряда аминокислотных остатков. В ходе работы над отчётом удалось познакомиться с возможностями, которые предоставляют серверы для оценки качества структур белковых молекул.

Введение

Глицеролкиназа (ЕС 2.7.1.30; АТФ: глицерин 3-фосфотрансфераза, ГК) катализирует Mg,АТФ-зависимое фосфорилирование глицерина с образованием глицерин 3-фосфата — важного интермедиата гликолиза:

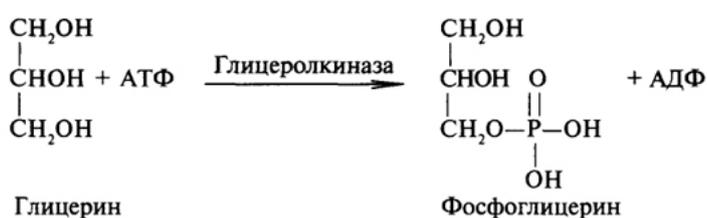


Рис. 1. Реакция фосфорилирования, катализируемая глицеролкиназой

ГК относится к суперсемейству АТФаз, белки которого имеют общий мотив $\beta\alpha\beta\alpha\beta\alpha$ и характеризуются значительными конформационными перестройками при связывании с субстратом. ГК широко представлена во всех живых организмах: бактериях, археях и эукариотах.

Лучше всего изучена ГК *Escherichia coli* (Ес-ГК), которая в зависимости от концентрации существует в виде активного димера или неактивного тетрамера. Димер в свою очередь образован двумя асимметричными субъединицами, одна из которых находится в закрытой конформации, а другая — в необходимой для катализа открытой конформации. Конформация изменяется при движении элементов структуры между доменами. Фруктозо-1,6-бисфосфат является аллостерическим ингибитором Ес-ГК: он селективно связывается с тетрамером и ингибирует активность фермента, мешая движению междоменного участка белка.

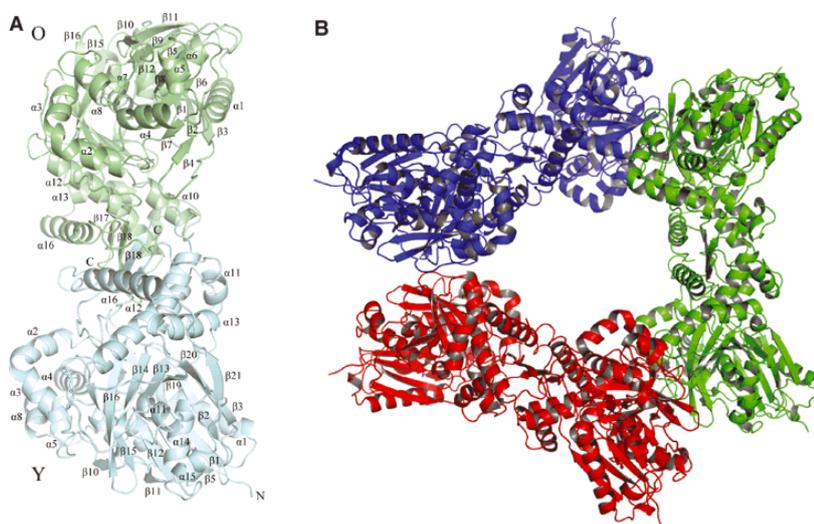


Рис. 2. Структура Тк-ГК. (А) Структура димера. (В) Структура псевдогексамера.

ГК из термофильной археи *Thermococcus kodakaraensis* (Тк-ГК), которая будет рассматриваться далее, всего на 57% идентична Ес-ГК по аминокислотной последовательности и характеризуется повышенной термостабильностью: ферментативная активность Тк-ГК сохраняется при нагреве до 90°C в течение 60мин. Считается, что такое свойство обусловлено заменой 83 остатков С-конца, но точного обоснования на данный момент нет. Кроме того, Тк-ГК не чувствительна к ингибированию фруктозо-1,6-бисфосфатом.

Димер Тк-ГК состоит из двух субъединиц О и Y (рис. 2А). Каждая субъединица в свою очередь состоит из двух доменов, содержащих типичный для данного суперсемейства мотив $\beta\alpha\beta\alpha\beta\alpha$, разделенных междоменной щелью. В кристалле Тк-ГК формирует псевдогексамер, состоящий из трех димеров (рис. 2В). Было показано, что в растворе даже при высоких концентрациях Тк-ГК в отличие от Ес-ГК существует только в виде димера. Тем не менее, наличие или отсутствие физиологической роли у гексамера все еще является открытым вопросом (1).

Результаты и обсуждение

Общая информация о модели

Модель 2ZF5 представляет собой димер (497 а.о., 55.97кДа), состоящий из субъединицы О и Y, связанных лигандов в структуре нет. Также в pdb-файле авторы представили менее вероятную модель гексамера, о которой говорилось выше.

Модель была получена в 2008 году учеными Koga, Katsumi, You, Matsumura, Takano and Kanaya из Осаки, Япония. Для разрешения фазовой проблемы авторы использовали метод молекулярного замещения ранее полученной ими же структурной моделью Тк-ГК (2). Общая информация и значения индикаторов качества модели приведены в таблице 1.

Таблица 1. Общая информация о модели

Метрика	Значение	
Параметры элементарной ячейки	$a = 217.48 \text{ \AA}$	$\alpha = 90^\circ$
	$b = 217.48 \text{ \AA}$	$\beta = 90^\circ$
	$c = 66.48 \text{ \AA}$	$\gamma = 120^\circ$
Число измеренных рефлексов	45 935	
Полнота набора рефлексов	45 094 (98.17%)	
Разрешение	2.40 \AA	
Диапазон разрешений структурных факторов	2.4 \AA - 41.1 \AA	
Псевдотрансляции	Не обнаружены	

Значения индикаторов качества модели

R-фактор — мера качества атомной модели, полученной из данных PCA. R-фактор показывает, насколько амплитуды структурных факторов, рассчитанных по кристаллографической модели, сходны с экспериментальными. Значение R-фактора

уменьшается при увеличении количества параметров, используемых в модели, поэтому это не самый надежный индикатор качества модели (9). Тем не менее, хорошими считаются значения R-фактора ~ 0.2 (6), и в случае обсуждаемой модели он меньше этого значения и составляет 0.174, то есть модель хорошо описывает полученные экспериментальные данные.

Сравнение обсуждаемых далее индикаторов качества с их типичными значениями показаны на рис. 3 (9), а параметры, подсчитанные сервисом MolProbity (4), приведены в таблице 2.

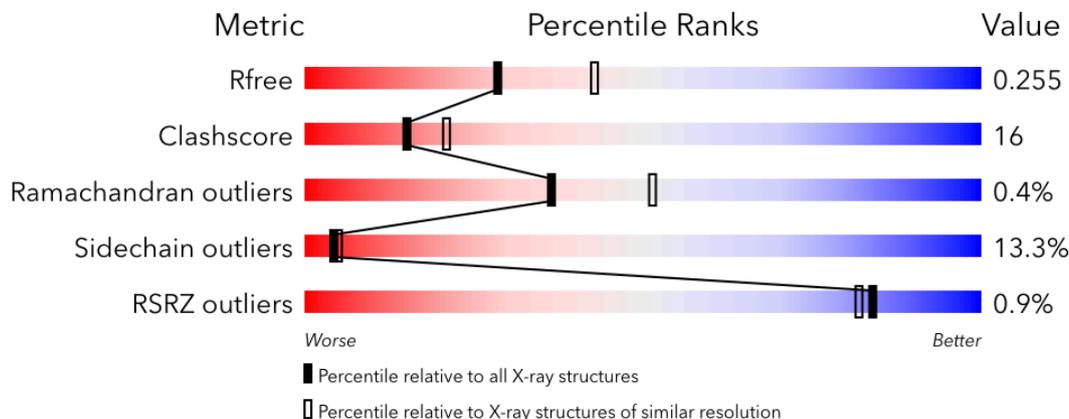


Рис. 3. Показатели качества модели по сравнению со средними показателями структур с тем же разрешением.

All-Atom Contacts		Clashscore, all atoms:		14.33	84 th percentile* (N=331, 2.40Å ± 0.25Å)
Protein Geometry	Poor rotamers	116	14 %		Goal: <0.3%
	Favored rotamers	606	75 %		Goal: >98%
	Ramachandran outliers	4	0 %		Goal: <0.05%
	Ramachandran favored	923	94 %		Goal: >98%
	MolProbity score [^]	2.95			33 rd percentile* (N=8058, 2.40Å ± 0.25Å)
	Cβ deviations >0.25Å	36	4 %		Goal: 0
	Bad bonds:	39 / 8006	0 %		Goal: 0%
	Bad angles:	63 / 10844	1 %		Goal: <0.1%

Таблица 2. Статистика по различным параметрам геометрии белка.

R_{free} рассчитывается так же, как и R-фактор, но на маленькой случайной выборке рефлексов, не используемой при оптимизации. R_{free} показывает, как хорошо модель предсказывает «новые» экспериментальные данные, и позволяет отследить уменьшение R-фактора за счет переобучения модели. В случае идеальной модели R_{free} близок к R-фактору и разница между ними < 0.05 , обычно же R_{free} составляет около 0.26 (8). Для исследуемой модели R_{free} составляет 0.255, то есть разница $|R - R_{free}| = 0.081$, что также говорит о хорошем качестве модели.

Real-space R-value (RSR) — мера сходства карты электронной плотности, рассчитанной по кристаллографической модели, с картой электронной плотности, рассчитанной по экспериментальным данным. Преимуществом этой метрики является то, что она может быть рассчитана и для произвольного набора атомов, что позволяет с ее помощью оценивать качество модели для конкретных а.о. (9). На рис. 4 показаны значения RSR для обеих цепей.

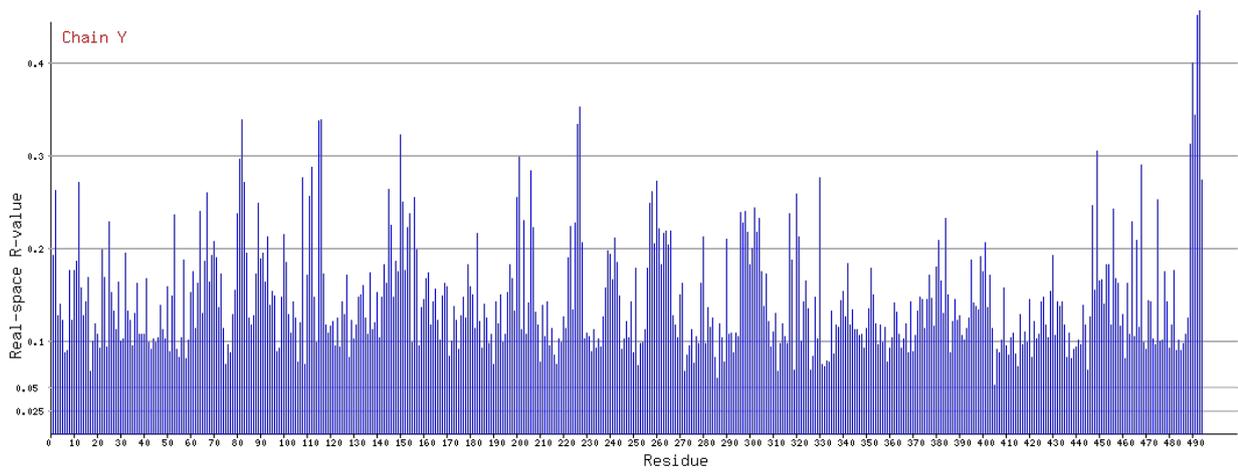
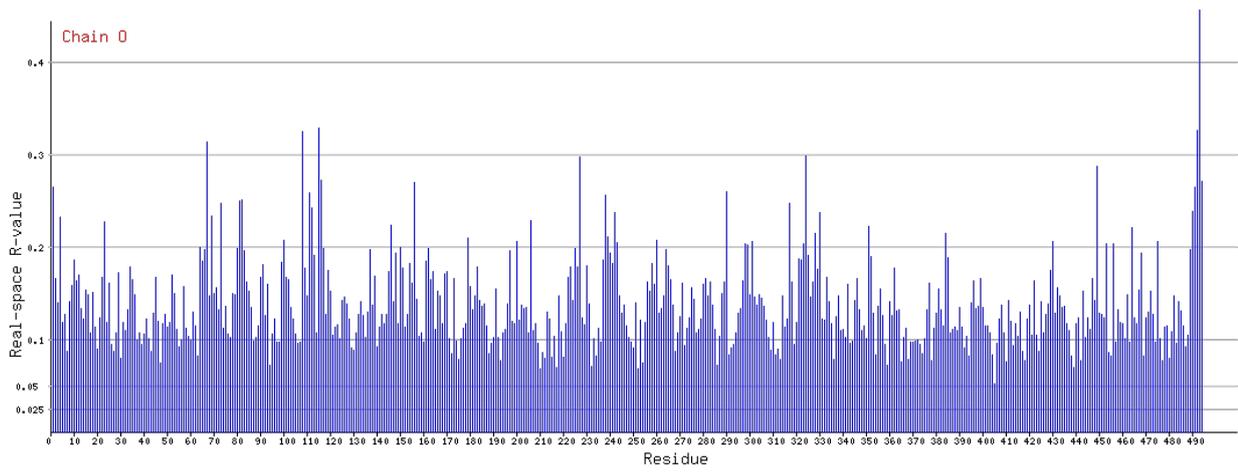


Рис. 4. Показатели RSR для остатков цепи O (вверху) и Y (внизу).

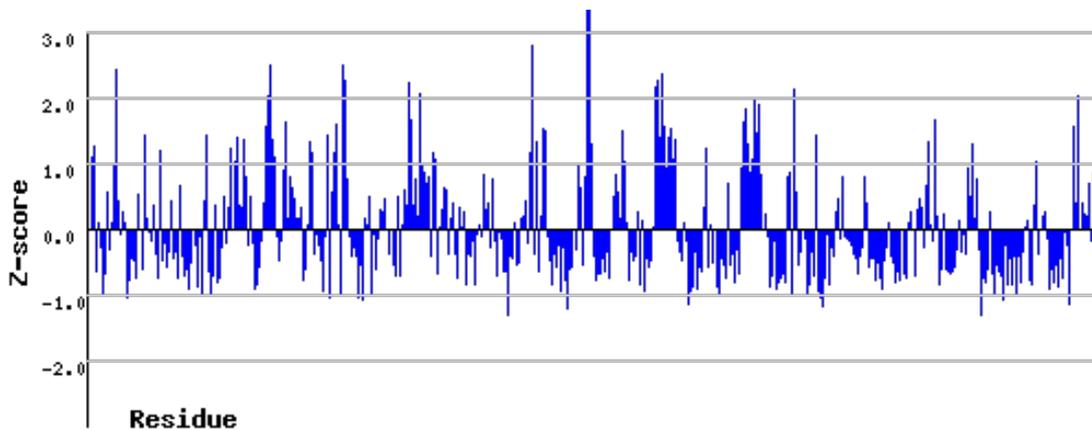
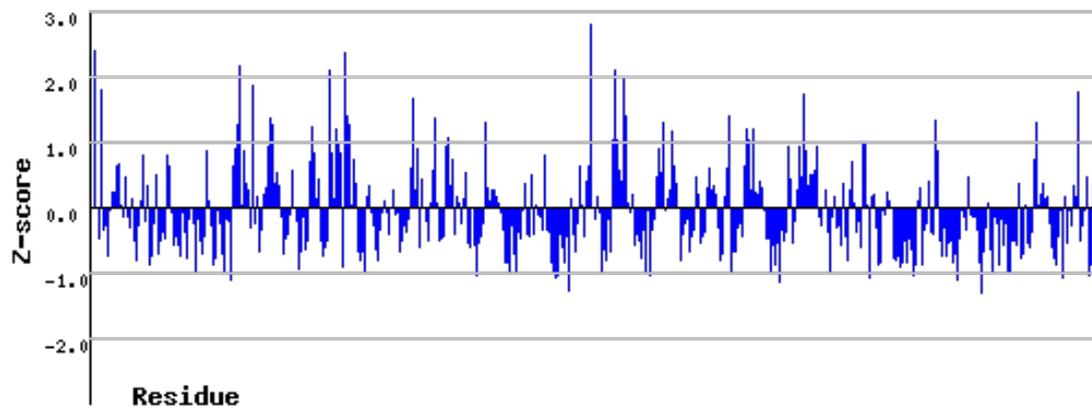


Рис. 5. Показатели RSR-Z для остатков цепи O (вверху) и Y (внизу).

RSR Z-score (RSR-Z) позволяет учесть зависимость электронной плотности от разрешения типа а.о. Сначала для RSR значений для данного разрешения и данного а.о. рассчитывается среднее μ и стандартное отклонение σ , затем рассчитывается Z-score: $(RSR - \mu) / \sigma$. Чем больше значение RSR-Z для данного а.о., тем хуже а.о. вписан в электронную плотность. Приемлемыми считаются значения $RSR-Z < 2$ (9). В обсуждаемой модели число маргинальных по этому параметру а.о. небольшое относительно других структур. На рис. 5 показаны значения RSR-Z для обеих цепей.

Clashscore — число пар атомов модели, слишком близко расположенных друг к другу (наложение Ван-дер-Ваальсовых радиусов $\geq 0.4\text{\AA}$), на 1000 атомов. На рис. 3 видно, что относительно других структур данный показатель говорит о низком качестве модели. Однако результаты проверки сервисом MolProbity (см. таблицу 3) говорят об обратном: у 84% структур с тем же разрешением данный параметр больше, то есть у рассматриваемой модели он хороший.

В целом, по значениям параметров можно сделать вывод, что качество модели ниже среднего.

Маргинальные остатки

Карта Рамачандрана — график торсионных углов ϕ (между N_{i-1} , C_i , $C_{\alpha,i}$, N_i) и ψ (между C_i , $C_{\alpha,i}$, N_i , C_{i+1}) для всех а.о. белка. Торсионные углы определяют конформации а.о. и всего белка. По оси Oх откладываются значения угла ϕ , а по оси Oу — угла ψ , так что итоговый график иллюстрирует какие комбинации углов возможны.

В современном варианте карт Рамачандрана используются отдельные графики для аминокислот, обладающих особенными локальными стерическими характеристиками. Например, торсионные углы в Gly из-за того, что его боковая цепь одним атомом водорода, не имеют некоторых ограничений, типичных для остальных аминокислот. Поэтому Gly часто попадает в неразрешенные области на общих графиках. Pro, напротив, имеет более жесткие ограничения из-за своей структуры (5).

Для определения маргинальных остатков использовались карты Рамачандрана, построенные сервисами MolProbity (рис. 4) и EDS (рис. 5). Далее будут кратко описаны результаты работы обоих сервисов, а затем приведена таблица с выбранными остатками.

Итак, на рис. 4 представлена карта Рамачандрана, полученная сервером MolProbity (4). На карте голубым показаны благоприятные (**favorable**) комбинации торсионных углов, а синим — разрешенные (**allowed**) комбинации. Согласно результатам работы данного сервиса, в рассматриваемой модели 93.8% а.о. находятся в благоприятных областях, 99.6% — в разрешенных. Четыре маргинальных а.о., не попадающих в эти области, показаны цветом и значения их торсионных углов приведены в таблице 5 («найденны MP»).

На сайте EDS (7) для структуры найдено целых 30 маргинальных остатков (3.3%). Соответствующая карта Рамачандрана для всего белка показана на рис. 5, а характеристики нескольких остатков также приведены в таблице 5 («найденны EDS»). Остаток Y490 встречался в обоих результатах.

В таблице 5 также приведены значения других параметров, подсчитанных сервисом MolProbity (4). Для дальнейшего анализа выбрано пять остатков, выделенных красным цветом, как остатков с наибольшими значениями RSR и сильными отклонениями по нескольким другим параметрам. Также для интереса был взят остаток V490 не с самым большим RSR, но из другой области белка. Маргиналов одновременно и по карте Рамачандрана, и по окружению в отчете (10) найдено не было. Также ни для одного из найденных остатков не была предсказана инверсия боковой цепи при добавлении водородов в структуру.

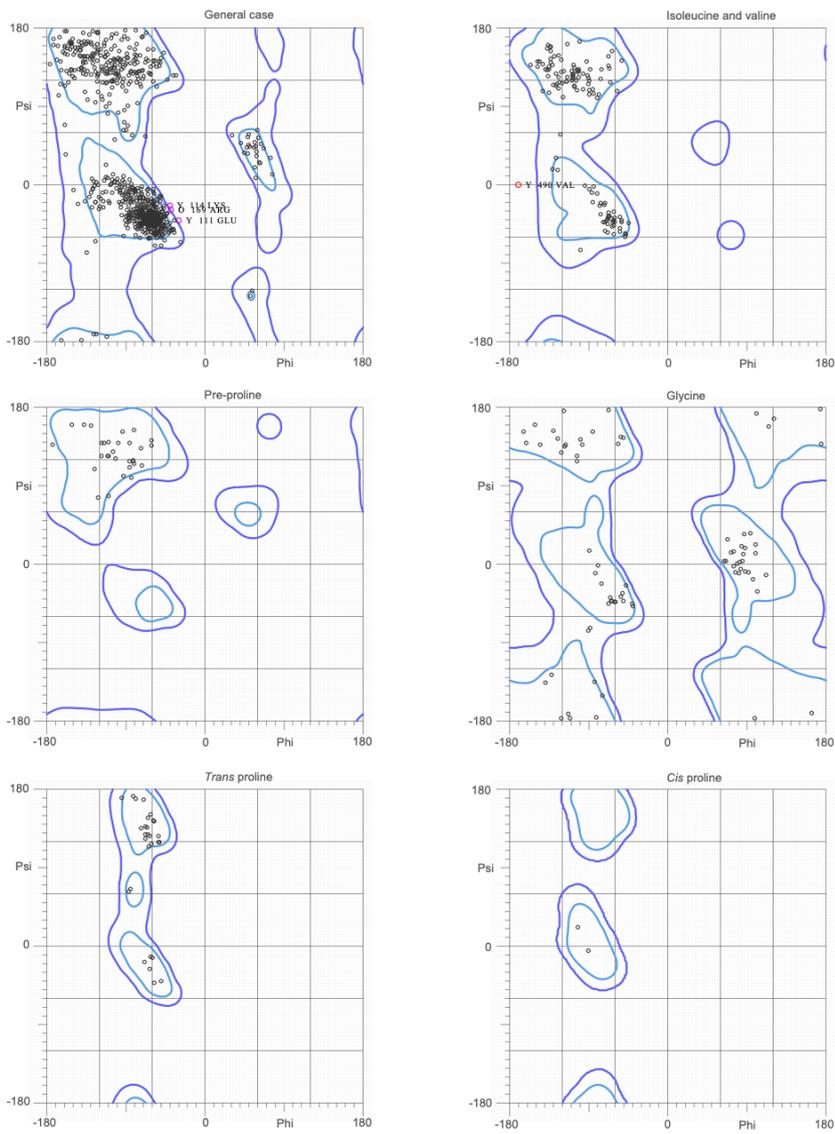


Рис. 4. Карты Рамачандрана, полученные сервисом MolProbity [4], для разных типов остатков модели: всех остатков, кроме названных далее; изолейцина и валина; остатков перед пролином; глицина; транс-пролина; цис-пролина.

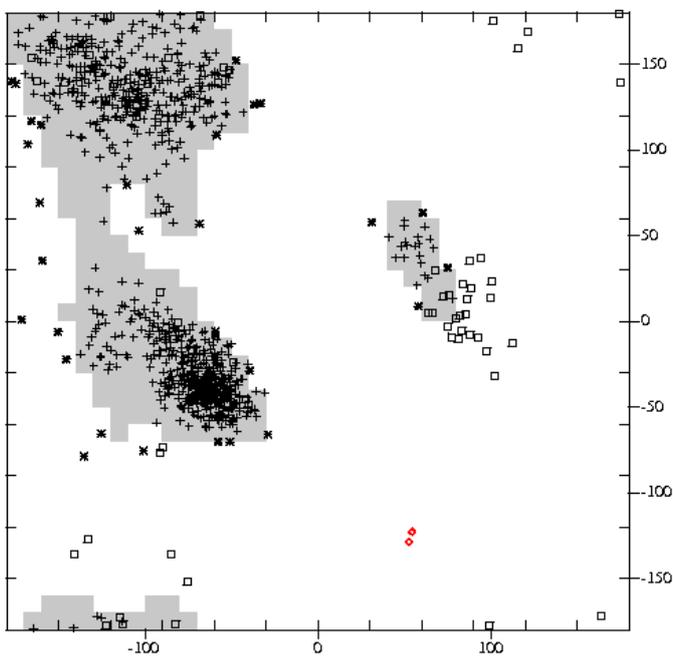


Рис. 5. Карта Рамачандрана, полученная сервисом EDS [4], для всего белка.

Таблица 5. Значения торсионных углов некоторых маргинальных остатков (EDS).

Позиция	а.о.	$\phi,^\circ$	$\psi,^\circ$	RSR	Clashes > 0.4 Å	Rotamer	Найден
O67	ARG	31.3	57.7	0.314	-	Outlier (0.1%)	EDS
O169	ARG	-39.5	-29.1	0.171	-	-	MP
O391	LEU	-145.3	-22.3	0.114	-	Outlier (0.3%)	EDS
O428	GLU	-68.5	56.7	0.139	0.93Å (OE1 with O16 ARG NH1)	-	EDS
O464	LYS	-159.4	35.3	0.221	-	Outlier (0%)	EDS
O490	VAL	-100.9	-75.6	0.239	-	Outlier (0%)	EDS
Y111	GLU	-30.5	-41.8	0.256	-	-	MP
Y114	LYS	-40.3	-24.9	0.099	0.57Å (HG2 with Y129 PRO HB2)	-	MP
Y226	LEU	-28.9	-66.1	0.334	-	Outlier (0.1%)	EDS
Y489	LYS	-103.4	52.7	0.313	1.05Å HG3 with Y489 LYS O	Allowed (0.4%)	EDS
Y490	VAL	-171.3	0.8	0.400	0.56Å HG12 with Y487 TRP O	-	Оба
Y492	ASP	-167.8	103.3	0.451	-	Allowed (0.5%)	EDS
Y493	SER	-165.8	116.6	0.544	-	Outlier (0%)	EDS

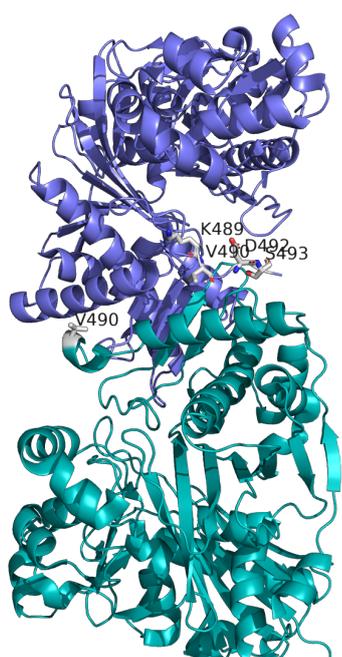


Рис. 6. Расположение рассматриваемых 5 маргинальных а.о. в молекуле белка.

Исследование выбранных маргинальных остатков

В данном пункте будет произведен анализ выбранных остатков по разным параметрам.

На *рис. 6* показано расположение всех пяти остатков в молекуле белка. Видно, что все они располагаются на поверхности, причем V490 находится в альфа-спирали, а четыре остальных — в петле.

V490

На *рис. 7* слева показана водородная связь V490 с молекулой воды из pdb-файла, а справа — электронная плотность остатка. Боковая цепь покрывается электронной плотностью только на уровне подрезки < 0.3 , на рисунке уровень равен 1. Гидрофобная боковая цепь невыгодно экспонирована в раствор и C_{β} -атом отклонен от ожидаемого положения на 0.26\AA (порогом считается 0.25\AA). Параметр *occipancy* в pdb-файле для всех атомов данного остатка равен 1.

Этот остаток был определен как маргинальный сервисом EDS, в то время как по данным MolProbity он лежит в допустимой области карты Рамачандрана (квантиль 0.46%). Возможно, это связано с тем, что в первом случае маргиналы определяются на основании общего для всего белка графика, тогда как во втором Ile и Val рисуются на отдельном графике. На мой взгляд, имеющиеся данные недостаточно убедительны для классификации остатка как маргинального: скорее всего низкое покрытие электронной плотностью связано с большой подвижностью остатка в силу его расположения в петле.

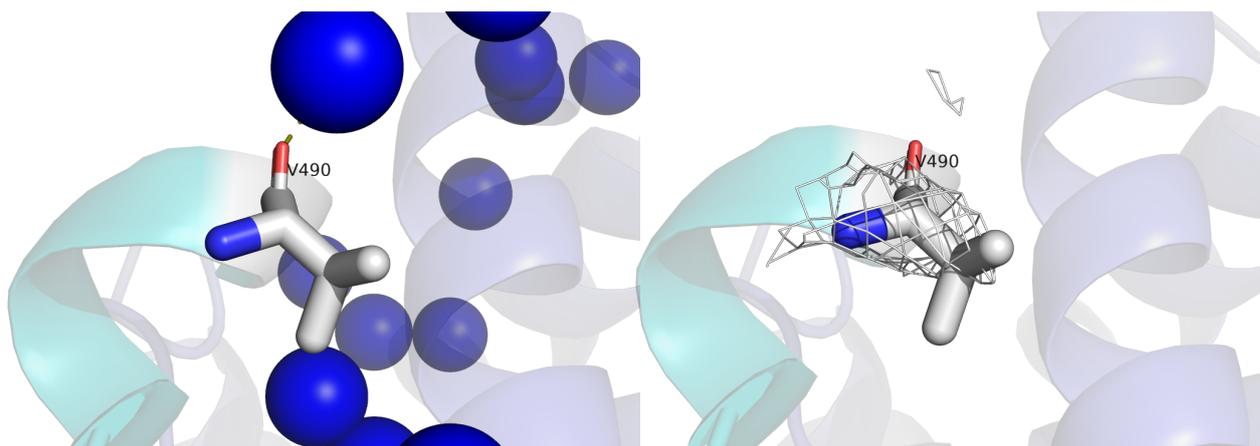


Рис. 7. Водородная связь V490 с водой (*слева*) и его электронная плотность ($\sigma = 1$).

K489, V490, D492, S493

На *рис. 8* показаны рассматриваемые четыре остатка и их электронная плотность при двух разных уровнях подрезки. Видно, что боковые цепи покрываются электронной плотностью на уровне подрезки 0.2, но на уровне 1 атомы оголяются. При этом S493 и D492 не покрыты практически полностью, в то время как у K489 и V490 не покрыты только концы боковых цепей.

Последние два остатка являются маргинальными по наложению радиусов (см. *таблицу 5*), остаток D492 — по существенному отклонению углов (худший: N-CA-C на 4.8σ), а остаток S493 — сразу по нескольким параметрам: по существенному отклонению углов (худший: C-N-CA на 7.0σ), длин связей (худшая: *worst is* N—CA на 5.0σ) и конформации боковой цепи (квантиль 0%, угол 16.1°), и, кроме того, он находится в невыгодном *цис*-положении. Наложения радиусов после добавления водородов изображены на *рис. 9*. Параметр *occipancy* в pdb-файле для всех атомов во всех четырех а.о. равен 1, а температурные факторы во всех случаях колеблются в пределах 70-100 (причем он растет при росте порядкового номера остатка), что довольно много. Молекул воды вблизи данных остатков в структуре нет, контактами с другими остатками полярные боковые цепи также не стабилизированы.

Остаток V490 был определен как маргинальный обоими сервисами, поэтому его мы будем считать действительно неверно расшифрованным. Остальные три остатка были найдены только сервисом EDS, в то время как по данным MolProbity они лежат в допустимой области карты Рамачандрана (квантили в пределах 0.49-1.05%). На мой взгляд, остатки S493 и K489 также плохо расшифрованы, учитывая значительные отклонения их параметров от средних. Тем не менее, в случае остатка D492 я бы сказала, что имеющиеся данные недостаточно убедительны для классификации его как маргинального.

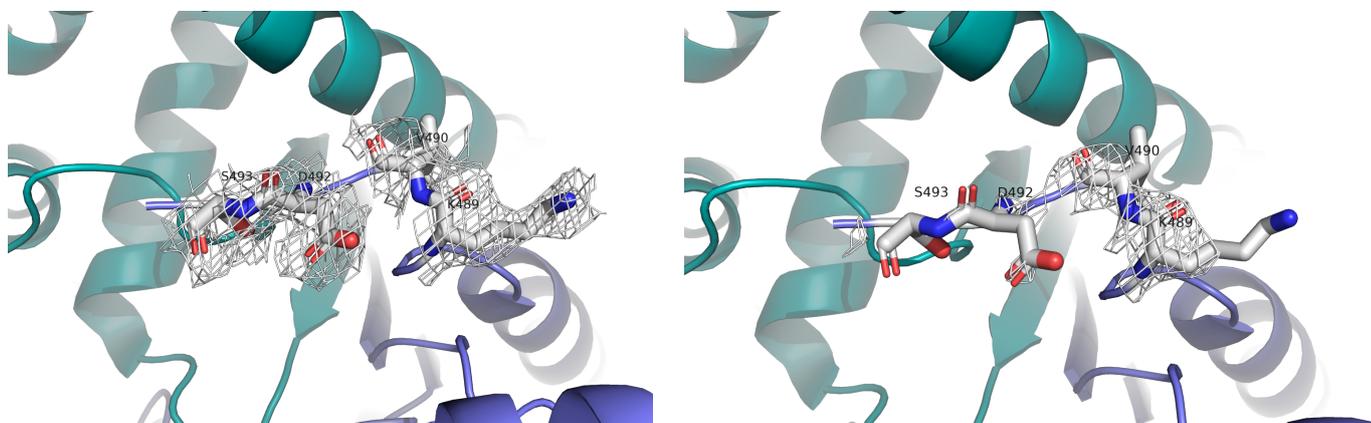


Рис. 8. Электронная плотность остатков при $\sigma = 0.2$ (слева) и $\sigma = 1$ (справа).

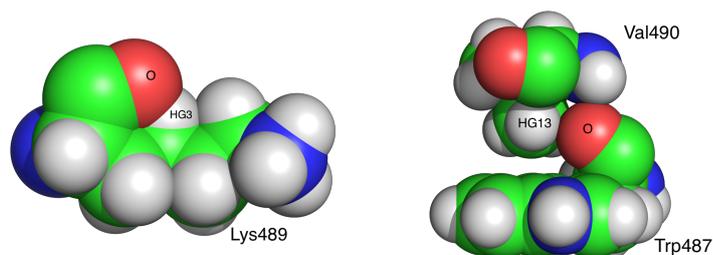


Рис. 9. Наложение радиусов при добавлении атомов водорода для остатков Lys489 (слева) и Val490 (справа).

Сравнение модели из PDB с моделью из PDB_redo.

На данный момент существует улучшенная модель в PDB_redo (11). Сравнение параметров двух моделей приведено в *таблице 6*, а в *таблице 7* приведены характеристики геометрии белка на основании новой модели (ср. с *таблицей 2*). Видно, что несмотря на то, что R-фактор стал больше, значение более показательного параметра R_{free} уменьшилось, что в итоге привело и к уменьшению абсолютной разницы между ними, а значит — к увеличению качества модели.

Показатели, перечисленные в *таблице 7* также говорят в пользу улучшения качества модели. Раньше зеленым было отмечено значение только первого показателя (clashscore), а сейчас — целых четырех. Все маргинальные остатки на карте Рамачандрана перешли в допустимую область (см. *рис. 10*).

Таблица 6. Сравнение значений параметров качества старой (PDB) и новой (PDB_REDO) моделей.

Метрика	PDB	PDB_REDO
R	0,1759	0,1872
R_{free}	0,2539	0,2296
$R - R_{free}$	0,078	0,0424

Таблица 7. Показатели качества новой модели по сравнению со средними показателями структур с тем же разрешением.

All-Atom Contacts	Clashscore, all atoms:	2.37		100 th percentile* (N=331, 2.40Å ± 0.25Å)
	Poor rotamers	41	5 %	Goal: <0.3%
Protein Geometry	Favored rotamers	720	89 %	Goal: >98%
	Ramachandran outliers	0	0 %	Goal: <0.05%
	Ramachandran favored	961	98 %	Goal: >98%
	MolProbity score [^]	1.62		99 th percentile* (N=8058, 2.40Å ± 0.25Å)
	Cβ deviations >0.25Å	7	1 %	Goal: 0
	Bad bonds:	0 / 8006	0 %	Goal: 0%
	Bad angles:	6 / 10844	0 %	Goal: <0.1%

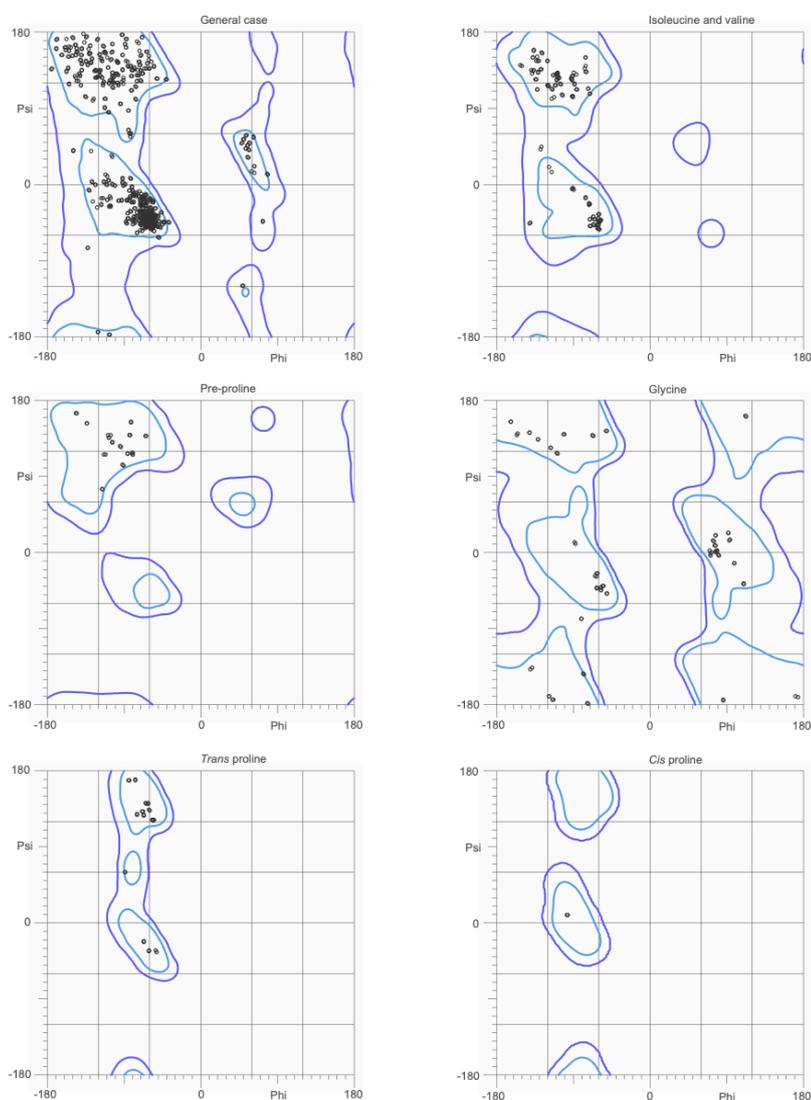


Рис. 10. Карты Рамачандрана, полученные сервисом MolProbity [4], для разных типов остатков новой модели: всех остатков, кроме названных далее; изолейцина и валина; остатков перед пролином; глицина; транс-пролина; цис-пролина.

Заклучение

Качество исходной модели было ниже среднего. После построения новой модели ситуация значительно улучшилась, так что теперь модель характеризуется, можно сказать, практически наилучшим из возможных по таким данным качеством.

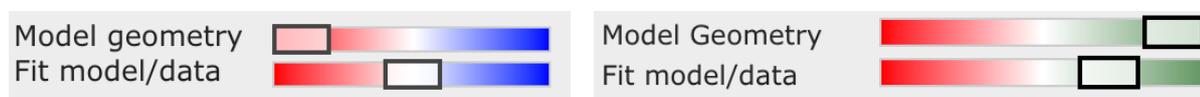


Рис. 11. Умозрительное сравнение качества моделей PDB (слева) и PDB_REDO (справа).

Список литературы

- [1] Koga, Y. , Katsumi, R. , You, D. , Matsumura, H. , Takano, K. and Kanaya, S. (2008), Crystal structure of highly thermostable glycerol kinase from a hyperthermophilic archaeon in a dimeric form. The FEBS Journal, 275: 2632-2643. doi:10.1111/j.1742-4658.2008.06410.x
- [2] Koga Y, Morikawa M, Haruki M, Nakamura H, Imanaka T & Kanaya S (1998) Thermostable glycerol kinase from a hyperthermophilic archaeon: gene cloning and characterization of the recombinant enzyme. Protein Eng 11, 1219–1227
- [3] Wlodawer, Alexander, et al. "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures." Febs Journal 275.1 (2008): 1-21.
- [4] <http://molprobity.biochem.duke.edu/>
- [5] https://proteopedia.org/wiki/index.php/Ramachandran_Plots
- [6] http://reference.iucr.org/dictionary/Free_R_factor
- [7] <https://eds.bmc.uu.se/cgi-bin/eds/rama?pdbCode=2zf5>
- [8] <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/r-value-and-r-free>
- [9] <http://www.ebi.ac.uk/pdbe/modval6>
- [10] https://files.rcsb.org/pub/pdb/validation_reports/zf/2zf5/2zf5_full_validation.pdf
- [11] <https://pdb-redo.eu/db/2zf5>