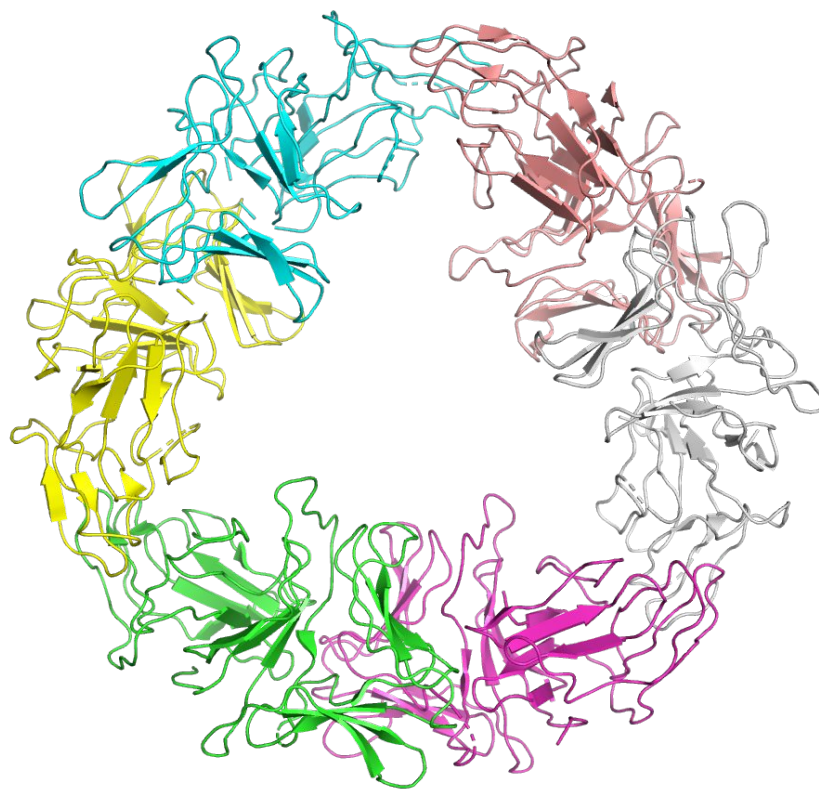


Validation.

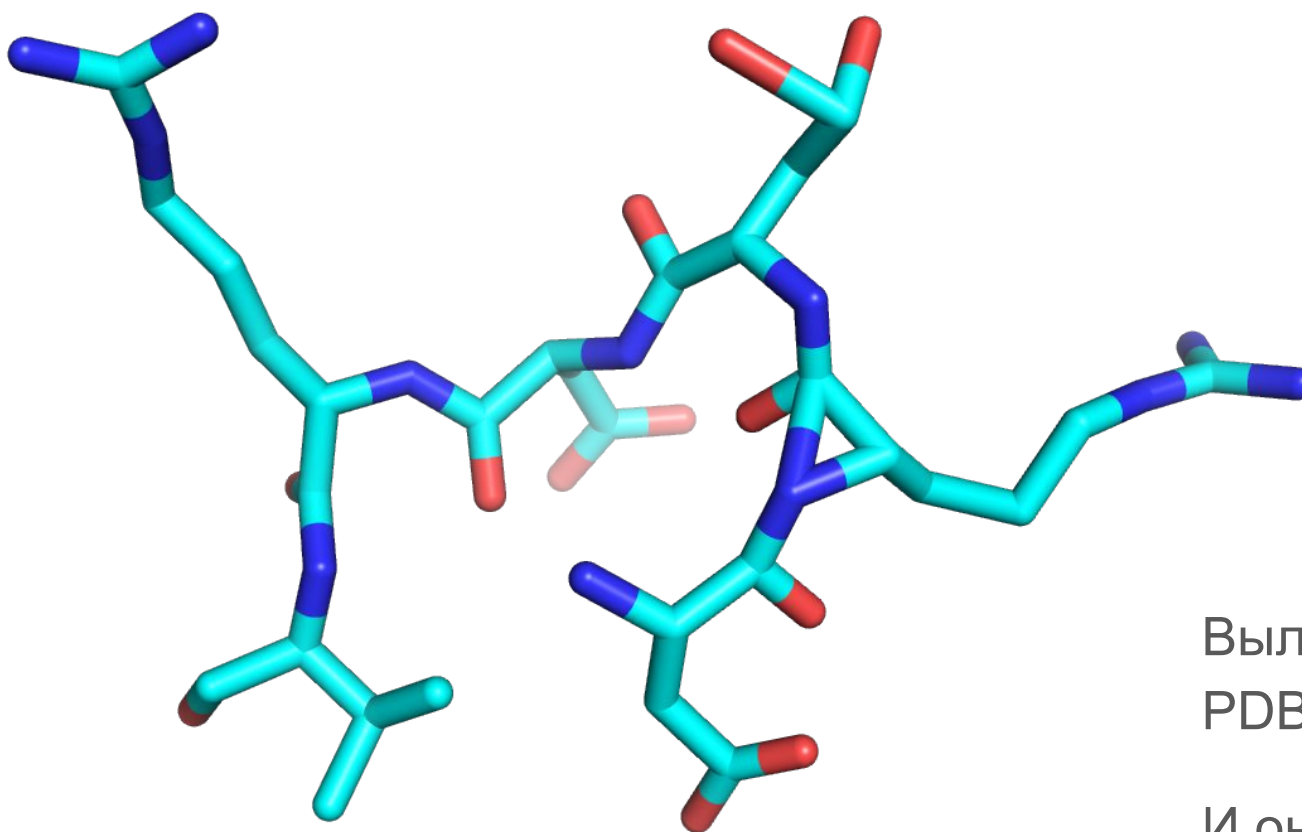
Показатели качества РСА-  
модели

Зачем вообще нужны структуры биологических макромолекул?



# Откровенная лажа в PDB

Fetuin-binding protein: модель имеет неадекватные с точки зрения химии длины связей и углы



**1DLP**  
**3.30 Å**

Выложена в  
PDB в 1999г.

И она всё ещё  
там.

# Валидация - контроль адекватности результатов

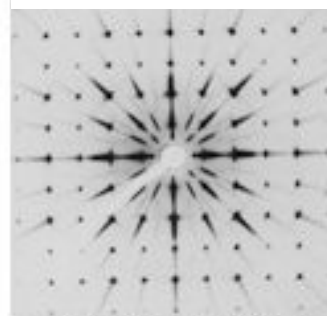
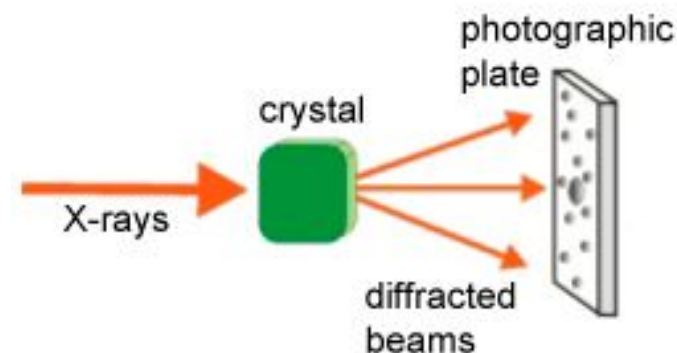
В терминах РСА:

- Контроль качества экспериментальных данных
- Контроль соответствия модели и экспериментальной электронной плотности
- Контроль соответствия модели нашим представлениям о физико-химических свойствах молекул



# РСА эксперимент

1. Получение кристалла белка
2. Получение дифракционной картины
3. Вычисление модулей структурных факторов ( $F$ )
4. Определение фаз (решение фазовой проблемы) ( $\varphi$ )
5. Построение функции электронной плотности в ячейке
6. Интерпретация электронной плотности, построение черновой модели
7. Сравнение электронной плотности модели с экспериментальной, оптимизация модели
8. Финальная модель

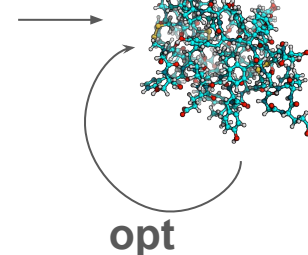


→ **F**

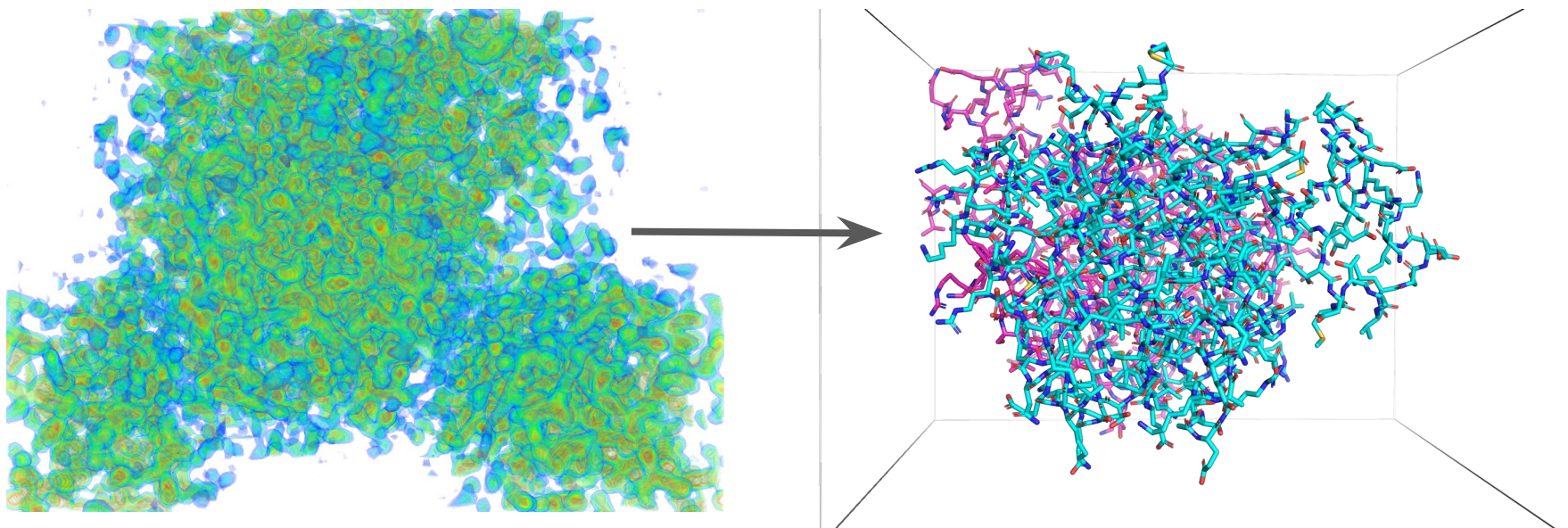
**SIR/MIR/AD/MAD/...** →  **$\varphi$**

**F,  $\varphi$**  →  **$\rho(x,y,z)$**

**$\rho(x,y,z)$**



Важно! Из РСА эксперимента получается ЭП,  
которая потом интерпретируется в координаты  
атомов



# Где можно ошибиться?

1. Получение кристалла белка
2. Получение дифракционной картины
3. Вычисление модулей структурных факторов (F)
4. Определение фаз структурных факторов ( $\phi$ )
5. Построение функции электронной плотности в ячейке
6. Интерпретация электронной плотности, построение черновой модели
7. Сравнение электронной плотности модели с экспериментальной, оптимизация модели
8. Финальная модель

Nunn et al., 1995.  
Бактериоферритин  
вместо LH1

Hoier et al., 1994.  
Ошибка в  
определении группы  
симметрии

'great pentaretraction'  
5 статей по ABC-  
транспортерам.  
 $I(h, k, l) \rightarrow |F(-h, -k, -l)|$ ,  
 $I(-h, -k, -l) \rightarrow |F(h, k, l)|$

# Где можно ошибиться?

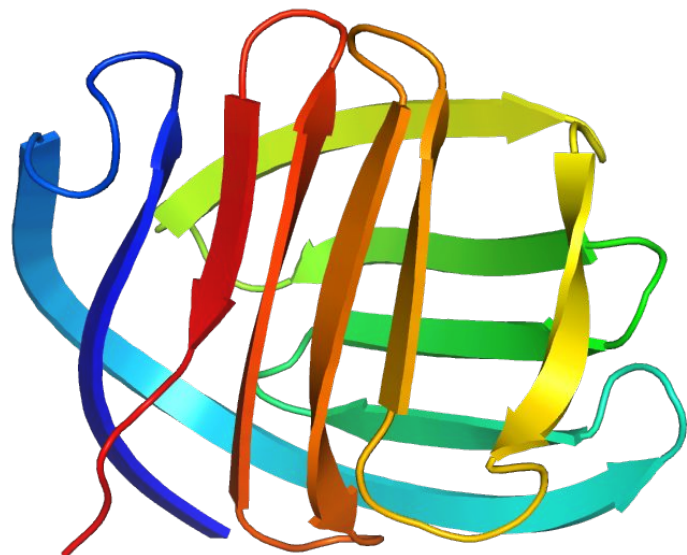
1. Получение кристалла белка
2. Получение дифракционной картины
3. Вычисление модулей структурных факторов (F)
4. Определение фаз структурных факторов ( $\varphi$ )
5. Построение функции электронной плотности в ячейке
- 6. Интерпретация электронной плотности, построение черновой модели**
- 7. Сравнение электронной плотности модели с экспериментальной, оптимизация модели**
- 8. Финальная модель**

Given the same data, no two crystallographers will ever produce identical final models. Their different biases and skill and experience levels will manifest themselves especially during manual model building but also during model refinement (e.g. different ways to parameterize a model and the use of different refinement programs and protocols).

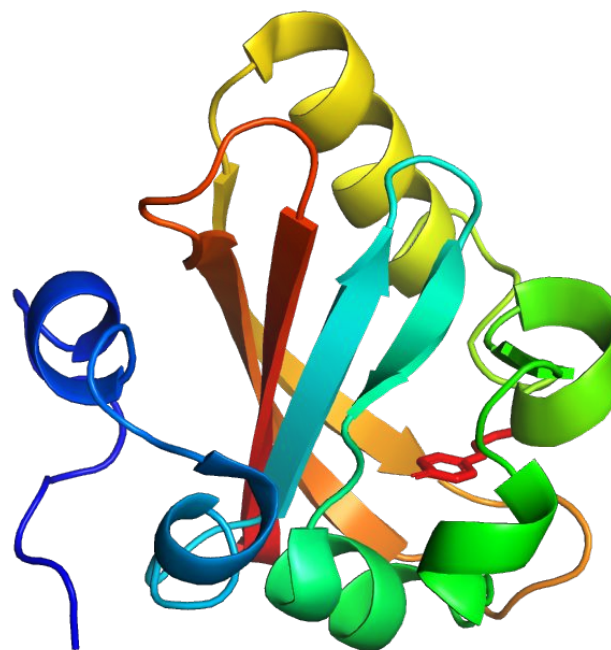
← ←  
Подвержены  
субъективному влиянию  
исследователя

# Примеры ошибок

Photoactive yellow protein: структура полностью ошибочна



**1PHY**  
2.4 Å, 1989

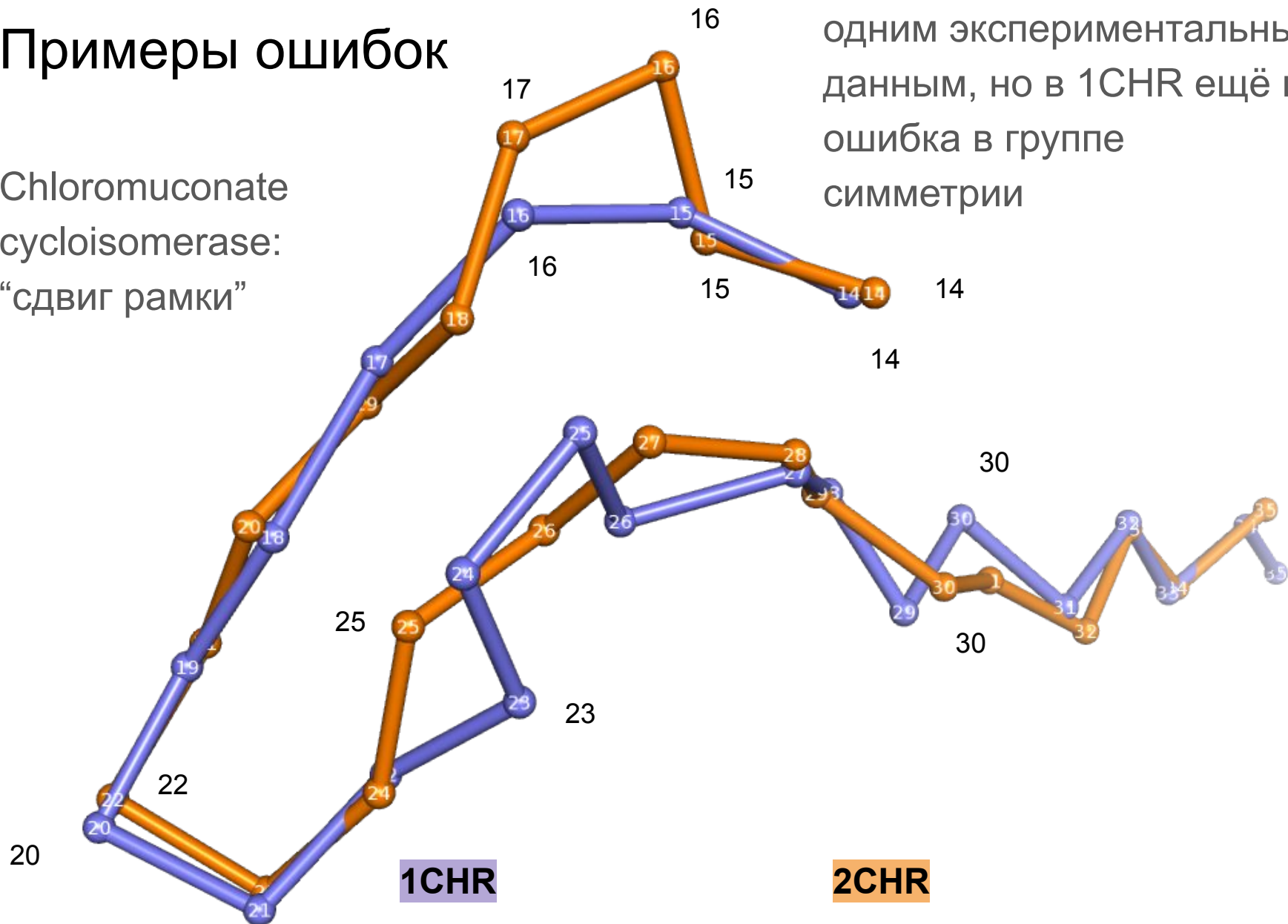


**2PHY**  
1.4 Å, 1995

# Примеры ошибок

Chloromuconate  
cycloisomerase:  
“сдвиг рамки”

Модели построены по  
одним экспериментальным  
данным, но в 1CHR ещё и  
ошибка в группе  
симметрии



# Примеры ошибок

Chloromuconate  
cycloisomerase:  
“сдвиг рамки”

Модели построены по  
одним экспериментальным  
данным, но в 1CHR ещё и  
ошибка в группе  
симметрии

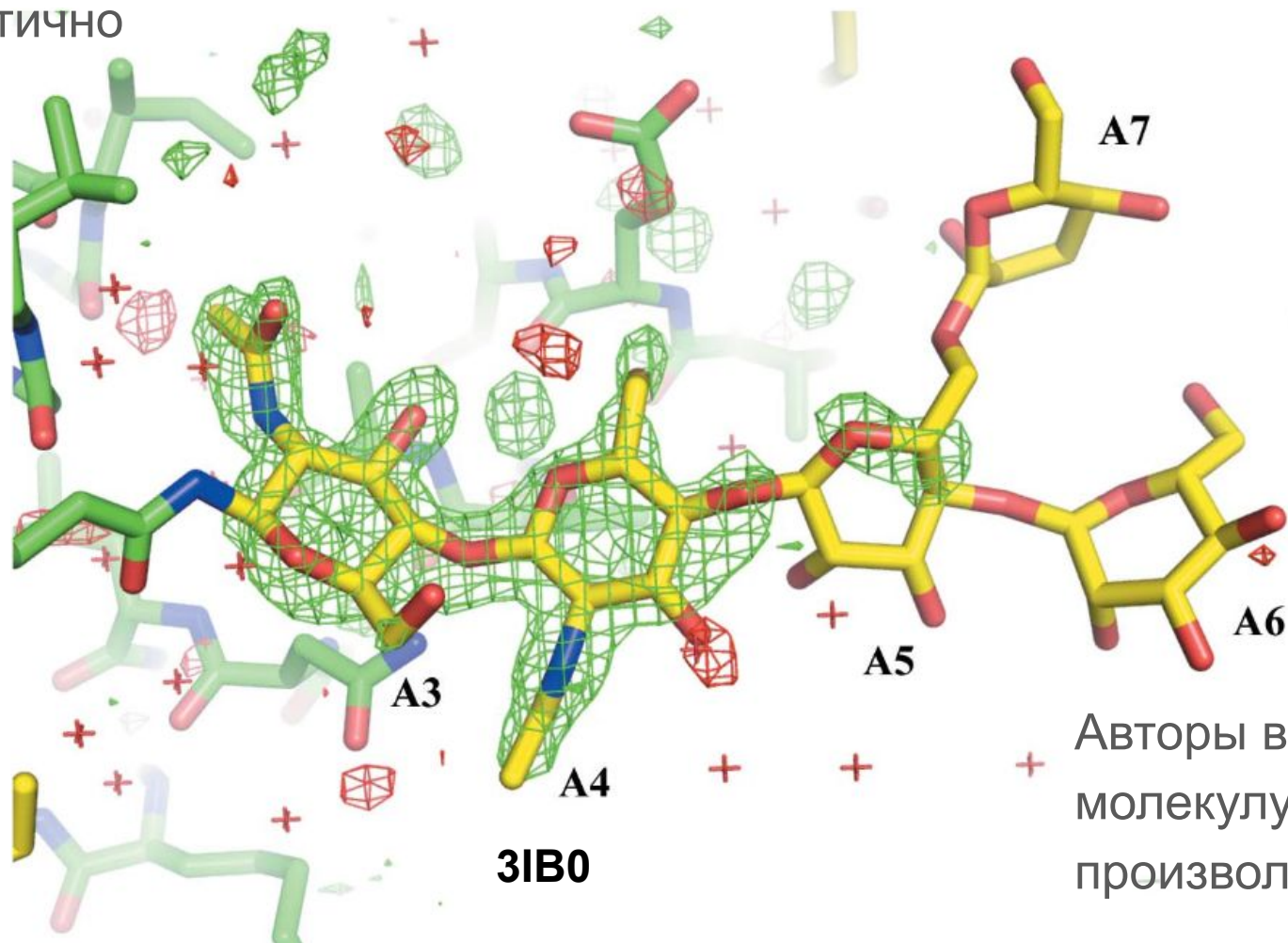
```
          *           20           *           40           *           60           *
1chr : MKIDAIEAVIVDVPT-KR-PIQMSITTVH QSYVIVRVYSE GLVGVGEGGSSVGGPVM SAECAETIKIIVERYL : 71
2chr : MKIDAIEAVIVDVPTKRPIQMSITT-VHQ QSYVIVRVYSE-GLVGVGEGGSSVGGPVM SAECAETIKIIVERYL : 71

          80           *           100
1chr : APHLLGTDAFNVSGALQTMARAVTGNA : 98
2chr : APHLLGTDAFNVSGALQTMARAVTGNA : 98
```



# Примеры ошибок

Гликозилированные остатки лактотрансферрина: модель вписана в экспериментальную электронную плотность лишь частично

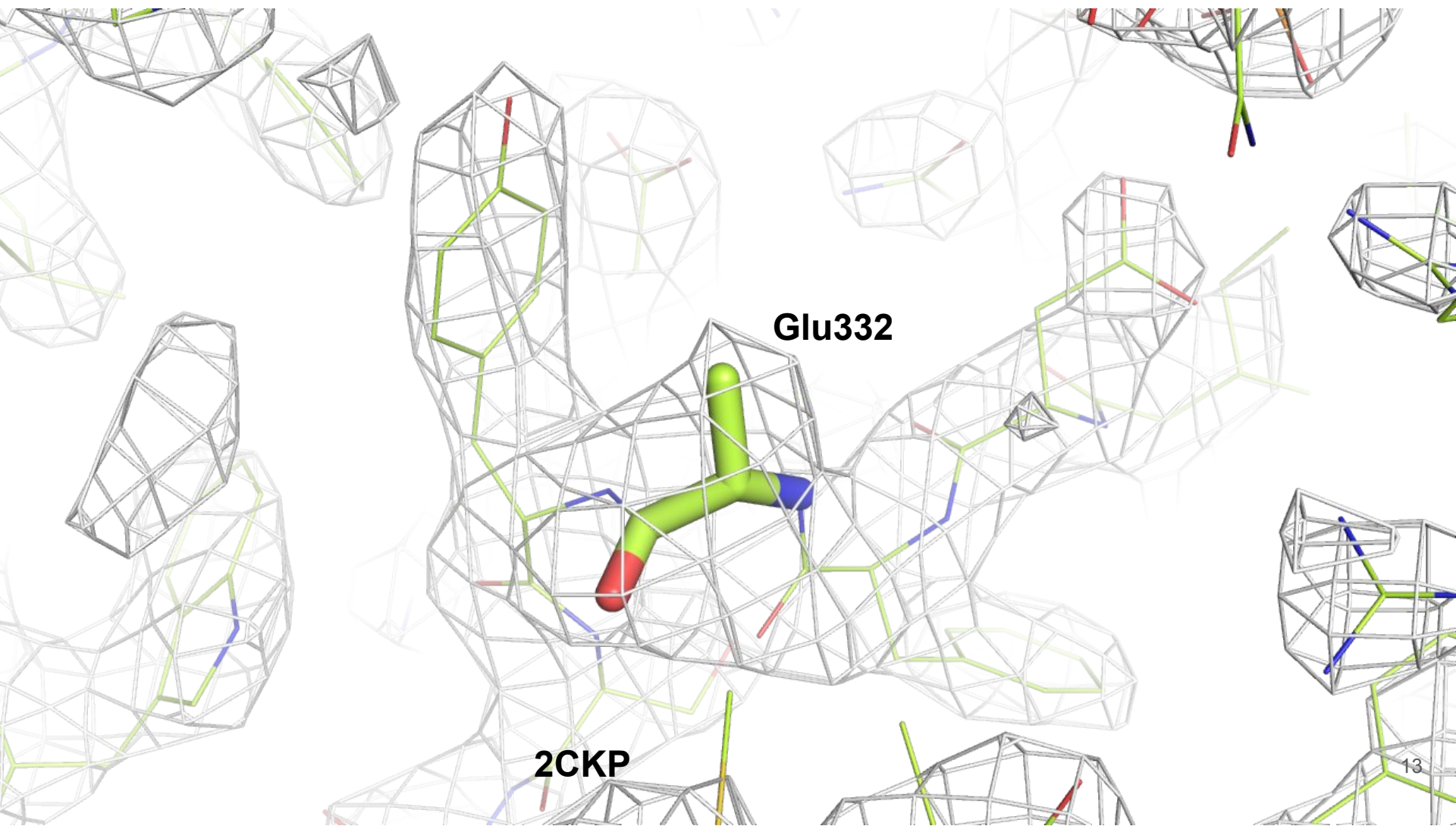


Авторы вписывают молекулу достаточно произвольно



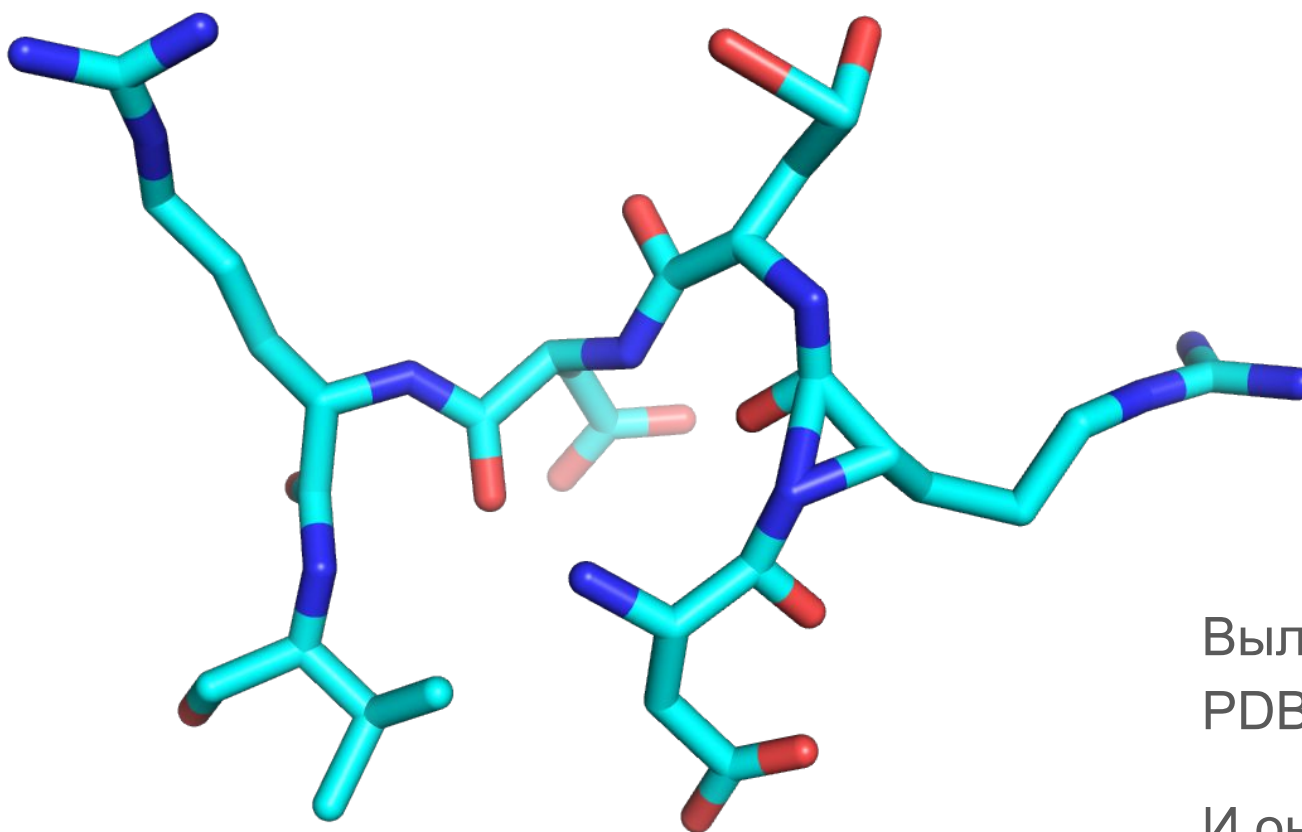
# Примеры ошибок

Недоразрешённые остатки



# Примеры ошибок

Fetuin-binding protein: модель имеет неадекватные с точки зрения химии длины связей и углы



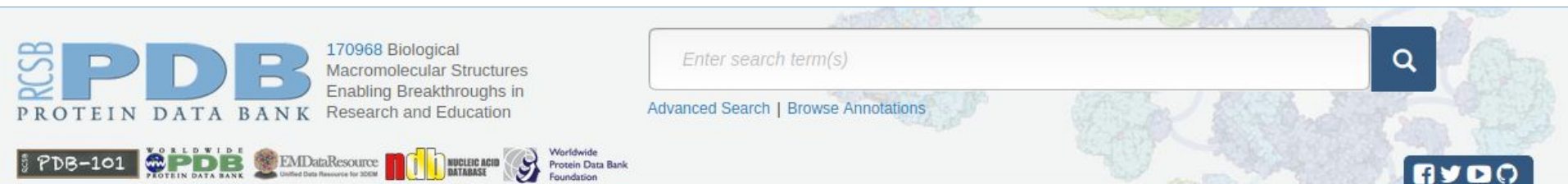
**1DLP**  
**3.30 Å**

Выложена в  
PDB в 1999г.

И она всё ещё  
там.

# Структура в PDB содержит ошибку. Что дальше?

- Структура и соответствующие статьи будут отозваны, если удастся опровергнуть правильность структуры



The screenshot shows the top navigation bar of the Protein Data Bank (PDB) website. On the left, there are logos for RCSB PDB, PDB-101, Worldwide PDB, EMDataResource, Nucleic Acid Database, and Worldwide Protein Data Bank Foundation. The main header features the text "170968 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education". A search bar with the placeholder "Enter search term(s)" and a magnifying glass icon is on the right. Below the search bar are links for "Advanced Search" and "Browse Annotations". Social media icons for Facebook, Twitter, YouTube, and LinkedIn are in the bottom right corner.

## 2OU1

Structures of apolipoprotein A-II and a lipid surrogate complex provide insights into apolipoprotein-lipid interactions

DOI: [10.2210/pdb2OU1/pdb](https://doi.org/10.2210/pdb2OU1/pdb)

Deposited: 2007-02-09 Released: 2007-03-06

Deposition Author(s): Kumar, M.S., Carson, M.C., Hussain, M.M., Murthy, H.M.K.

Entry **2OU1** was removed from the distribution of released PDB entries (status Obsolete) on **2018-05-09**.

Details:

### UAB Statement on Protein Data Bank Issues

Published on December 08, 2009

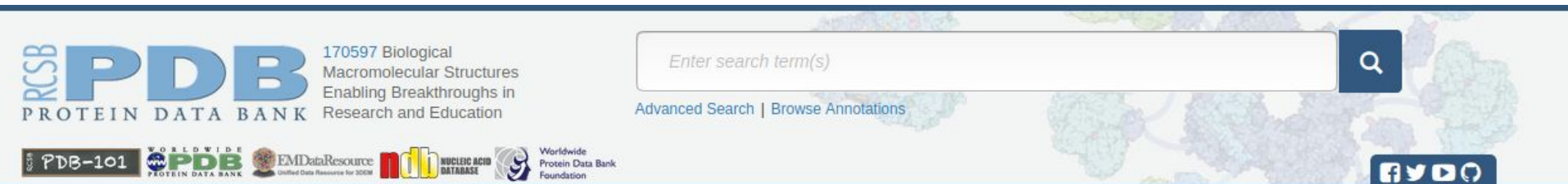
The University of Alabama at Birmingham has requested that the **Research Collaboratory for Structural Bioinformatics Protein Data Bank** remove certain protein structure files deposited by a former UAB employee. UAB also has identified nine publications related to the same protein structures that should be retracted from various scientific journals, and is making those journals aware of this matter.



[https://en.wikipedia.org/wiki/H.M.\\_Krishna\\_Murthy](https://en.wikipedia.org/wiki/H.M._Krishna_Murthy)

# Структура в PDB содержит ошибку. Что дальше?

- Структура будет заменена более свежей и правильной или подвергнута ревизии



The screenshot shows the top navigation bar of the Protein Data Bank (PDB) website. On the left, the PDB logo is displayed with the text 'RCSB PDB PROTEIN DATA BANK' and '170597 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education'. To the right of the logo is a search bar with the placeholder text 'Enter search term(s)' and a magnifying glass icon. Below the search bar are links for 'Advanced Search' and 'Browse Annotations'. At the bottom of the header, there are several partner logos including PDB-101, Worldwide PDB, EMDataResource, Nucleic Acid Database, and Worldwide Protein Data Bank Foundation. On the far right, there are social media icons for Facebook, Twitter, YouTube, and LinkedIn.

1PHY

CRYSTALLOGRAPHIC STRUCTURE OF A PHOTORECEPTOR PROTEIN AT 2.4 ANGSTROMS RESOLUTION

DOI: [10.2210/pdb1PHY/pdb](https://doi.org/10.2210/pdb1PHY/pdb)

Deposited: 1989-08-04 Released: 1990-01-15

Deposition Author(s): Mcree, D.E., Tainer, J.A., Getzoff, E.D.

Entry **1PHY** was removed from the distribution of released PDB entries (status Obsolete) on **1995-10-15**.  
It has been replaced (superseded) by [2PHY](#).



# Структура в PDB содержит ошибку. Что дальше?

- Структура так и останется в PDB, потому что новых структур никто не делал, а для старых не выложены данные из эксперимента

RCSB  
**PDB**  
PROTEIN DATA BANK

170597 Biological  
Macromolecular Structures  
Enabling Breakthroughs in  
Research and Education

Enter search term(s)



[Advanced Search](#) | [Browse Annotations](#)



Structure Summary

3D View

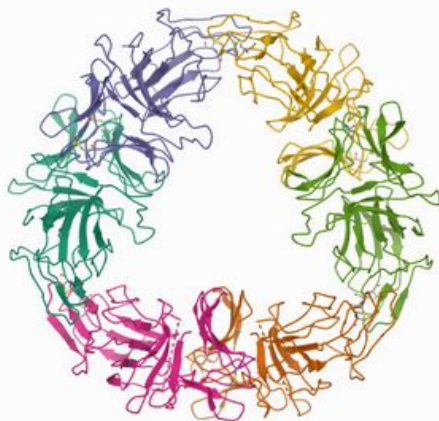
Annotations

Experiment

Sequence

Genome

Biological Assembly 1 ?



## 1DLP

STRUCTURAL CHARACTERIZATION OF THE NATIVE FETUIN-BINDING PROTEIN SCILLA CAMPANULATA AGGLUTININ (SCAFET): A NOVEL TWO-DOMAIN LECTIN

DOI: [10.2210/pdb1DLP/pdb](https://doi.org/10.2210/pdb1DLP/pdb)

Classification: **SUGAR BINDING PROTEIN**

Organism(s): *Hyacinthoides hispanica*

Mutation(s): No

Deposited: 1999-12-11 Released: 2000-02-10

Deposition Author(s): Wright, L.M., Reynolds, C.D., Rizkallah, P.J., Allen, A.K., VanDamme, E.J.M., Donovan, M.J., Peumans, W.J.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 3.30 Å

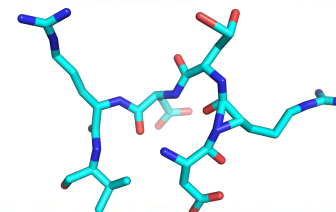
R-Value Free: 0.293

wwPDB Validation

3D Report

Full Report

Metric	Percentile Ranks	Value
Clashscore		84
Ramachandran outliers		16.4%



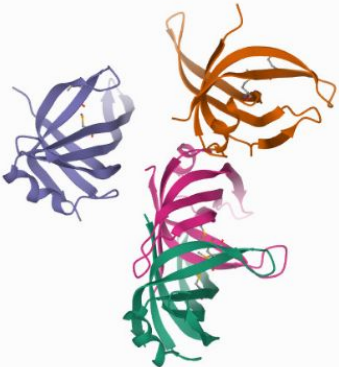
# wwPDB и валидация

Автоматическая система проверки wwPDB OneDep создаёт отчёт по каждой структуре

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

Structure Summary 3D View Annotations Experiment Sequence Genome

Biological Assembly 1 ?



**1WOC**  
Crystal structure of PriB  
DOI: 10.2210/pdb1WOC/pdb  
Classification: DNA BINDING PROTEIN  
Organism(s): Escherichia coli K-12  
Expression System: Escherichia coli BL21(DE3)  
Mutation(s): No ⓘ

Deposited: 2004-08-13 Released: 2005-01-25  
Deposition Author(s): Shioi, S., Ose, T., Maenaka, K., Abe, Y., Kohda, D., Katayama, T., Ueda, T.

Experimental Data Snapshot  
Method: X-RAY DIFFRACTION  
Resolution: 2.00 Å  
R-Value Free: 0.276  
R-Value Work: 0.221

wwPDB Validation 3D Report Full Report

Metric	Percentile Ranks	Value
Rfree		0.277
Clashscore		25
Ramachandran outliers		1.5%
Sidechain outliers		4.9%
RSRZ outliers		8.5%

Worse Better  
■ Percentile relative to all X-ray structures  
□ Percentile relative to X-ray structures of similar resolution

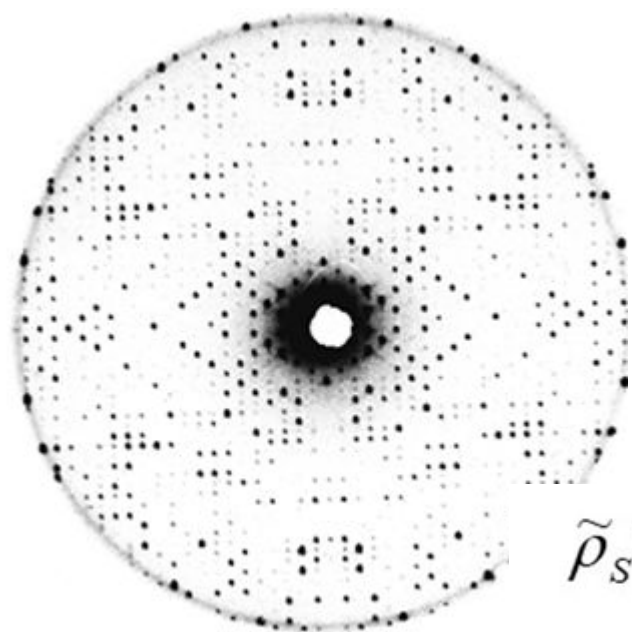
Find Similar Assemblies

Biological assembly 1 assigned by authors.

This is version 1.3 of the entry. See complete [history](#).

# Показатели качества модели в целом: Разрешение

- Вычисляется по набору структурных факторов и параметрам ячейки
- Характеризует совокупность экспериментальных данных – структурных факторов



$$\tilde{\rho}_S(x, y, z) \approx \sum_{(hkl) \in S} F_{hkl} \cos[2\pi(hx + ky + lz) - \varphi_{hkl}]$$

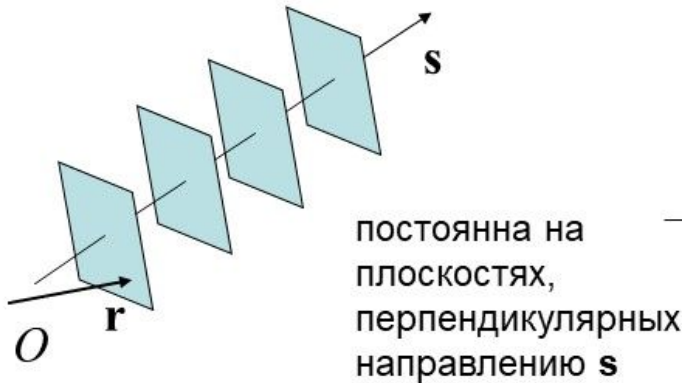
# Показатели качества модели в целом: Разрешение

Гармоника Фурье

$$H_s(\mathbf{r}) = \cos[2\pi(hx + ky + lz)] = \cos[2\pi(\mathbf{s}, \mathbf{r})]$$

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

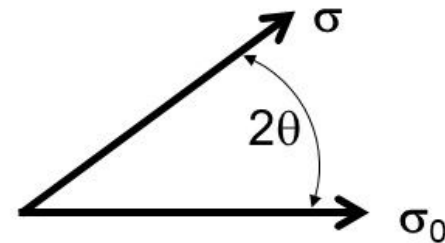
$$h = (\mathbf{s}, \mathbf{a}), k = (\mathbf{s}, \mathbf{b}), l = (\mathbf{s}, \mathbf{c})$$



**Разрешение**, соответствующее гармонике Фурье

$d = \frac{1}{|\mathbf{s}|}$  - расстояние между соседними максимумами в направлении  $\mathbf{s}$ ;

$$|\mathbf{s}| = \frac{2 \sin \theta}{\lambda} \quad d = \frac{\lambda}{2 \sin \theta}$$





# Показатели качества модели в целом:

## Разрешение

- Разрешение полного набора гармоник равно минимальному разрешению из всех гармоник набора
- В эксперименте получается измерить не все гармоники
- Полнота данных для данного разрешения - % гармоник с большим разрешением, которые удалось измерить в эксперименте
- Указывают также верхний порог разрешения - самое большое детектированное разрешение гармоники (почему так?)



Разрешение для данного набора можно принять таким:

1 Å с полнотой данных  $20/32 * 100\% = 62\%$

1.5 Å с полнотой  $19/22 * 100\% = 86\%$

2 Å с полнотой  $18/20 * 100\% = 90\%$

● Измеренный рефлекс

● Потерянный рефлекс

# Показатели качества модели в целом:

## Разрешение

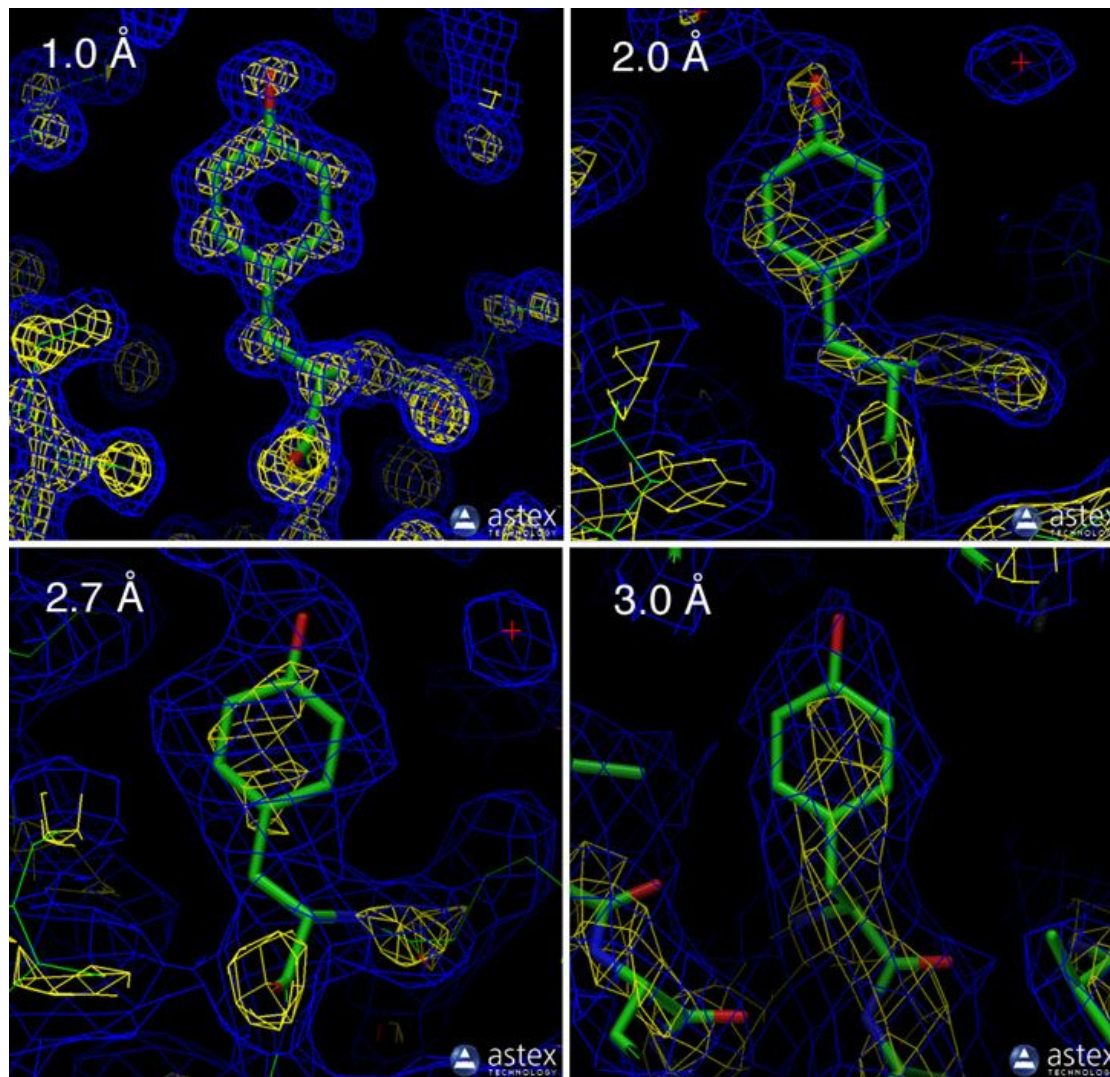
- Разрешение структуры определяется разрешением измеренной гармоника с самым малым разрешением и полнотой измеренного набора
- Есть небольшая доля субъективности в определении разрешения

```
REMARK 3 REFINEMENT TARGET : MAXIMUM LIKELIHOOD
REMARK 3
REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.36
REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) : 38.32
REMARK 3 DATA CUTOFF (SIGMA(F)) : NONE
REMARK 3 COMPLETENESS FOR RANGE (%) : 98.52
REMARK 3 NUMBER OF REFLECTIONS : 89715
REMARK 3
```

# Показатели качества модели в целом:

## Разрешение

- Чем лучше - меньше - разрешение (в Å), тем менее вероятны ошибки
- По данным PCA с плохим разрешением можно построить хорошую модель. При хорошем разрешении в модели могут быть ошибки



Тyr103 миоглобина из структур с разным разрешением

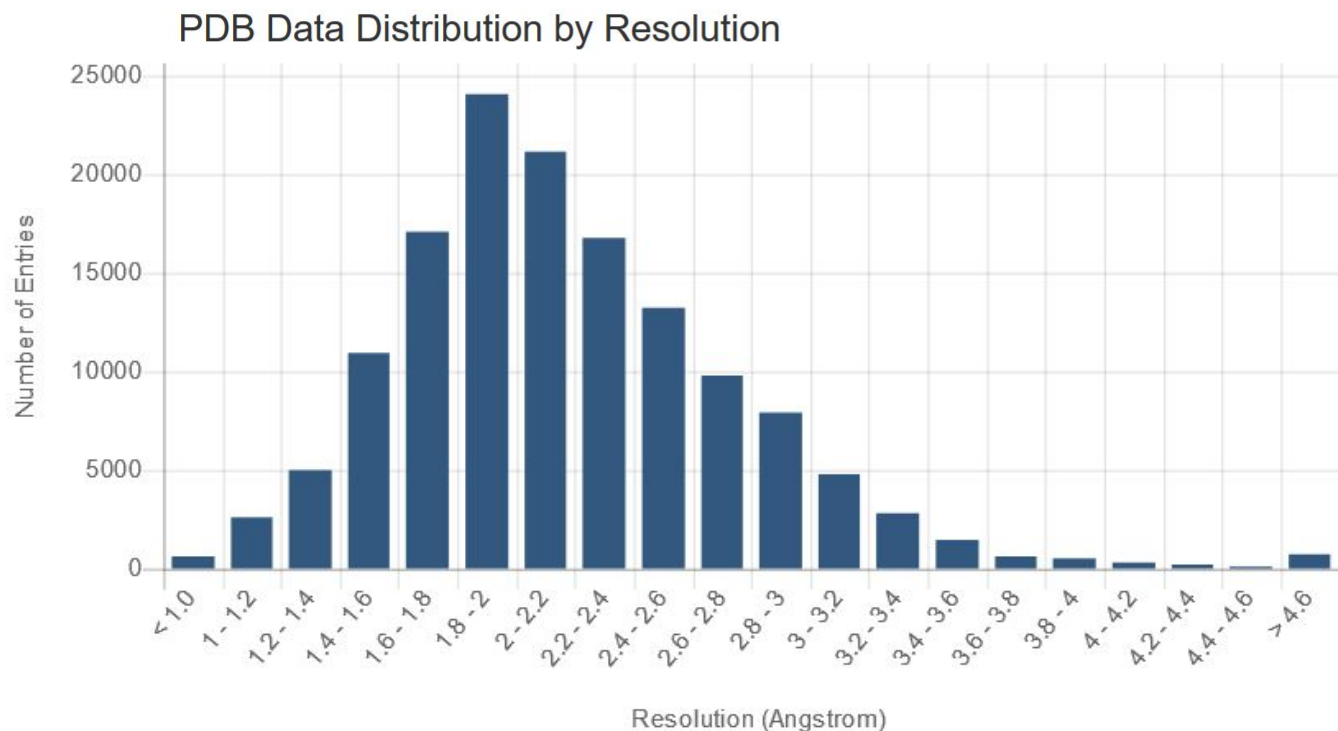
# Показатели качества модели в целом: Разрешение

Высокое разрешение:  $< 1.5 \text{ \AA}$

Хорошее разрешение:  $1.5 - 2.5 \text{ \AA}$

Удовлетворительное:  $2.5 - 3.5 \text{ \AA}$

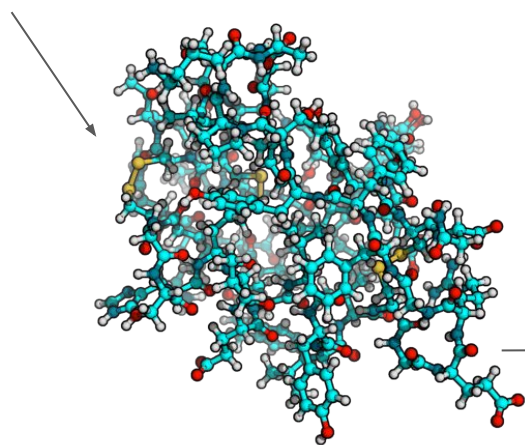
Низкое:  $> 3.5 \text{ \AA}$



# Показатели качества модели в целом: R-фактор

- Вычисляется по оптимизированной модели и измеренным структурным факторам
- Характеризуют соответствие модели экспериментальным данным – структурным факторам

**$F_{obs}(h,k,l)$**  - измерены в эксперименте



модель

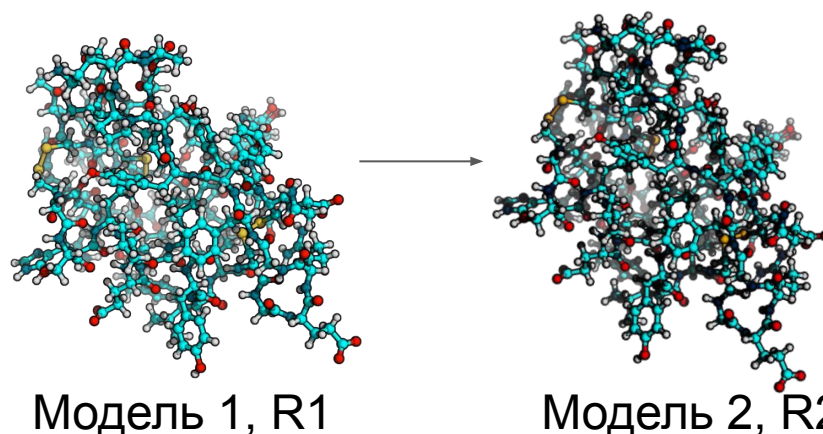
$$R = \frac{\sum_{hkl} |F_{hkl}^{calc} - F_{hkl}^{obs}|}{\sum_{hkl} F_{hkl}^{obs}} \neq 100\%$$

**$F_{calc}(h,k,l)$**  -  
рассчитаны из модели

# Показатели качества модели в целом: R-фактор

- В процессе оптимизации модели уточняется:
  - соответствие  $F_{obs}(h,k,l)$  и  $F_{calc}(h,k,l)$
  - соответствие длин связей и валентных углов модели физико-химическим свойствам молекул
- Минимизируем составной R-фактор:

$$R_{mixed} = w_{X-ray} R_{X-ray} + w_{dist} R_{dist} + w_{angle} R_{angle}$$



if  $R2 < R1$  -> M2, else: M1



# Показатели качества модели в целом: R-фактор

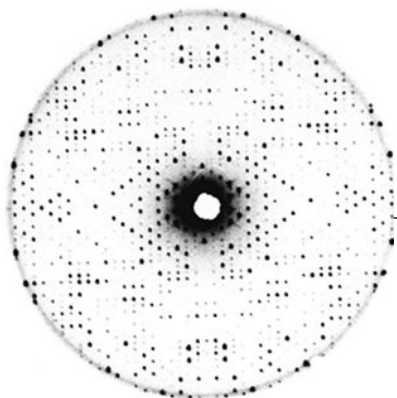
- Чем ниже R-фактор, тем больше структурные факторы модели соответствуют экспериментальным
- Проблема: переоптимизация. Можно подогнать даже очень неправильную модель.

**Table 1.** Comparison of some model and refinement statistics of CRABP II (intentionally traced backwards; model X), CRABP I refined conservatively (with NCS constraints and grouped temperature factors; model Y), and the CRABP I structure refined according to today's 'standard' refinement practices (model Z).

Model	X	Y	Z
Resolution range (Å)	8.0–3.0	8.0–2.9	6.0–2.9
R	0.214	0.251	0.169

# Показатели качества модели в целом:

$R_{\text{free}}$



h	k	l	F
3	5	5	207.9
3	5	6	255.9
3	5	7	328.1
3	5	8	298.8
3	5	9	583.4
3	5	10	833.9
3	5	11	1403.0

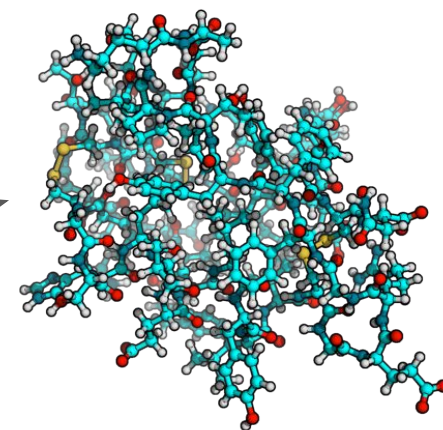
Набор измеренных рефлексов

h	k	l	F
3	5	5	207.9
3	5	6	255.9

Контрольный набор

h	k	l	F
3	5	7	328.1
3	5	8	298.8
3	5	9	583.4
3	5	10	833.9
3	5	11	1403.0

Рабочий набор



МОДЕЛЬ



## Показатели качества модели в целом:

$R_{\text{free}}$

- Вычисляется по той же формуле, что и R-фактор, но только для  $F_{\text{obs}}$  контрольных рефлексов и  $F_{\text{calc}}$  финальной модели после оптимизации
- Если модель правильная, то  $R_{\text{free}}$  окажется примерно равным R-фактору или немногим больше
- Если  $R_{\text{free}}$  сильно больше R-фактора, модель переоптимизирована

$$R = \frac{\sum_{hkl} |F_{hkl}^{\text{calc}} - F_{hkl}^{\text{obs}}|}{\sum_{hkl} F_{hkl}^{\text{obs}}} \neq 100\%$$

- Хорошие значения:  $R_{\text{free}} < 20\%$
- Плохие значения:  $R_{\text{free}} > 40\%$
- Значения  $(R_{\text{free}} - R) > 10\%$  настораживают в отношении переоптимизации (overfitting)

# Показатели качества модели в целом:

$R_{\text{free}}$

```
REMARK      3
REMARK      3  FIT TO DATA USED IN REFINEMENT.
REMARK      3  CROSS-VALIDATION METHOD           : THROUGHOUT
REMARK      3  FREE R VALUE TEST SET SELECTION       : RANDOM
REMARK      3  R VALUE          (WORKING + TEST SET)  : 0.15621
REMARK      3  R VALUE          (WORKING SET)        : 0.15185
REMARK      3  FREE R VALUE                               : 0.19471
REMARK      3  FREE R VALUE TEST SET SIZE (%)       : 10.1
REMARK      3  FREE R VALUE TEST SET COUNT         : 5989
REMARK      3
```

- Авторы получили в эксперименте 59 297 рефлексов
- Они утверждают, что 10.1% данных спрятали в сейф
- По оставшимся 53 308 рефлексам оптимизировали модель и получили  $R=15\%$
- После этого достали тайные рефлекссы из сейфа и рассчитали R-фактор по ним. Авторы получили  $R_{\text{free}} = 19\%$ .

# Показатели качества модели в целом:

$R_{\text{free}}$

**Table 1.** Comparison of some model and refinement statistics of CRABP II (intentionally traced backwards; model X), CRABP I refined conservatively (with NCS constraints and grouped temperature factors; model Y), and the CRABP I structure refined according to today's 'standard' refinement practices (model Z).

Model	X	Y	Z
Resolution range (Å)	8.0–3.0	8.0–2.9	6.0–2.9
R	0.214	0.251	0.169
$R_{\text{free}}$	0.617	0.320	0.323

# Показатели качества отдельных остатков

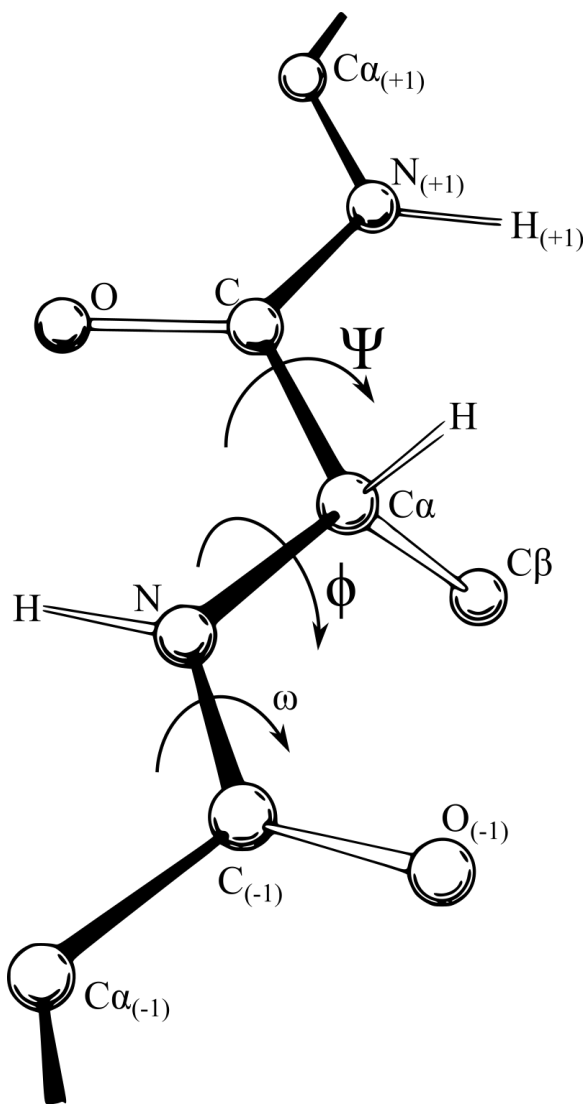
- Торсионные углы (Карты Рамачандрана и ротамеры боковых цепей)
- RSR
- Комфортность окружения остатка

Остатки с нетипичными значениями показателей - **маргинальные**.

Маргинальность остатка - ошибка расшифровки либо функционально значимая особенность.

**% маргинальных остатков** - ещё одна характеристика качества модели в целом

# Показатели качества отдельных остатков: укладка полипептидной цепи



Некоторые значения геометрических параметров энергетически выгодны, и потому встречаются часто. При оптимизации модели геометрические параметры подгоняются под такие табличные значения.

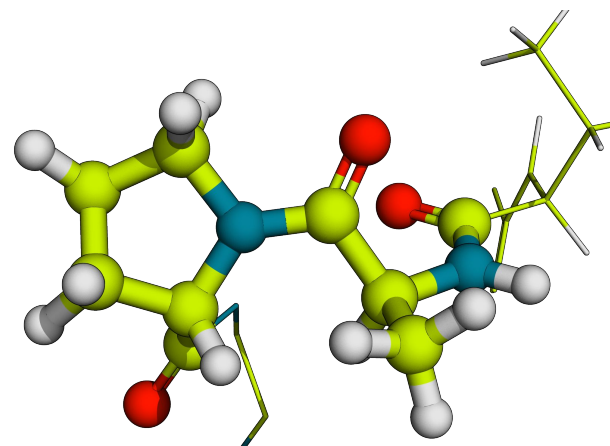
Укладка полипептидной цепи определяется тремя торсионными углами  $\phi$ ,  $\psi$ ,  $\omega$

(Торсионный угол принимает значения от  $-180^\circ$  до  $180^\circ$ )

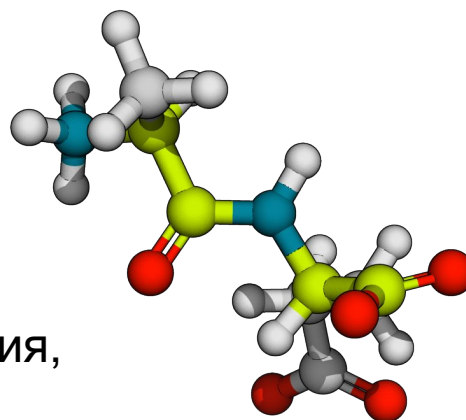
# Показатели качества отдельных остатков: укладка полипептидной цепи

$\omega$  принимает всего два значения -  $0^\circ$  или  $180^\circ$  (почему?)

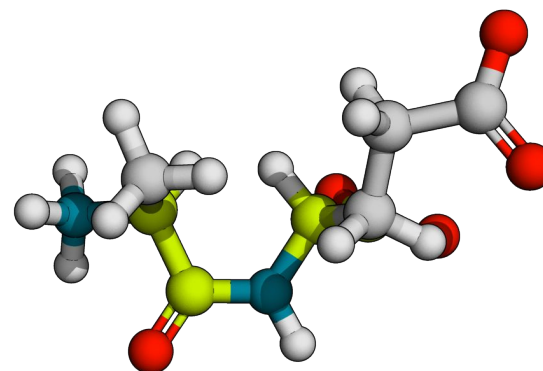
Транс-конформация наиболее частая для всех остатков кроме пролина.



Цис-конформация,  $\omega=0^\circ$

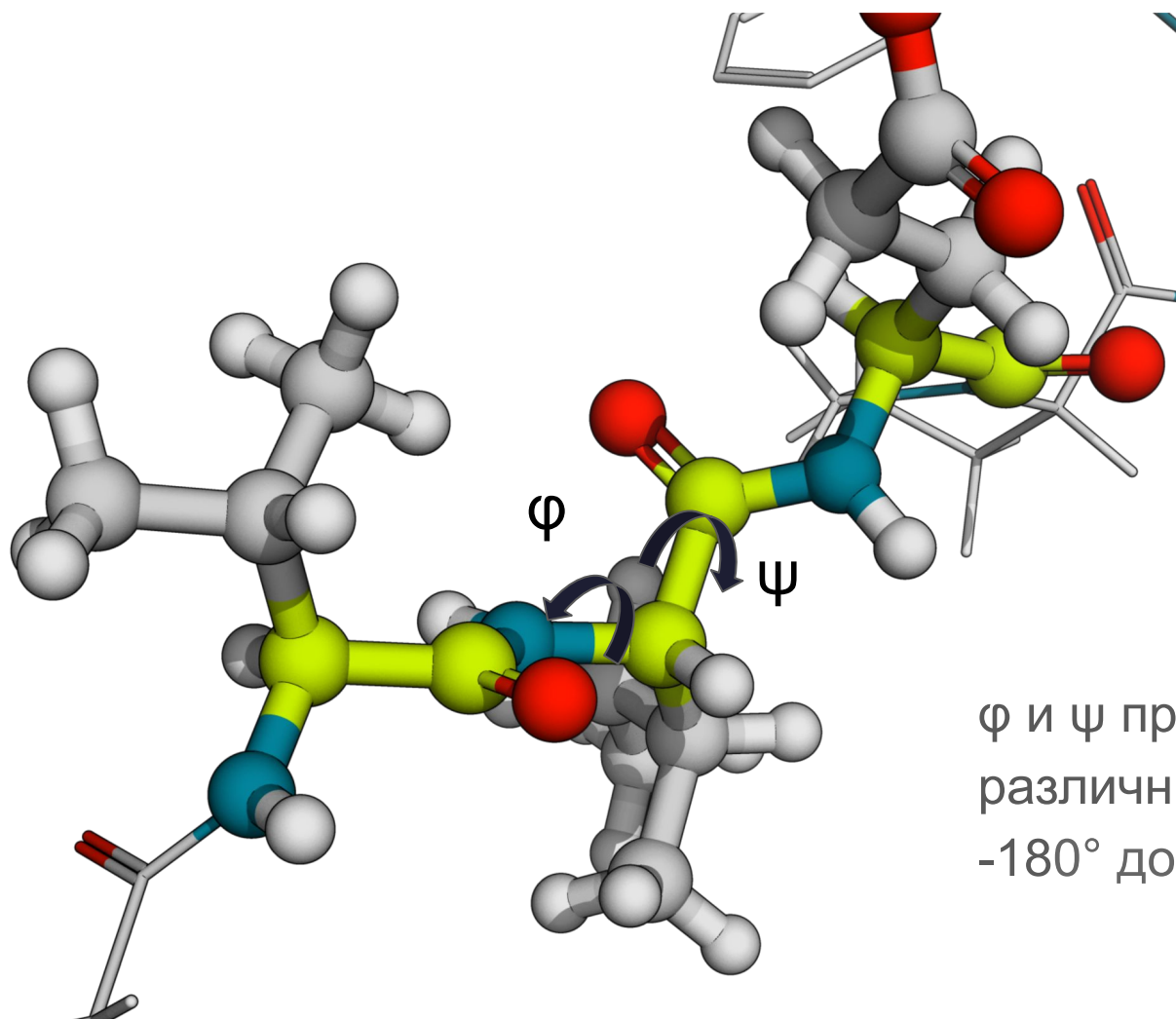


Транс-конформация,  
 $\omega=180^\circ$



Цис-конформация,  $\omega=0^\circ$

# Показатели качества отдельных остатков: карты Рамачандрана для $\phi$ и $\psi$



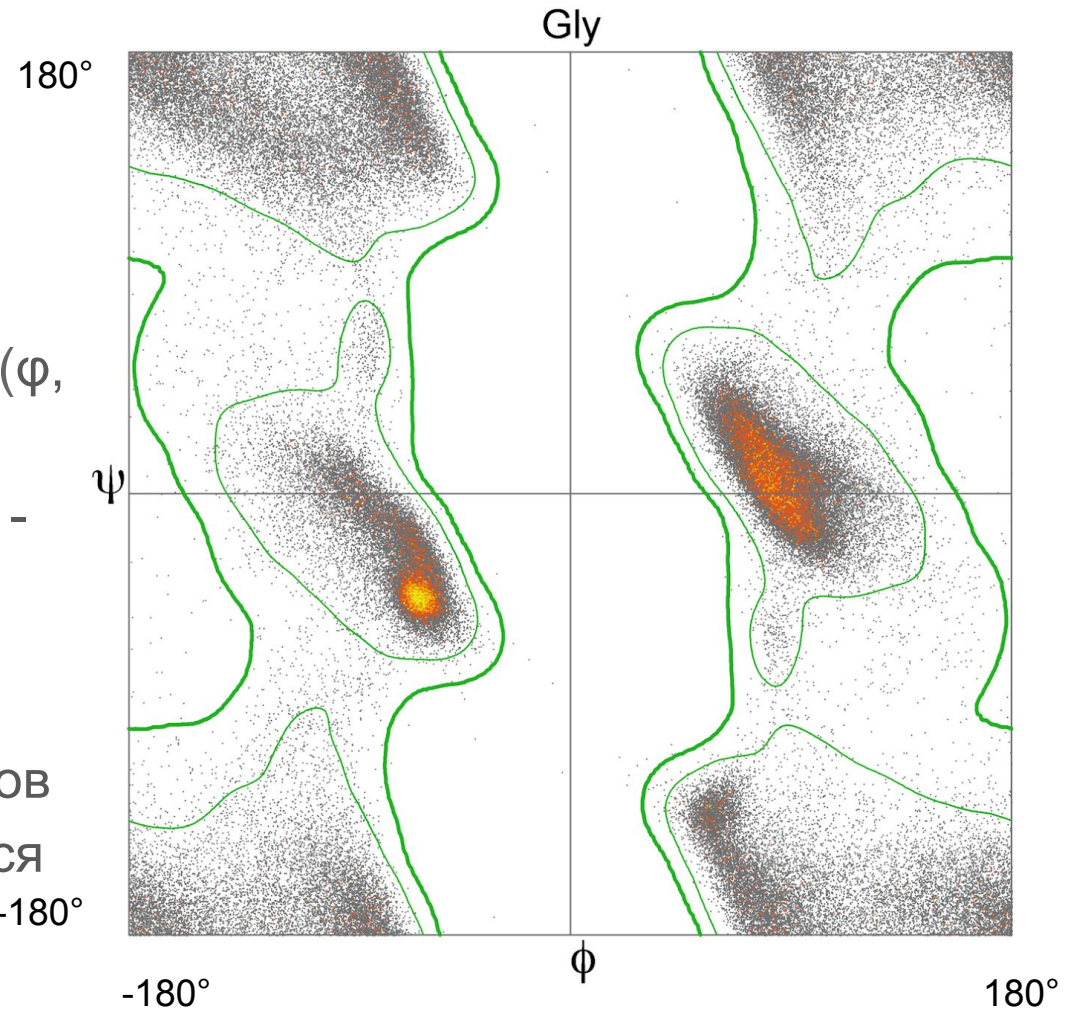
$\phi$  и  $\psi$  принимают  
различные значения от  
 $-180^\circ$  до  $180^\circ$

# Показатели качества отдельных остатков: карты Рамачандрана для $\phi$ и $\psi$

$\phi$  и  $\psi$  принимают различные значения от  $-180^\circ$  до  $180^\circ$ .

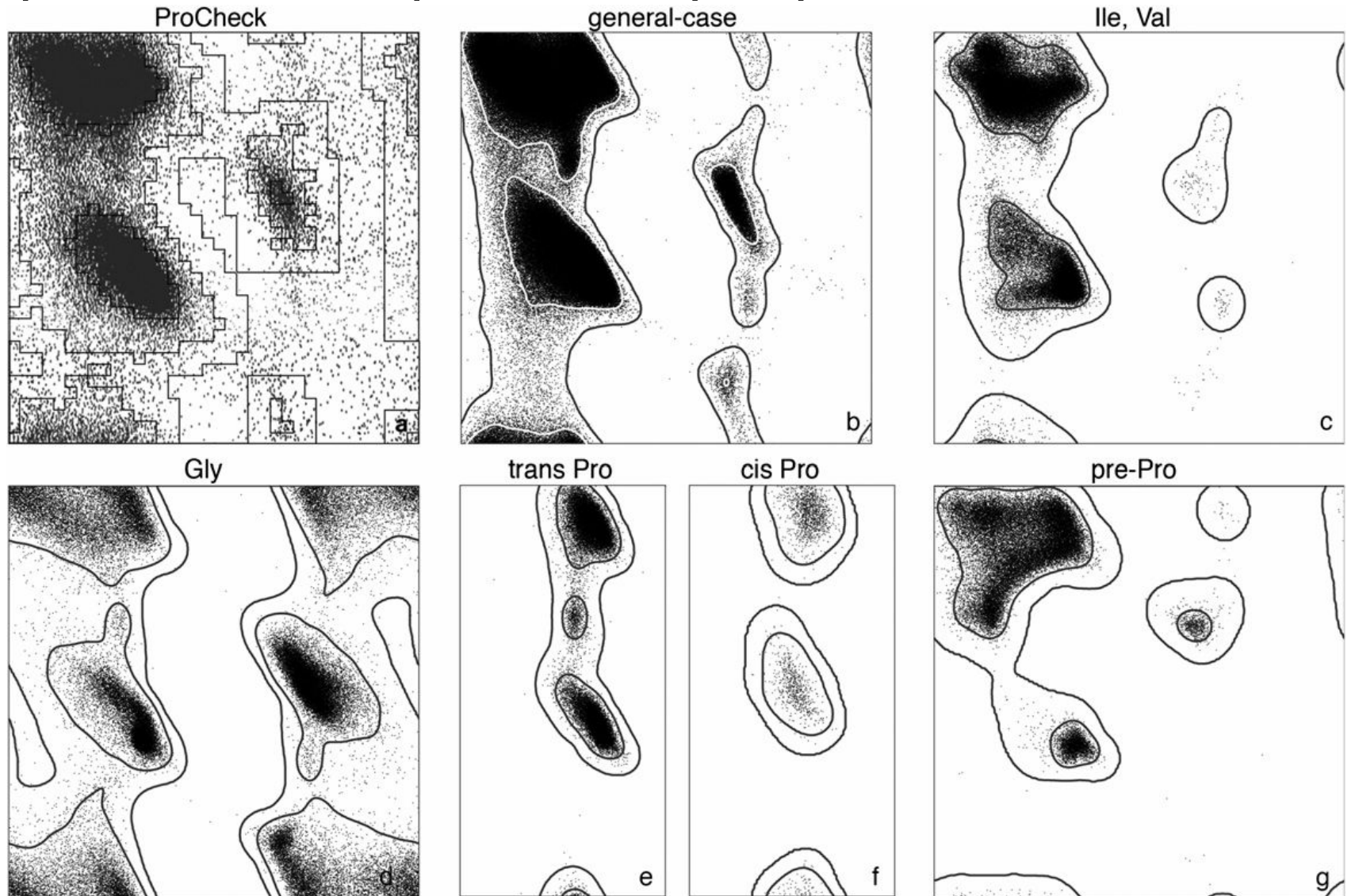
Если для каждой аминокислоты на плоскости ( $\phi$ ,  $\psi$ ) изобразить все наблюдаемые конформации - получится карта Рамачандрана.

Карты для отдельных остатков и для общих случаев строятся на основе статистики базы PDB.





# Показатели качества отдельных остатков: карты Рамачандрана для $\phi$ и $\psi$

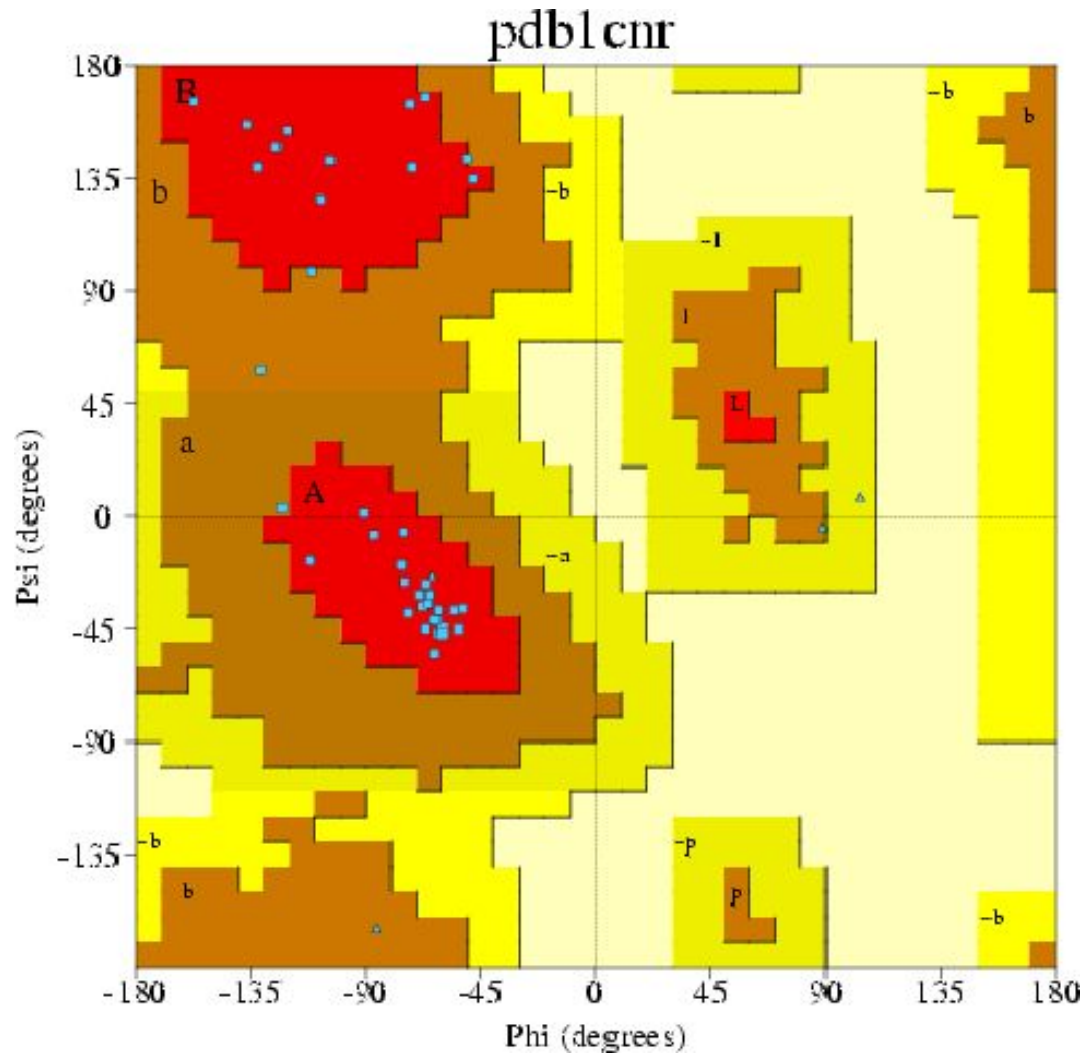


# Показатели качества отдельных остатков: карты Рамачандрана для $\phi$ и $\psi$

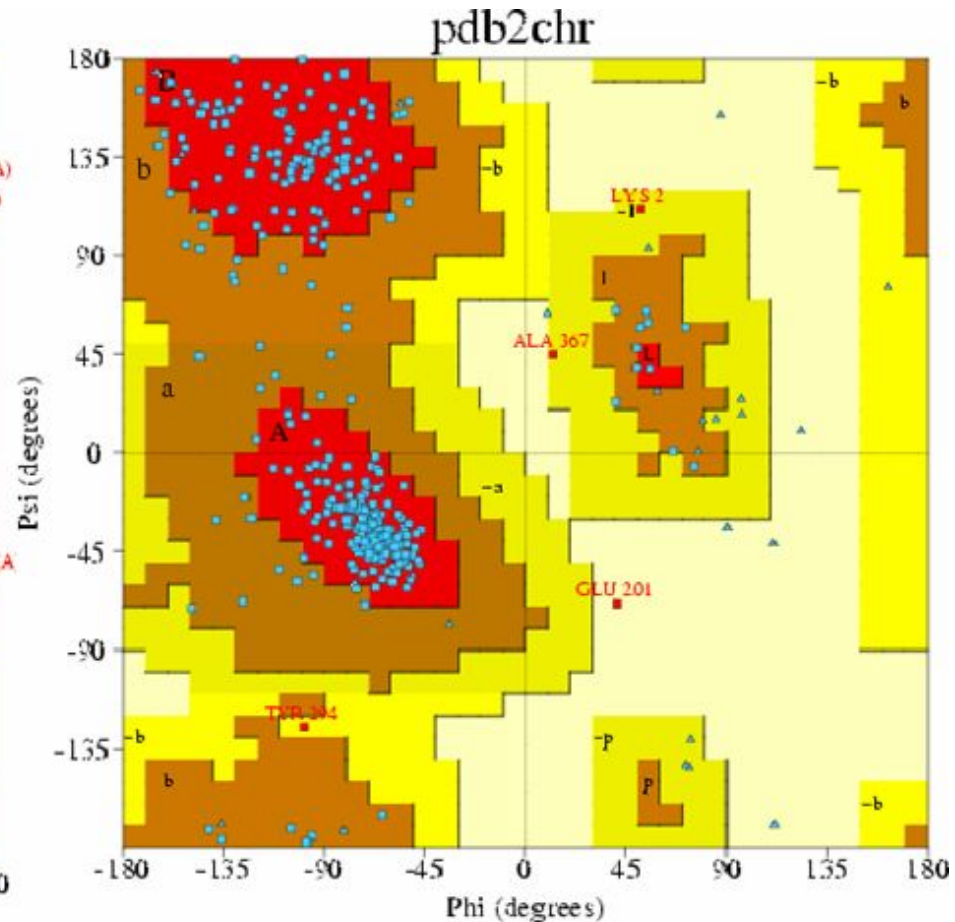
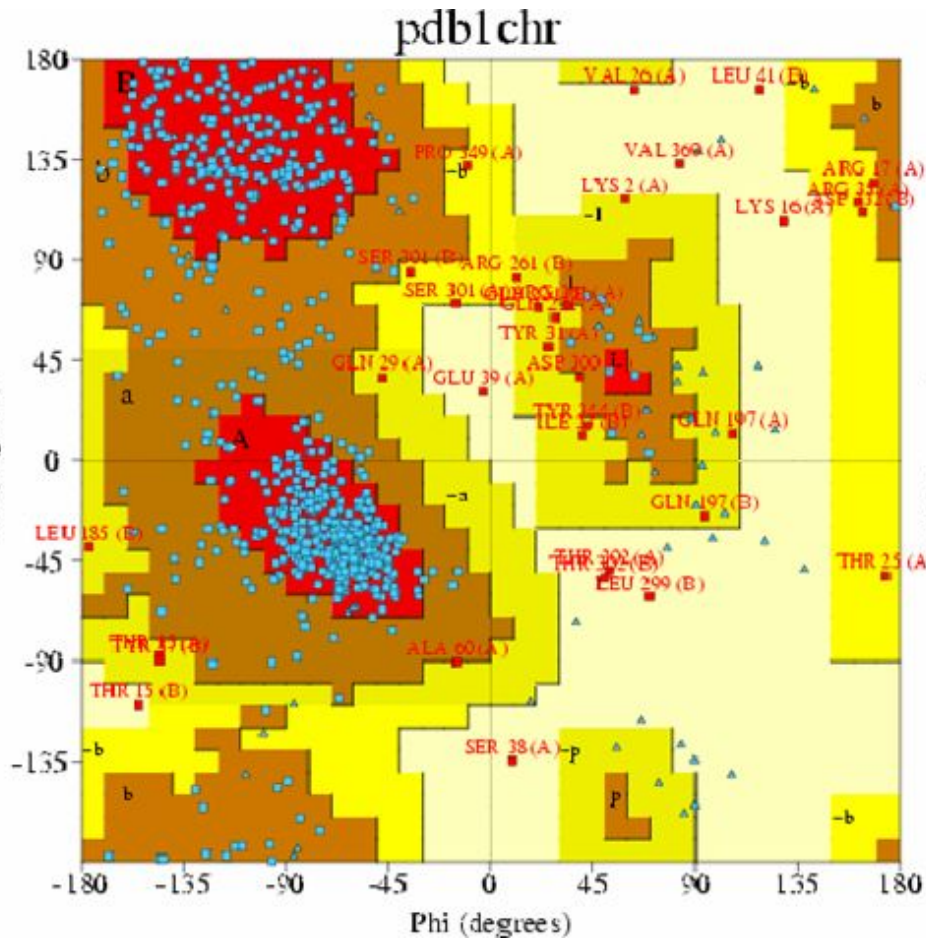
Выделяют предпочитаемую, разрешенную, допустимую, запрещённую области на карте.

Попадание остатка вне предпочитаемой области - маргинал.

В хорошей модели >90% остатков, не считая Gly, Pro, находятся в предпочитаемой области



# Показатели качества отдельных остатков: карты Рамачандрана для $\phi$ и $\psi$

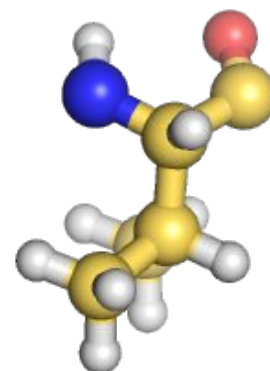
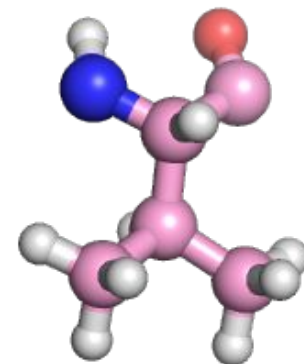
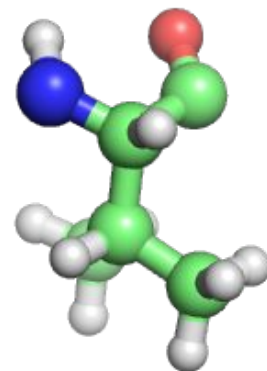


# Показатели качества отдельных остатков: торсионы боковых цепей

Боковые цепи имеют от 0 (Gly, Ala) до 4х (Lys, Arg) степеней свободы - вращений вокруг ковалентных связей боковой цепи

Соответствующие торсионные углы обозначаются  $\chi_1, \dots, \chi_4$ , отсчёт идет от связи  $C\alpha - C\beta$

Ротамеры - типичные конформации боковых цепей остатков. Задаются значениями  $\chi_1, \dots, \chi_4$  и доверительными интервалами к ним. Вычисляются на основе статистики базы PDB.



Ротамеры Val

# Показатели качества отдельных остатков: RSR

Пространственный R-фактор сравнивает соответствие электронной плотности модели и “экспериментальной” электронной плотности. Заменяет визуальный анализ того, как остаток вписан в электронную плотность.

Проблема: “экспериментальная” электронная плотность строится до оптимизации модели и больше никак не используется. Откуда её взять?

$$RSR = \frac{\sum_{A \in L} |\rho_{\text{эксп}} - \rho_{\text{модель}}|}{\sum_{A \in L} \rho_{\text{эксп}}} [\cdot 100\%]$$

(Суммирование по гриду в пространстве вокруг группы атомов)

**Хорошие значения:** RSR < 10%  
**Плохие:** RSR > 20%

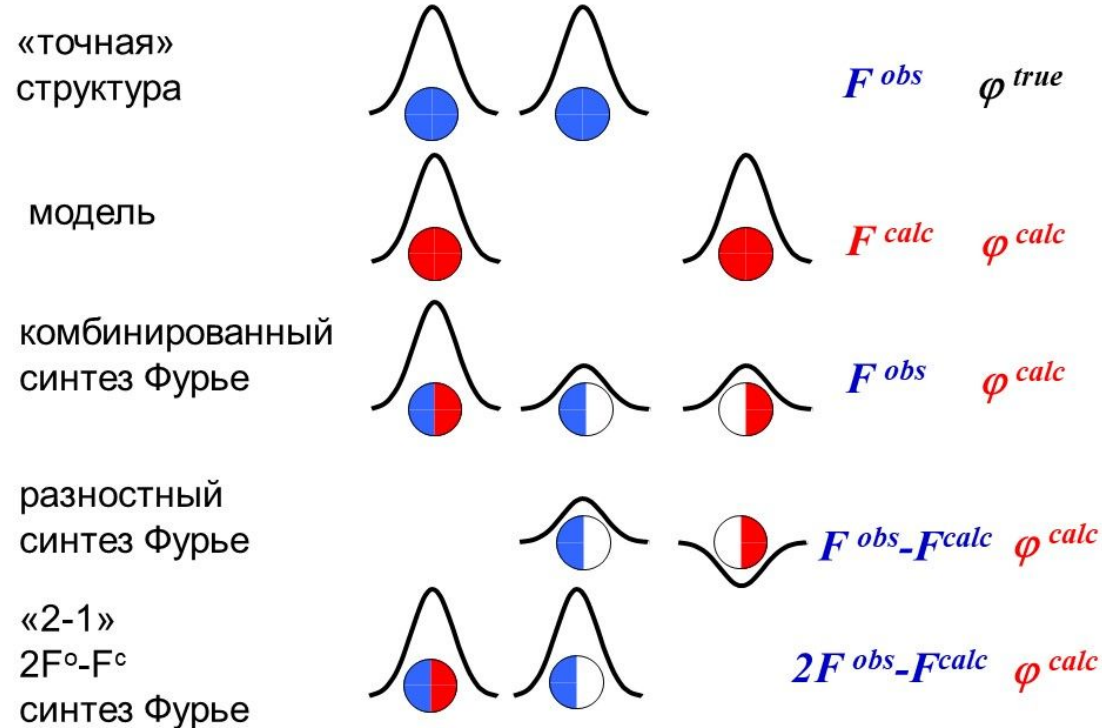
# Показатели качества отдельных остатков: RSR

“Экспериментальную” ЭП  
считают с помощью  
комбинированного синтеза  
Фурье:  $2F_{\text{эксп}} - F_{\text{модель}}$ ,  $\varphi_{\text{модель}}$

Откуда взять  $F_{\text{эксп}}$  для  
структуры в PDB?

С 2008 года при  
депонировании структуры  
в PDB авторы обязаны  
предоставлять файл  
структурных факторов

«Комбинированные» синтезы Фурье.



# Показатели качества отдельных остатков: RSR Z-score

Относительная оценка RSR для конкретного остатка.

RSR сравнивается со средним RSR для того же типа остатков (Ala) по выборке из PDB с примерно таким же разрешением

Если RSR плохой, а RSR-Z – хороший, то значит координаты атомов расшифрованы плохо, но не хуже, чем в других подобных структурах.

$$Z = \frac{(RSR - \langle RSR_{\text{resolution}} \rangle)}{\sigma(RSR_{\text{resolution}})}$$

**Хорошие значения:** RSR-Z < 2  
**Плохие:** RSR-Z > 2

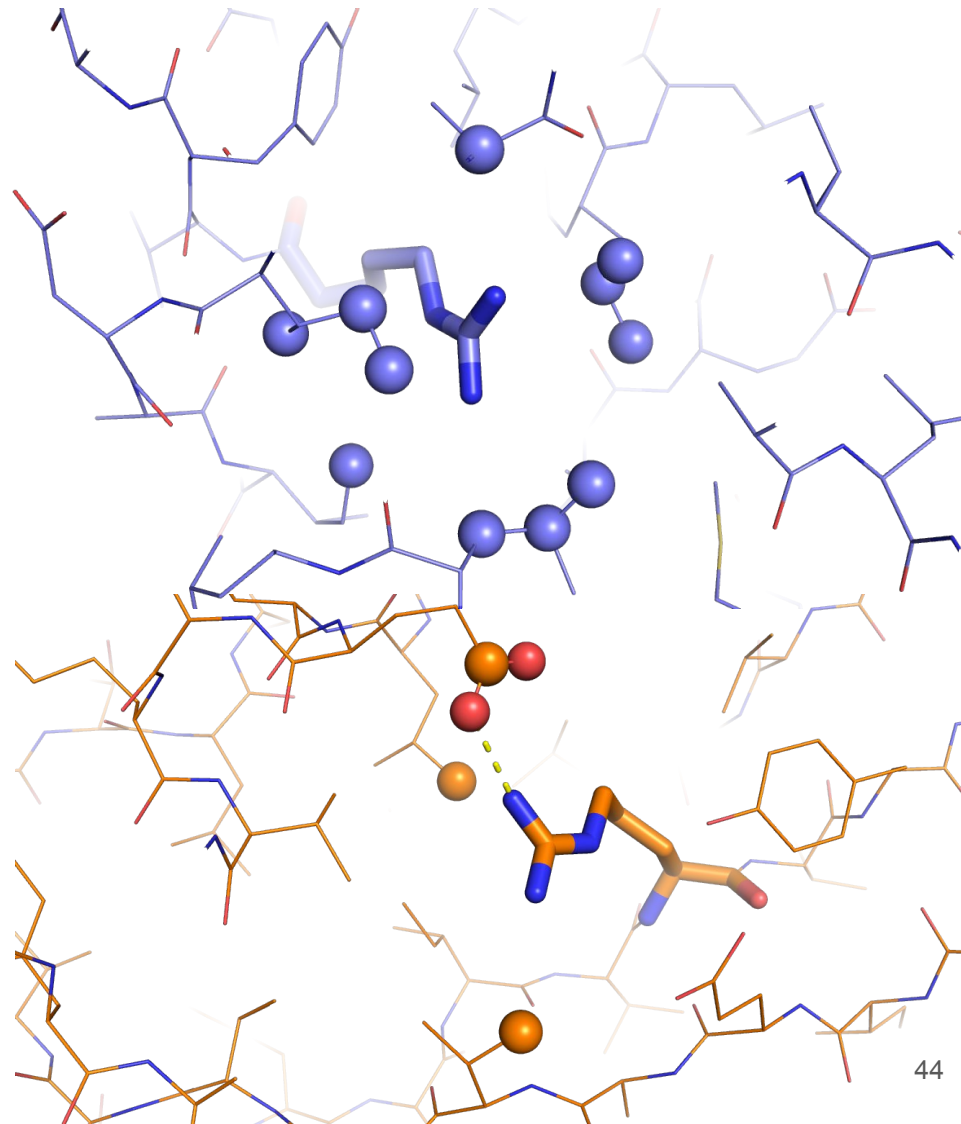


# Показатели качества отдельных остатков: “Комфортность” окружения атома

Основывается на физико-химических ограничениях для взаимодействующих групп атомов.

Полностью независим от процедуры оптимизации модели.

Зависим от экспертной оценки.



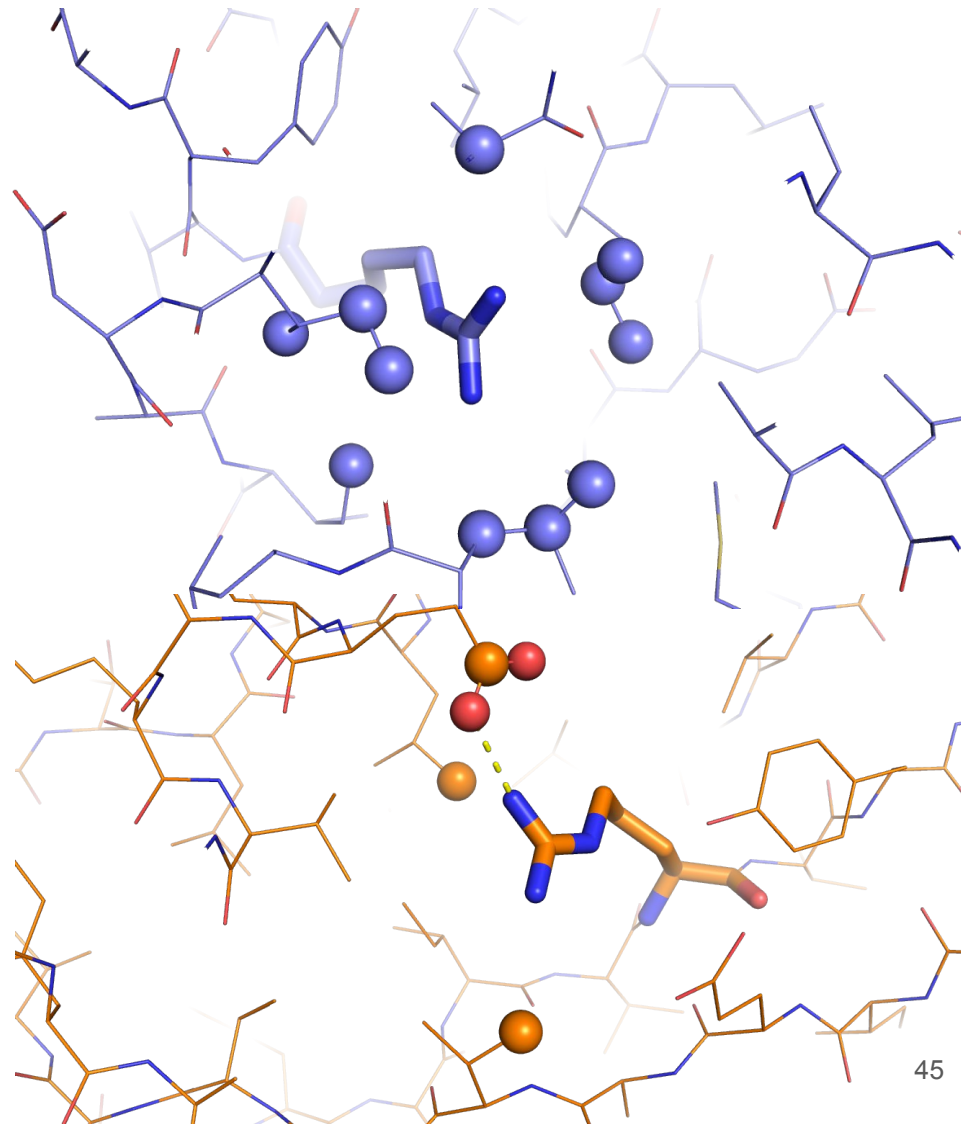
Arg35 в 1chr и 2chr



# Показатели качества отдельных остатков: “Комфортность” окружения атома

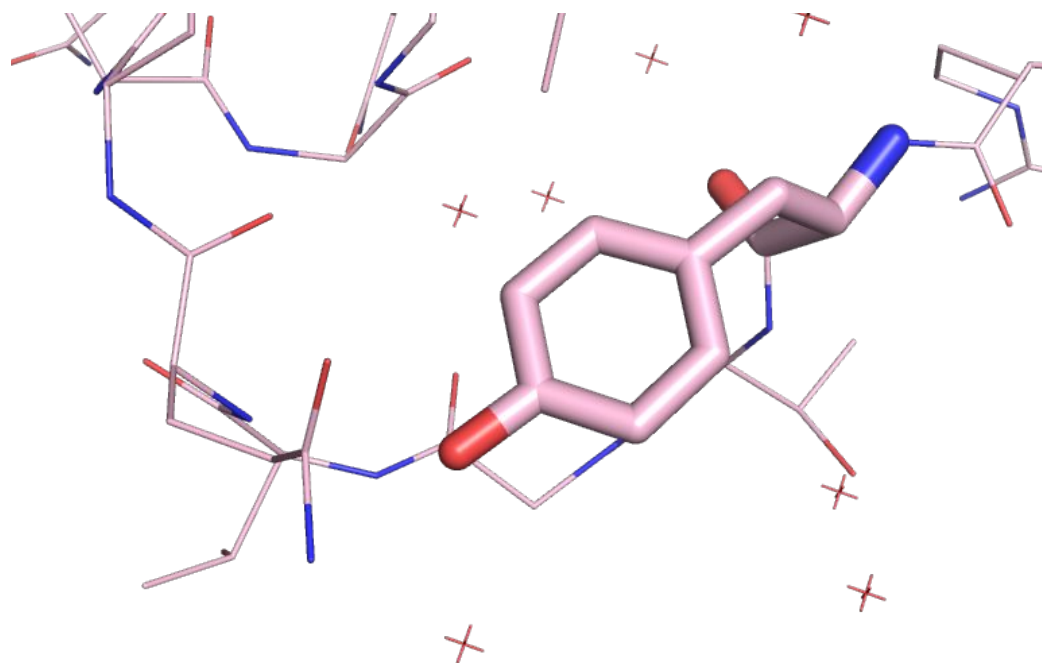
- Clash (перекрывание Ван-дер-Ваальсовых радиусов несвязанных атомов)
- Экспонированные гидрофобные остатки
- Доноры/акцепторы водородной связи без водородной связи
- Заряженные группы не скомпенсированы противоположными ионами

Arg35 в 1chr и 2chr



# Показатели качества отдельных остатков: Z-score комфортности окружения остатка

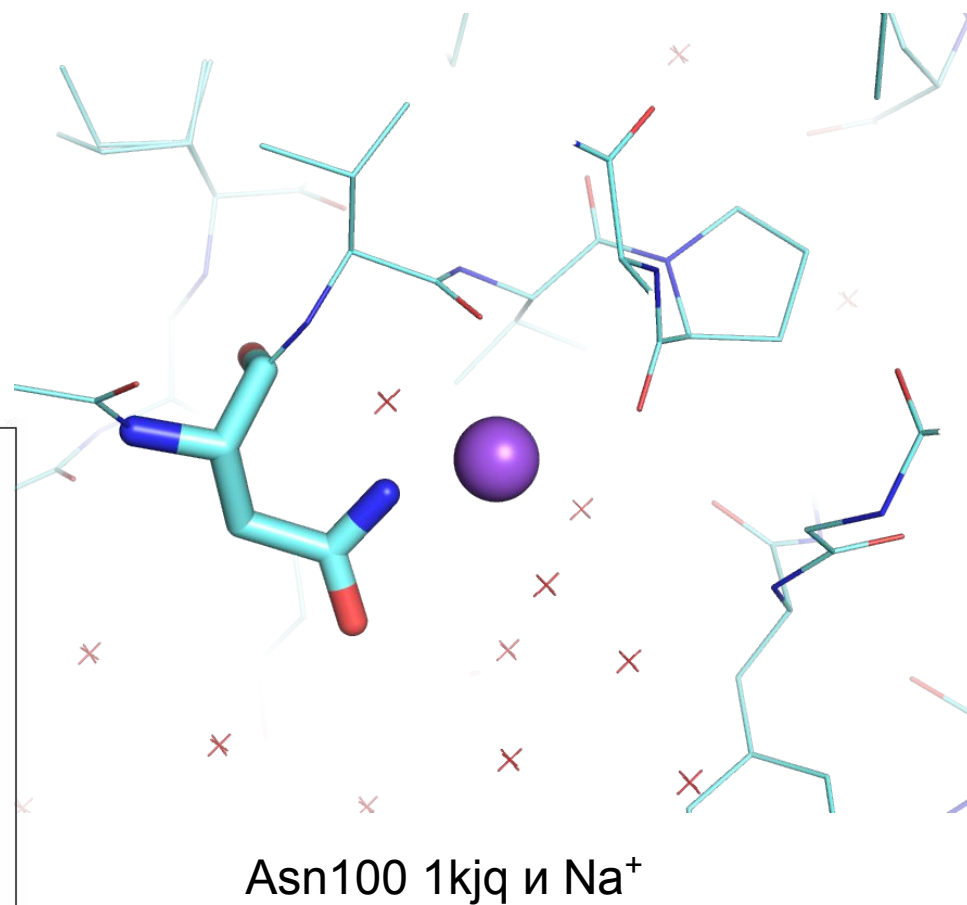
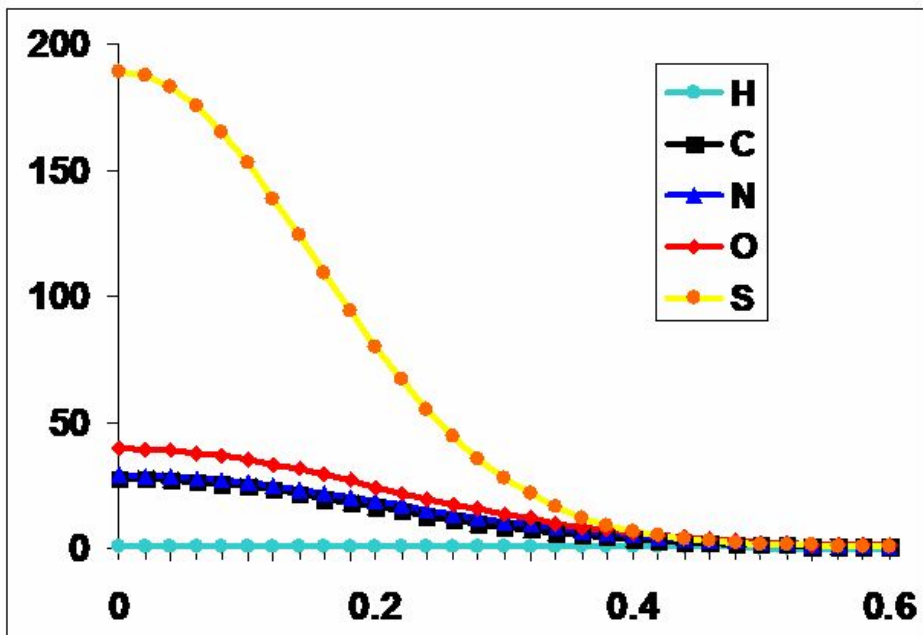
- В программе WhatCheck, например, рассчитывается Z-score для комфортности окружения каждой боковой цепи
- Маргиналы – Z-score < -5
- Маргиналов по окружению стоит проверять визуально: часто маргинальность объясняется выходом на поверхность глобулы, контактом с белком из соседней ячейки и др.



Tyr221 5TLN,  
Z-score=5.07

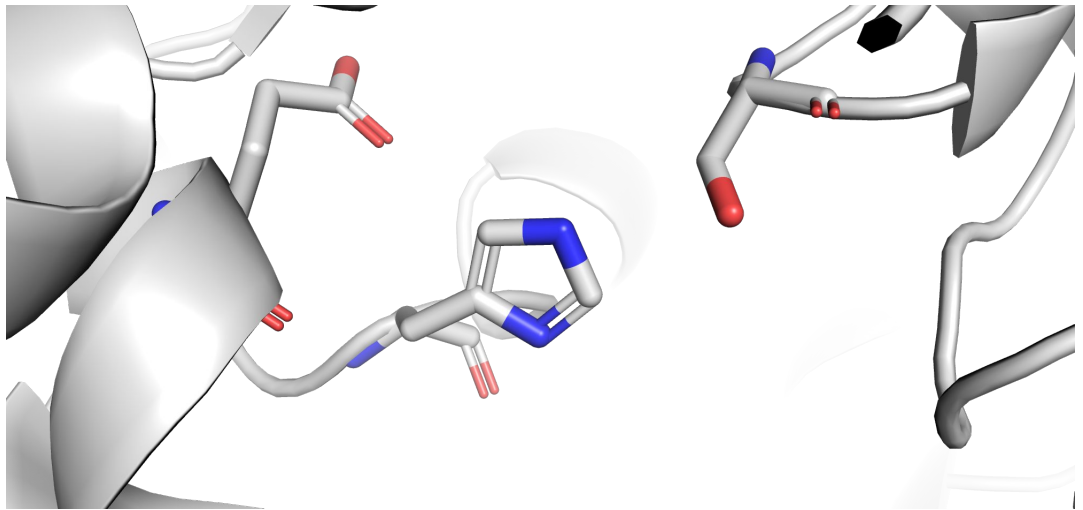
# Показатели качества отдельных остатков: Инверсии Asn/Gln/His

Электронная плотность O и N трудноразличима. Ориентацию Gln/Asn надо проверять.

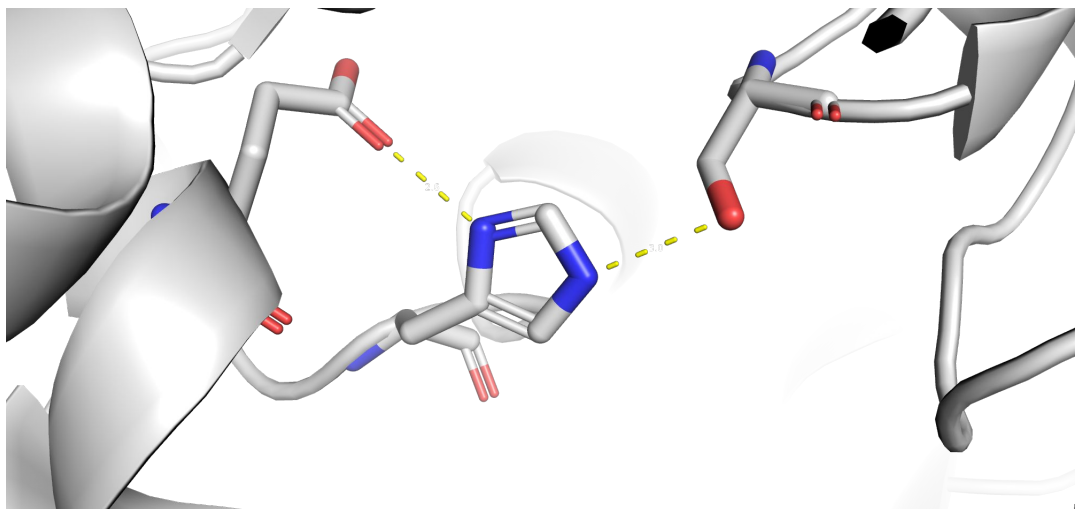


# Показатели качества отдельных остатков: Инверсии Asn/Gln/His

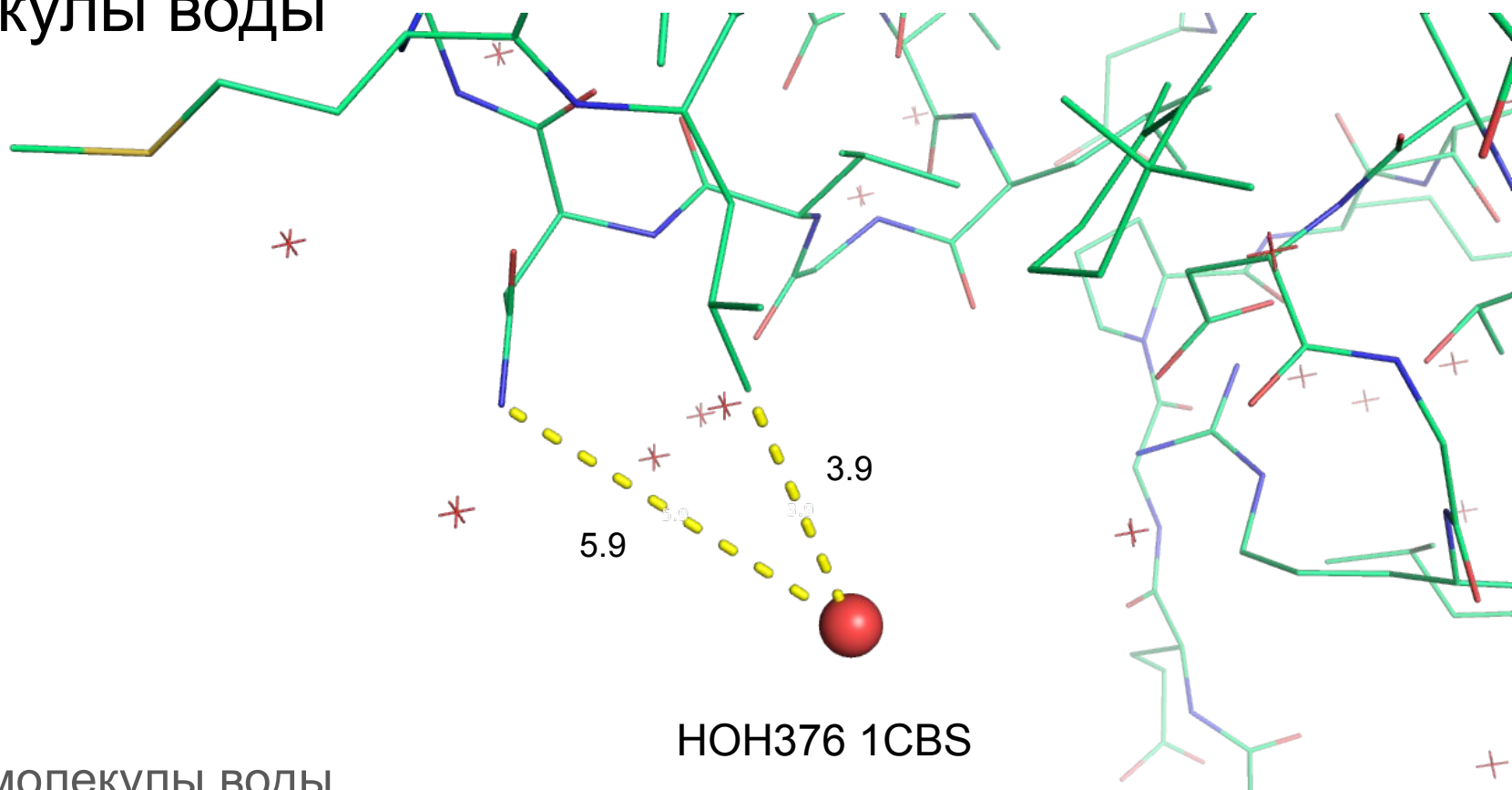
Каталитическая  
триада в 6ESJ



Функциональная  
каталитическая  
триада



# Показатели качества отдельных остатков: Молекулы воды

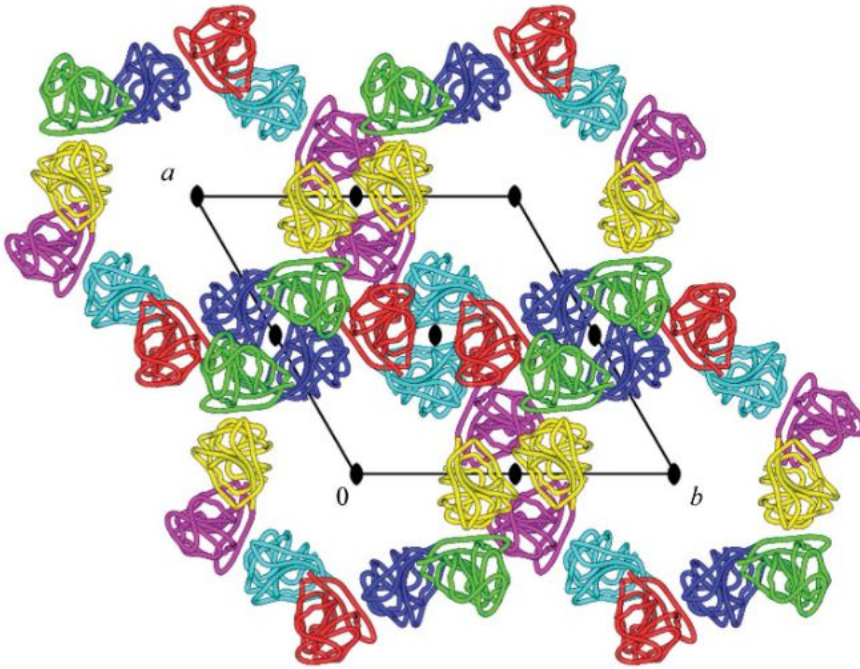


HOH376 1CBS

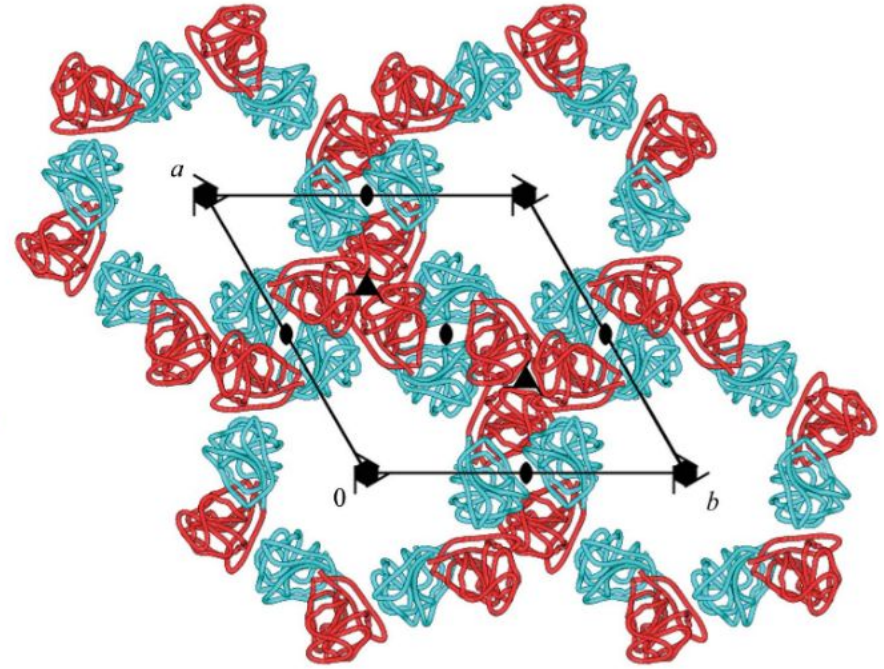
Не все молекулы воды кристалла не являются ошибочными. Конкретно эта контактирует с соседом по кристаллу.

```
REMARK 525 SOLVENT
REMARK 525
REMARK 525 THE SOLVENT MOLECULES HAVE CHAIN IDENTIFIERS THAT
REMARK 525 INDICATE THE POLYMER CHAIN WITH WHICH THEY ARE MOST
REMARK 525 CLOSELY ASSOCIATED. THE REMARK LISTS ALL THE SOLVENT
REMARK 525 MOLECULES WHICH ARE MORE THAN 5A AWAY FROM THE
REMARK 525 NEAREST POLYMER CHAIN (M = MODEL NUMBER;
REMARK 525 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE
REMARK 525 NUMBER; I=INSERTION CODE):
REMARK 525
REMARK 525 M RES CSSEQI
REMARK 525 HOH A 376          DISTANCE = 5.94 ANGSTROMS
```

# Тонкости: несколько биологических единиц в асимметричной ячейке



Если мономеры структурно отличаются - адекватно

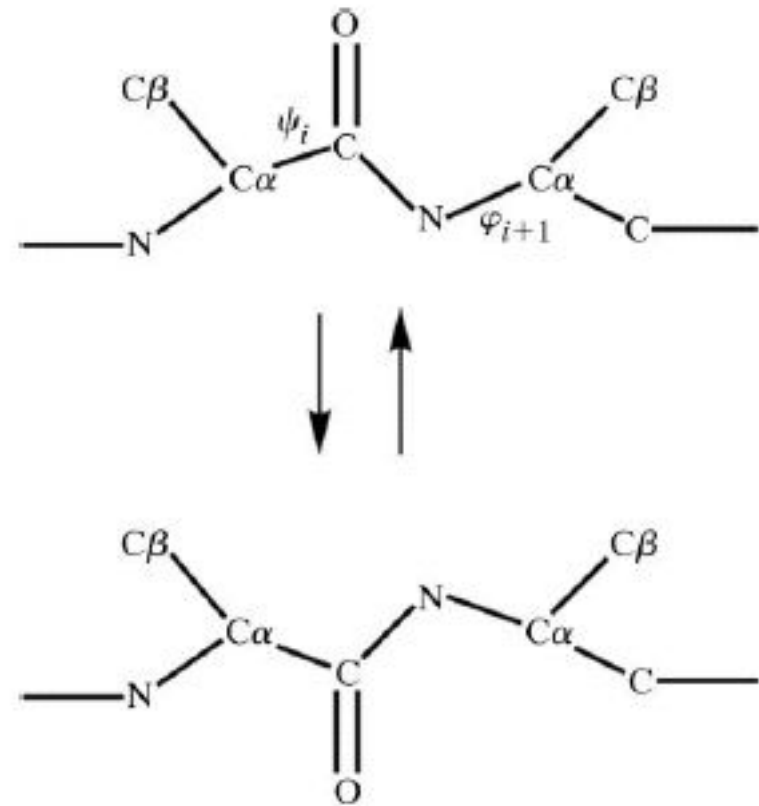


Если мономеры идентичны - ведёт к падению качества полученной ЭП



# Тонкости: инверсия пептидной цепи

Индикатором возможности инверсии служат два идущих подряд остатка на карте Рамачандрана, попавшие в неблагоприятные области



# Как валидировать структуру из PDB?

- Автоматический отчёт PDB
- Сторонние программы/сервисы:

MolProbity, WHAT\_CHECK,  
CheckMyMetal ...

- Метод пристального взгляда  
(экспертная оценка)

# Remediation

повторное построение моделей по исходным экспериментальным данным

