

# Validace strukturních dat

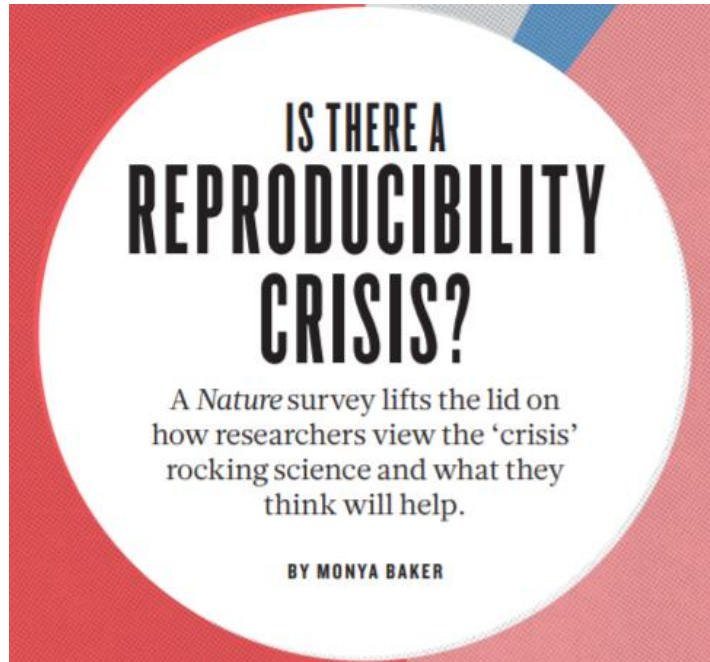
C3210 Strukturní Bioinformatika

# Proč validovat?

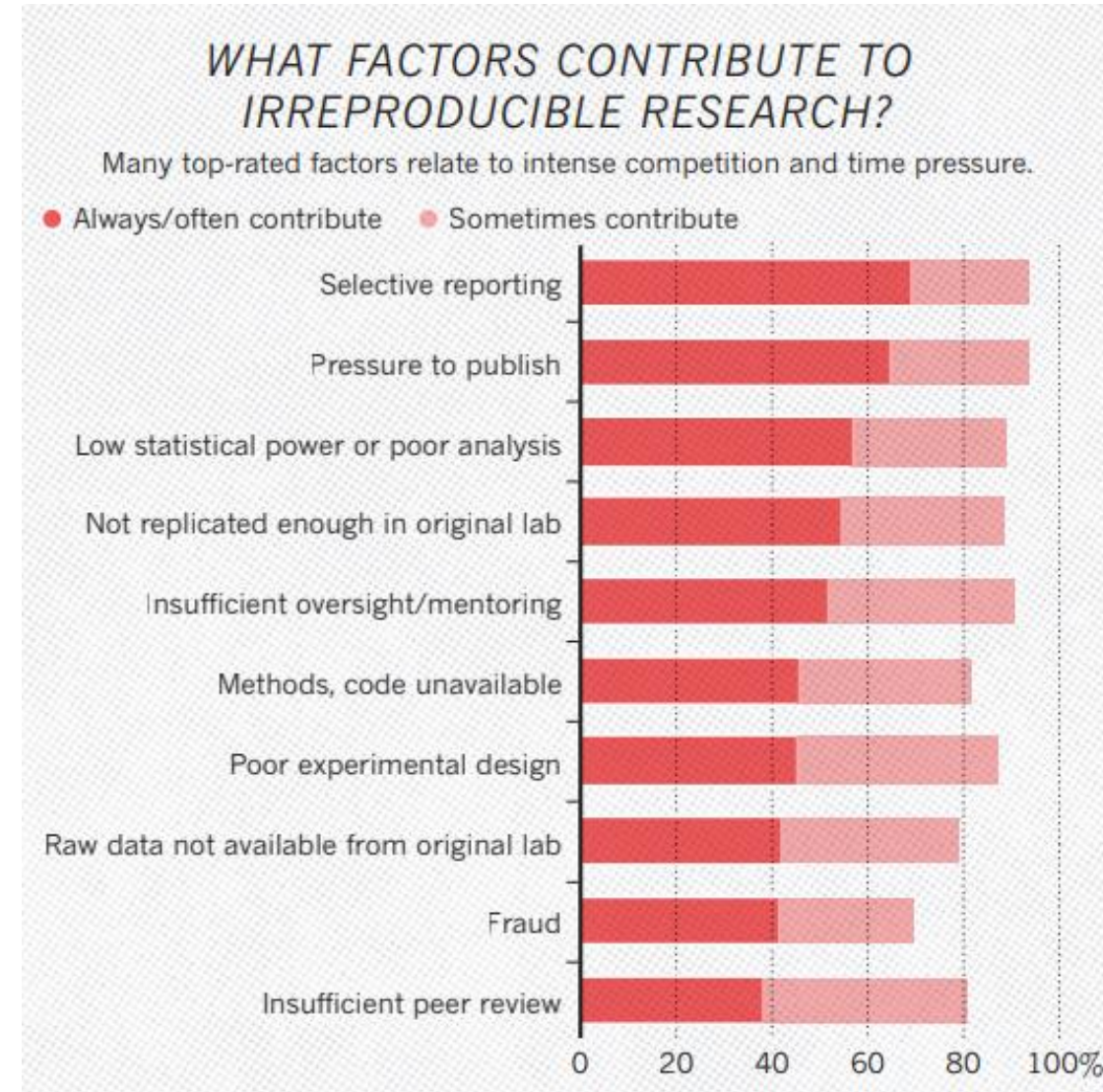
- Nespolehlivost
- „Garbage in garbage out“
- Selektivita dat



# Krise reprodukce výsledků

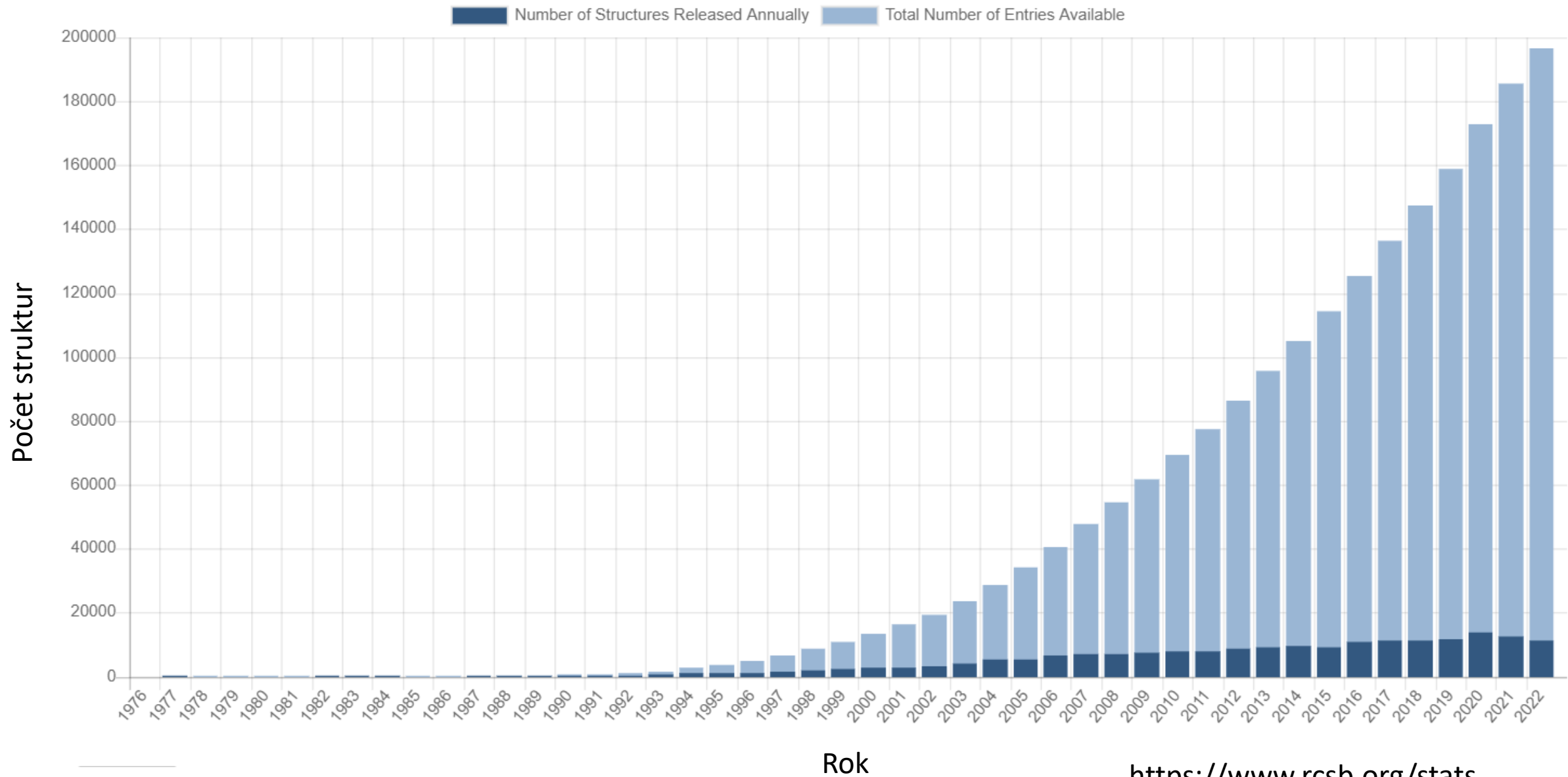


Více než 70 % respondentů odpovědělo, že mají zkušenost s neúspěšným pokusem o zopakování experimentů z jiných laboratoří.

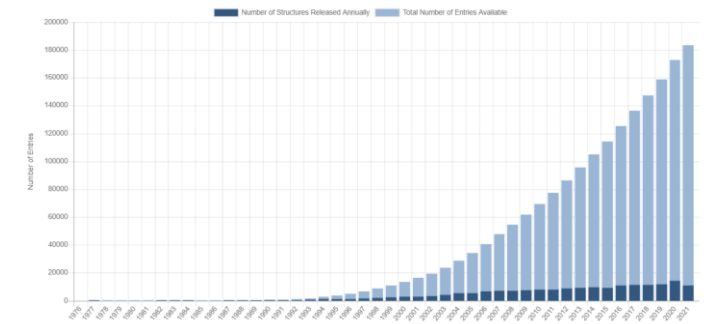
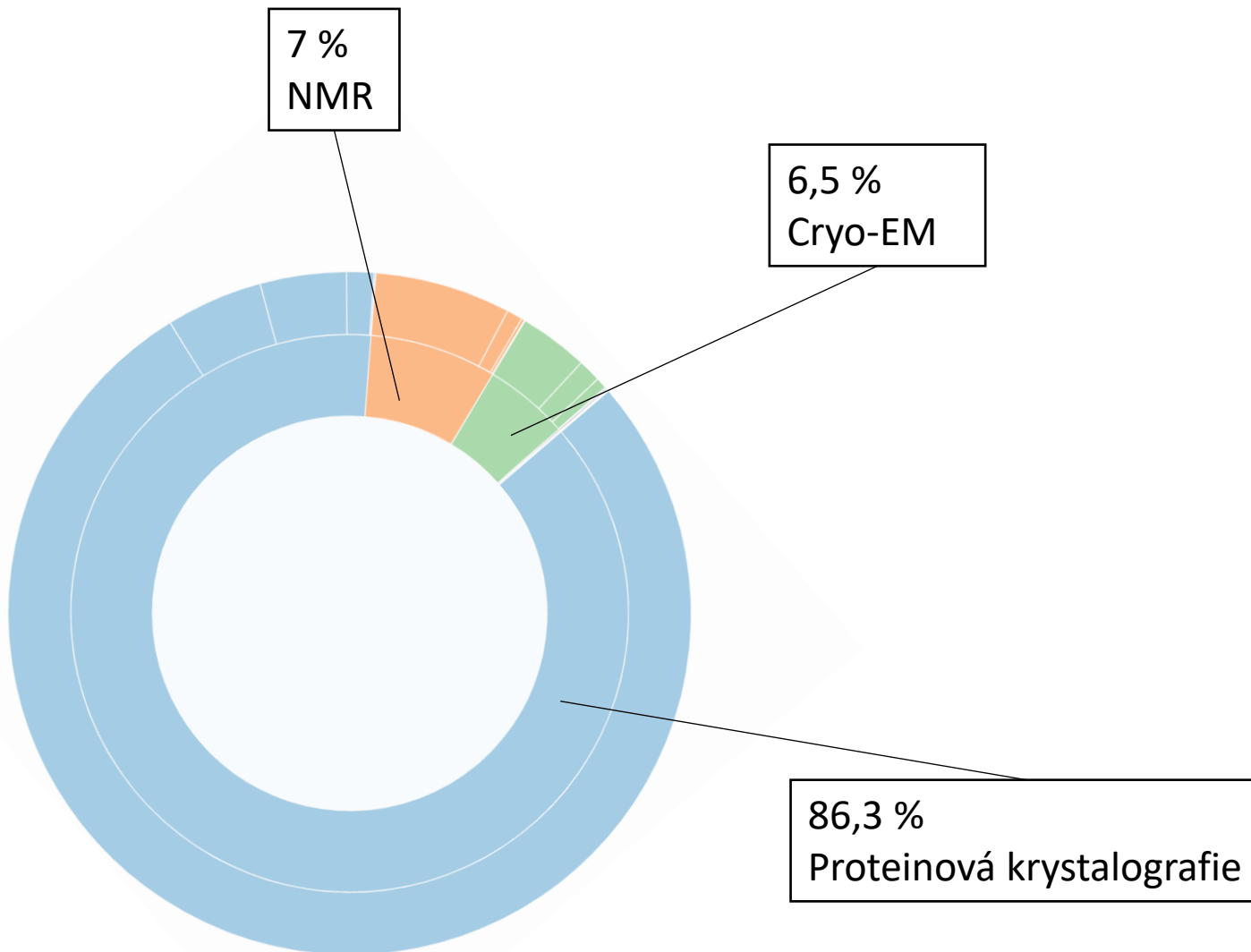


<https://www.nature.com/articles/533452a>

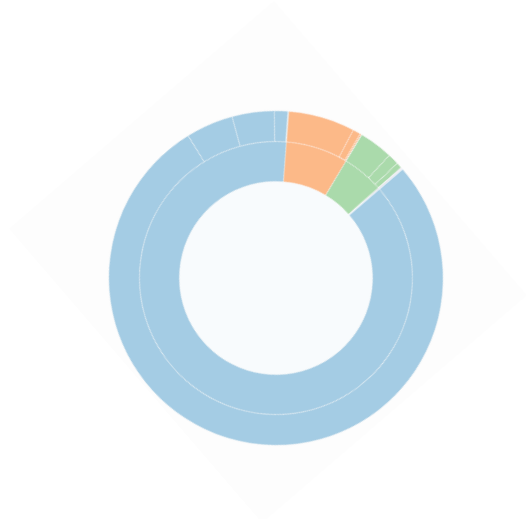
# Nárůst proteinových struktur v PDB databázi



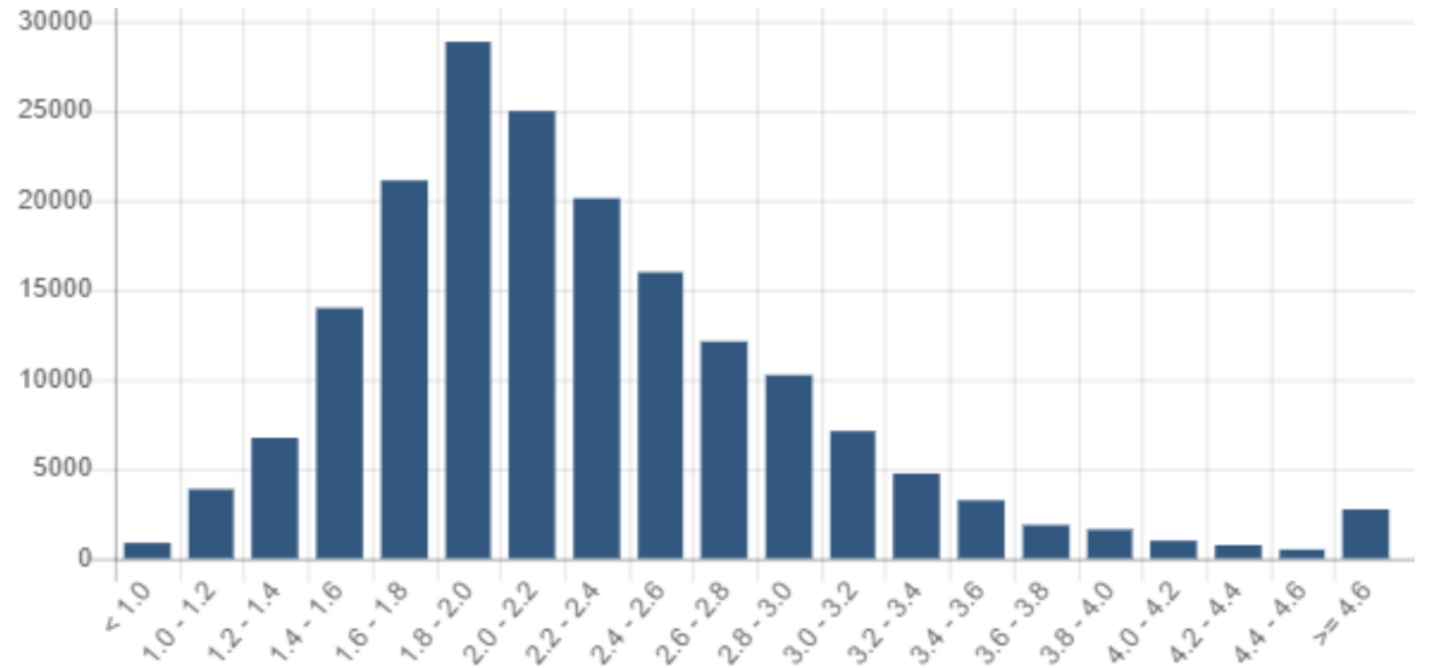
# Zastoupení vložených struktur podle metody



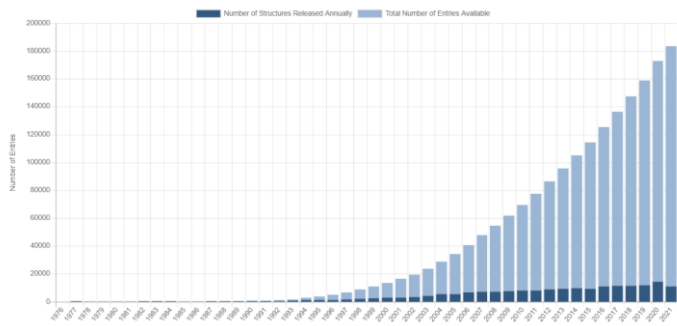
# Selekce struktur podle rozlišení



Počet struktur

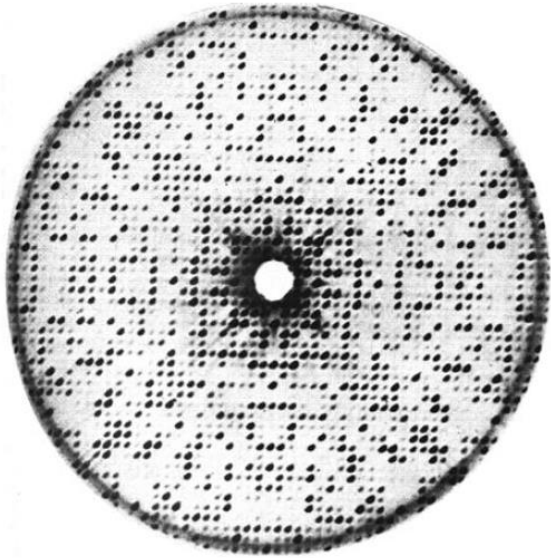


Rozlišení (Ångström [Å])



# Opakování – proteinová krystalografie

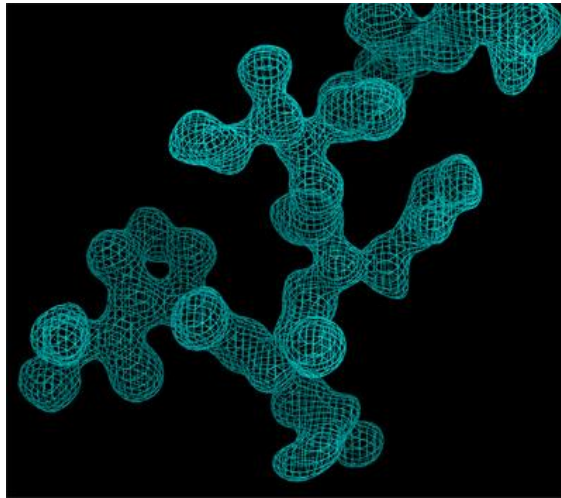
Krok 1



Získ difrakčních dat



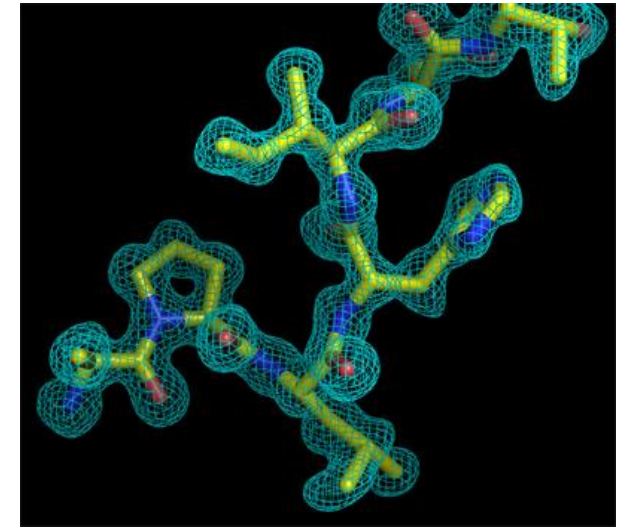
Krok 2



Výpočet elektronové hustoty



Krok 3



Interpretace elektronové hustoty pomocí modelu

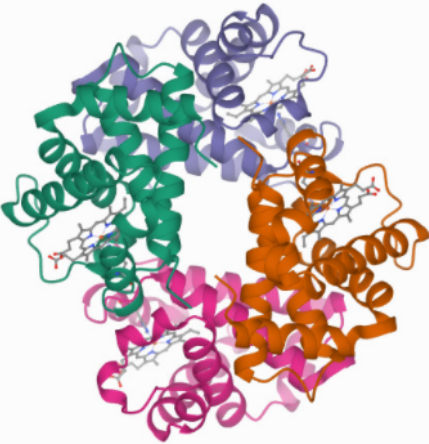
# Proteinová krystalografie

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers

MyPDB

Structure Summary 3D view Annotations Experiment Sequence Genome Ligands versions

Biological Assembly 1 ?



3D View: Structure | Electron Density | Ligand Interaction

Global Symmetry: Cyclic - C2 (3D View)  
Global Stoichiometry: Hetero 4-mer - A2B2

Pseudo Symmetry: Dihedral - D2 (3D View)  
Pseudo Stoichiometry: Homo 4-mer - A4

Find Similar Assemblies

Biological assembly 1 assigned by authors and generated by PISA (software)

## 1A3N

DEOXY HUMAN HEMOGLOBIN

DOI: [10.2210/pdb1A3N/pdb](https://doi.org/10.2210/pdb1A3N/pdb)

Classification: **OXYGEN TRANSPORT**

Organism(s): *Homo sapiens*

Mutation(s): No

Deposited: 1998-01-22 Released: 1998-04-29

Deposition Author(s): Tame, J., Vallone, B.

### Experimental Data Snapshot

Method: X-RAY DIFFRACTION

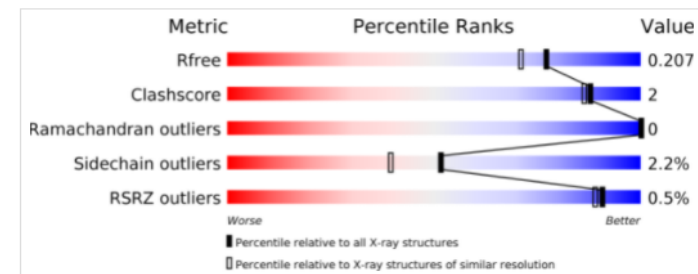
Resolution: 1.80 Å

R-Value Free: 0.220

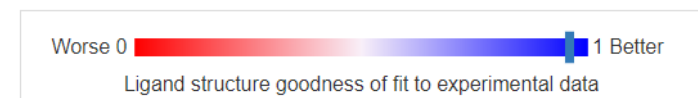
R-Value Work: 0.171

### wwPDB Validation

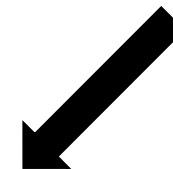
3D Report Full Report



### Ligand Structure Quality Assessment



This is version 1.2 of the entry. See complete [history](#).

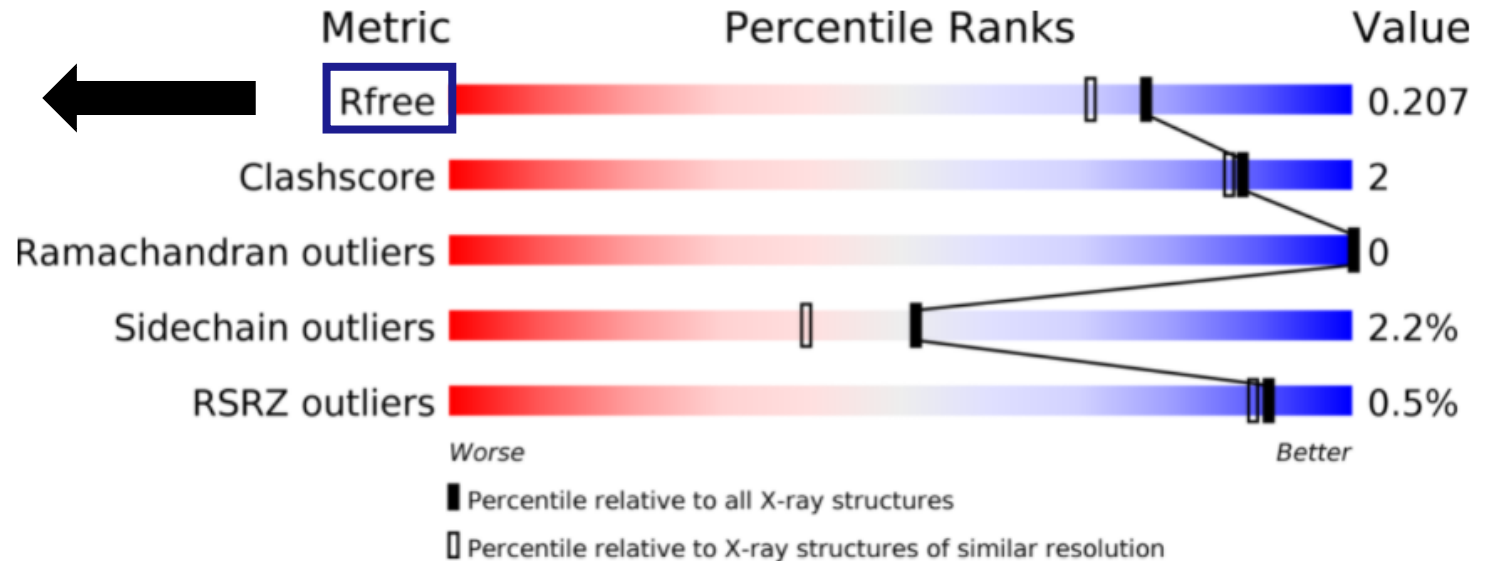




# Validace struktur v PDB

$$R_{free} = \frac{\sum_{hkl \in T} ||F_O| - k|F_C||}{\sum_{hkl \in T} |F_O|}$$

## Validační statistika wwPDB

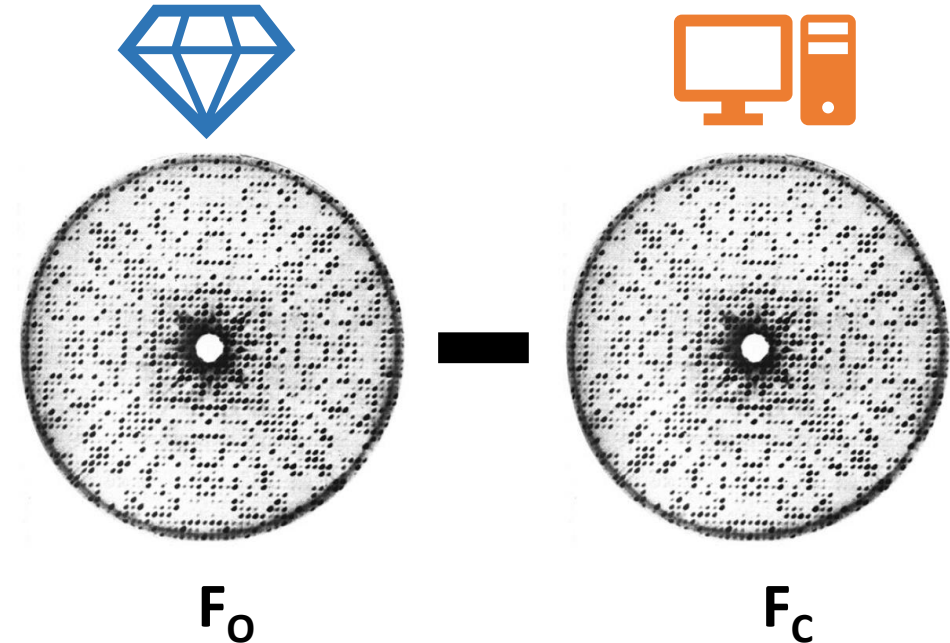


# Kvalita modelu (R-faktor)

Experimentálně  
pozorované  
difrakce

Vypočtené difrakce  
z našeho modelu

$$R = \frac{\sum_{hkl} ||F_O| - k|F_C||}{\sum_{hkl} |F_O|}$$



## Pro R-faktor platí:

- Dva možné zápisy:  $R = 0,2$  nebo  $R = 20 \%$
- R-faktor může mít teoreticky hodnoty od 0,0 po 1,0
- V praxi se budete střídat s hodnotami mezi 0,1 a 0,4
- Čím nižší R-faktor, tím lepší struktura



Dobře vyřešená struktura nabývá hodnoty  $R = \text{rozlišení}/10$   
tj. pro dobře vyřešenou strukturu s rozlišením 2 Å je optimální  
 $R = 0,2$



# Kvalita modelu (R-free)

Experimentálně  
pozorované  
difrakce

Vypočtené difrakce  
z našeho modelu

Množina **T** jsou náhodně vybraná data,  
která nejsou použita  
pro výpočet elektronové hustoty.

$$R = \frac{\sum_{hkl} ||F_O| - k|F_C||}{\sum_{hkl} |F_O|}$$

$$R_{free} = \frac{\sum_{hkl \in T} ||F_O| - k|F_C||}{\sum_{hkl \in T} |F_O|}$$

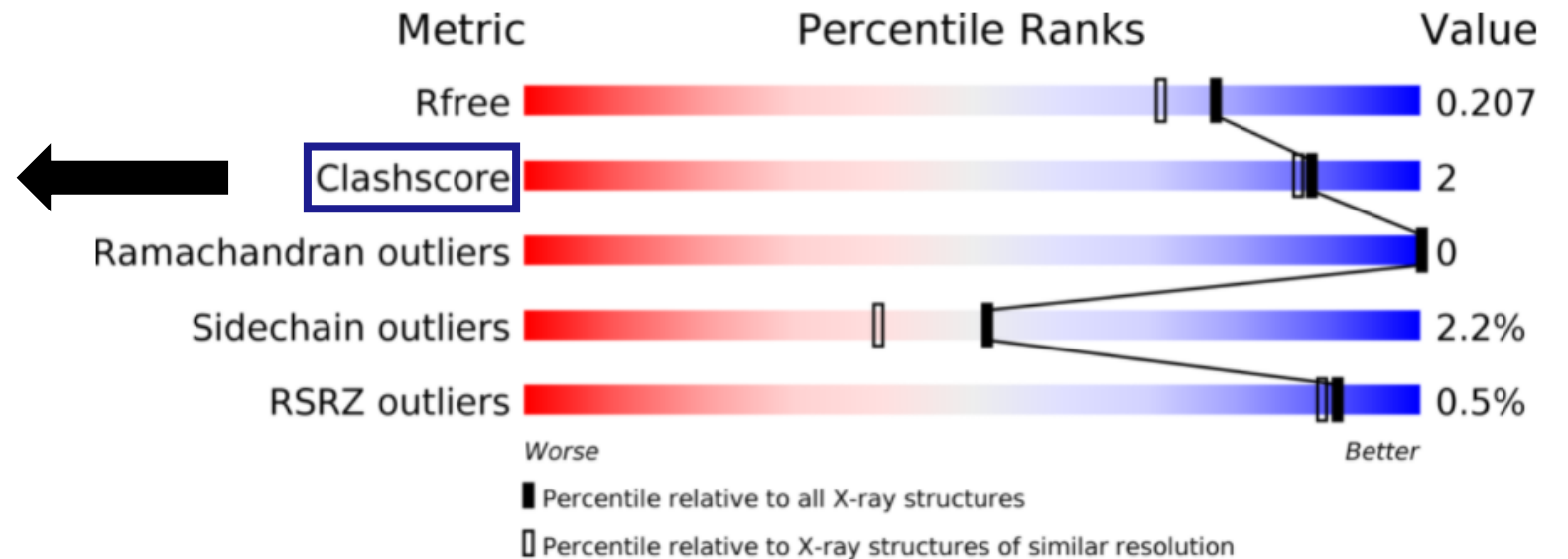
**Pro  $R_{free}$  platí:**

- Zpravidla je o něco větší než R-faktor, tj. pro strukturu s  $R = 0,2$  je rozumné mít  $R_{free} = 0,23$
- Pokud je velký rozdíl mezi  $R$  a  $R_{free}$  struktury, tato struktura je podezřelá a je nutná podrobnější inspekce

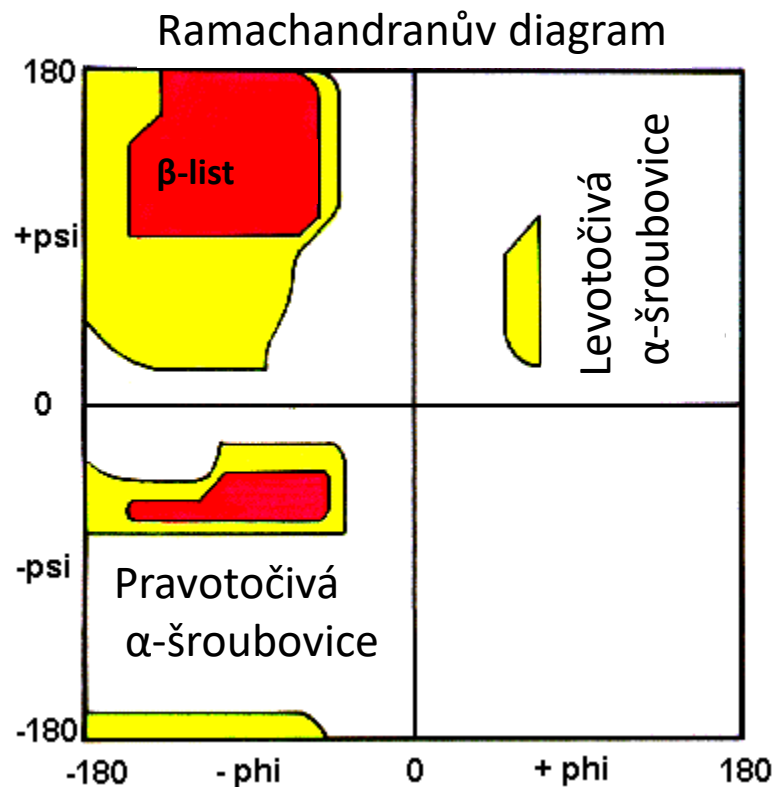
# Validace struktur v PDB

## Validační statistika wwPDB

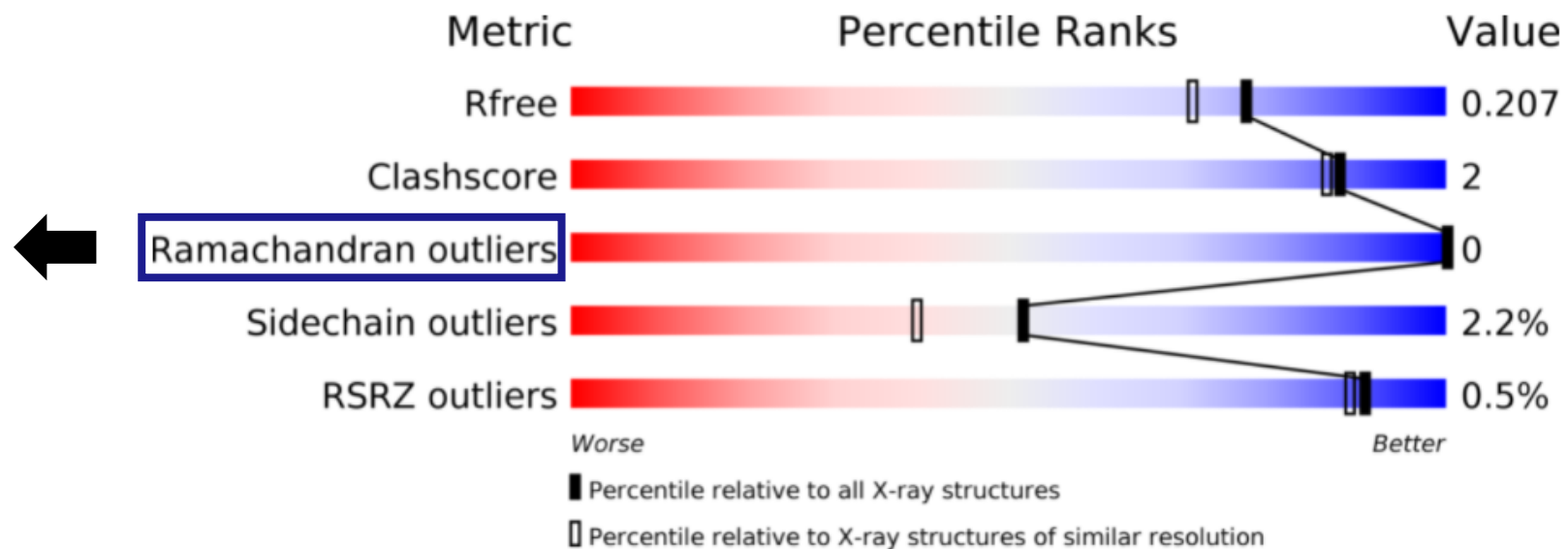
Počet párů atomů, které jsou neobvykle blízko sebe.



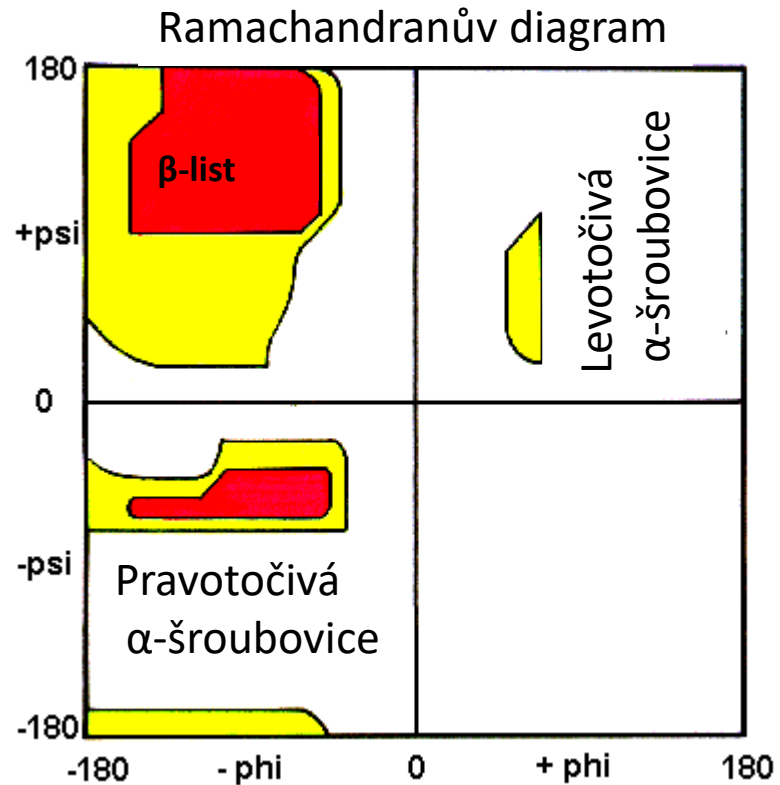
# Validace struktur v PDB



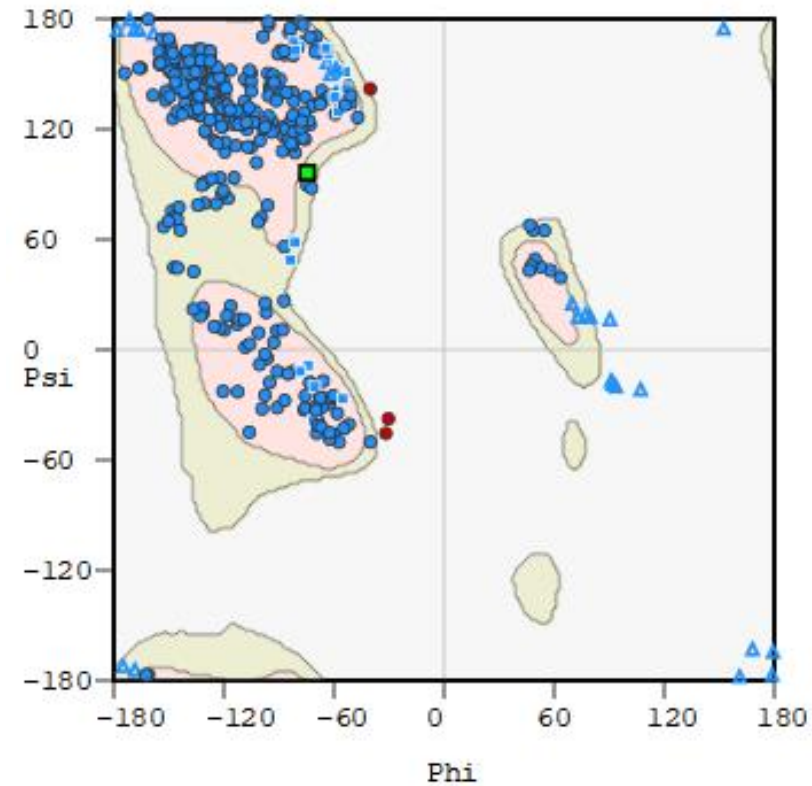
## Validační statistika wwPDB



# Kvalita modelu (Ramachandranův diagram)



Ramachandranův diagram  
modelu proteinu

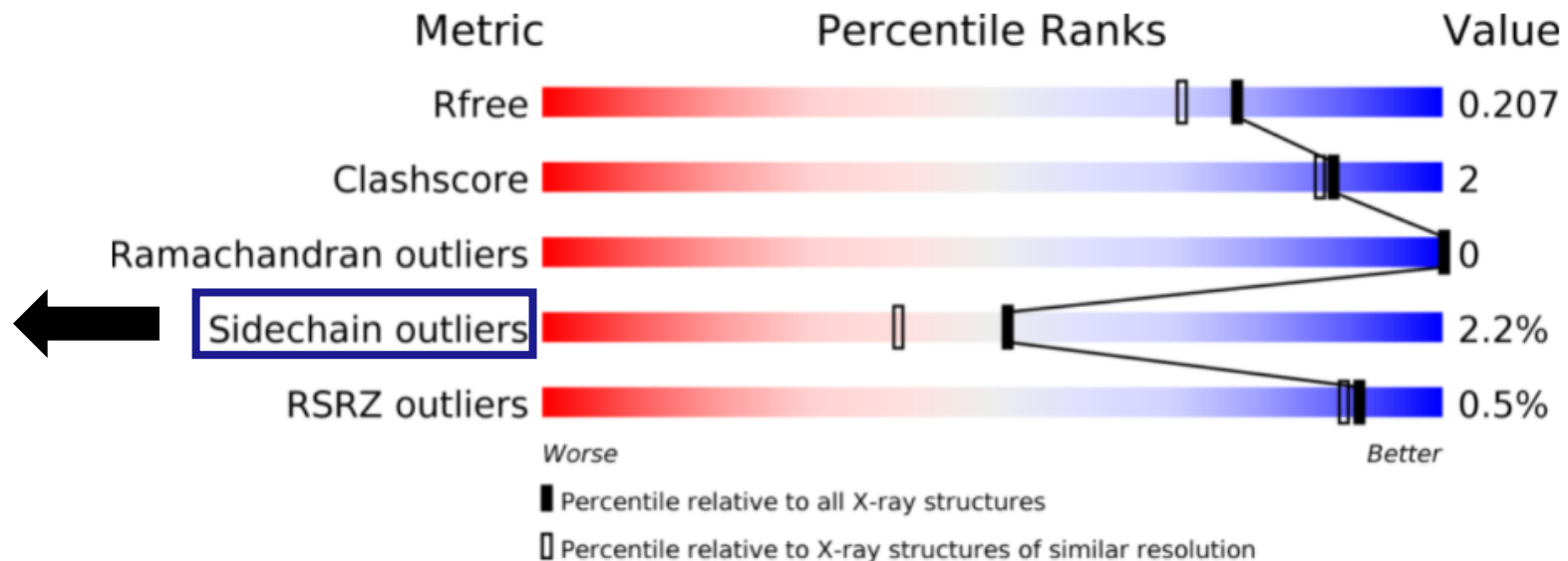


**Modře**- residua se správnou konformací torzních úhlů  
**Červeně**- residua s nesprávnou konformací torzních úhlů

# Validace struktur v PDB

## Validační statistika wwPDB

Procento aminokyselin,  
které mají neobvyklou konformaci  
bočního řetězce.



# Validace struktur v PDB

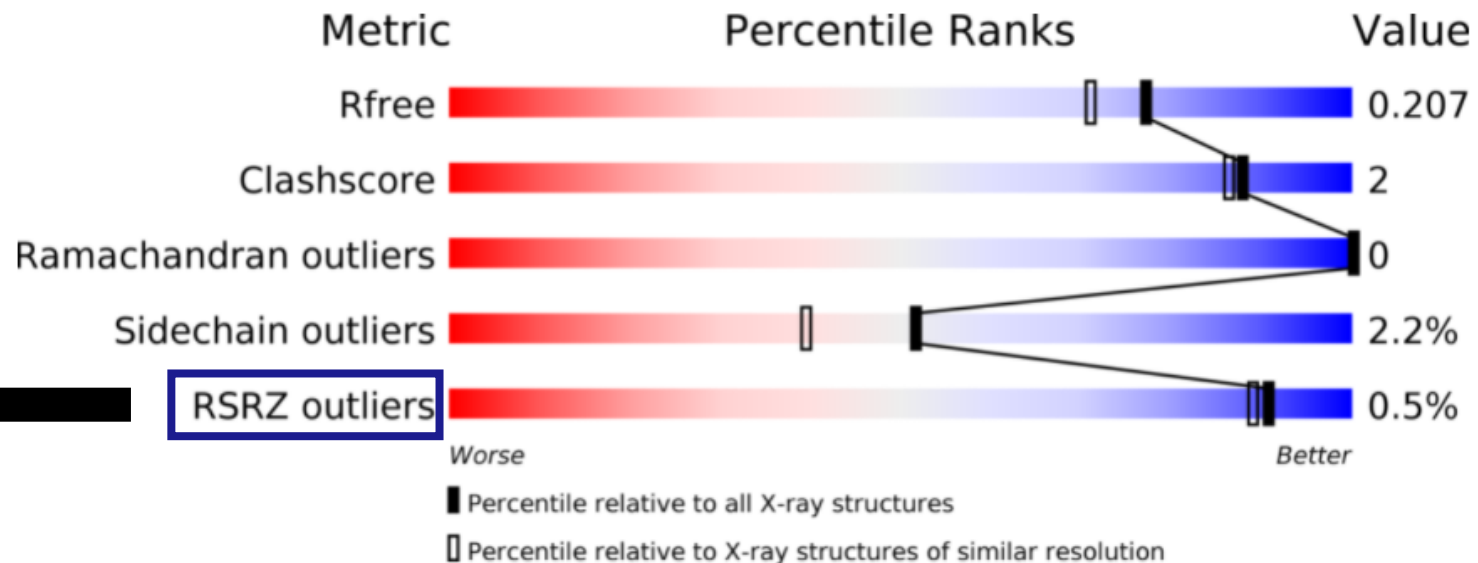
## Validační statistika wwPDB

„Real-space R-value“ je ukazatel,  
jak dobře odpovídá model  
elektronové hustotě.

Daný ukazatel reprezentuje procento  
reziduí, která nesedí dobře v el. hustotě.



RSRZ outliers



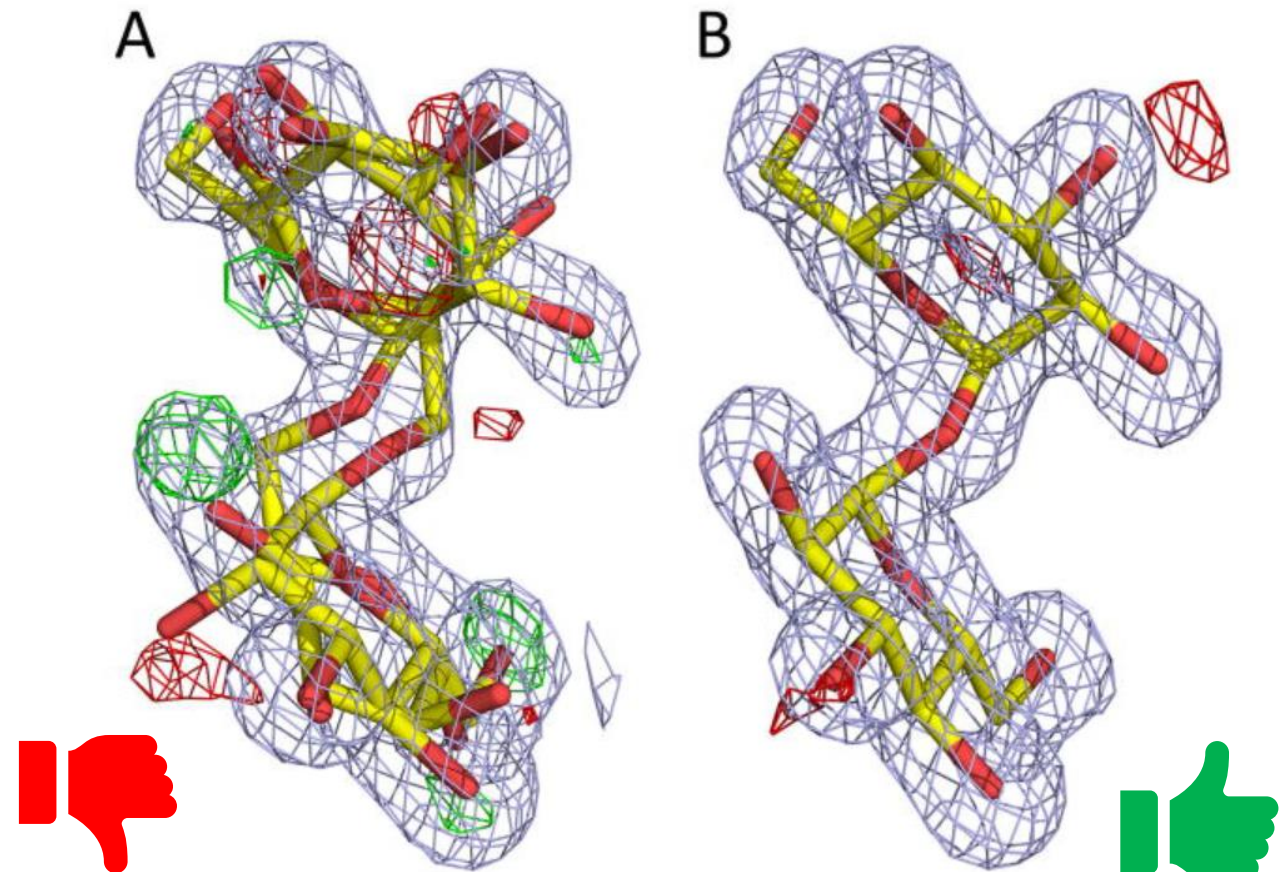
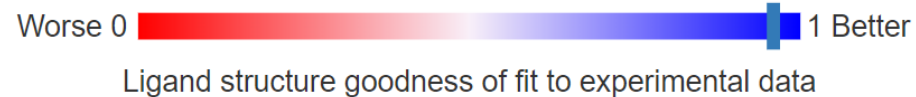


# Kvalita ligandu

- Manuální inspekce
- Nutnost znalosti metodiky

## Validační statistika wwPDB

### Ligand Structure Quality Assessment



**A)** Původně namodelovaná  $\alpha$ -1,6-mannobiosa. Autoři museli vložit alternativní orientaci molekuly, aby interpretovali el. hustotu. To se jim moc nepovedlo a dané dvě konformace molekuly nesedí dobře v el. hustotě.

**B)** Správně interpretovaná el. hustota pomocí molekuly threhalosy.

# Odkazy

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4531100/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3138648/>  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5799025/>
- <https://pubmed.ncbi.nlm.nih.gov/28291756/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5349432/>