CEITEC

Central European Institute of Technology
BRNO | CZECH REPUBLIC

Validation - ligands

Radka Svobodová

Can we still find wrong structures in PDB?

Example: Nipah G attachment glycoprotein (PDB ID 3D12, PNAS)
Contains 30 instances of 11 different carbohydrates, each with one ring and five chiral atoms.

Results:
- 13 of these ligands have incorrect chirality
- In a few cases, all chiral atoms exhibit incorrect chirality
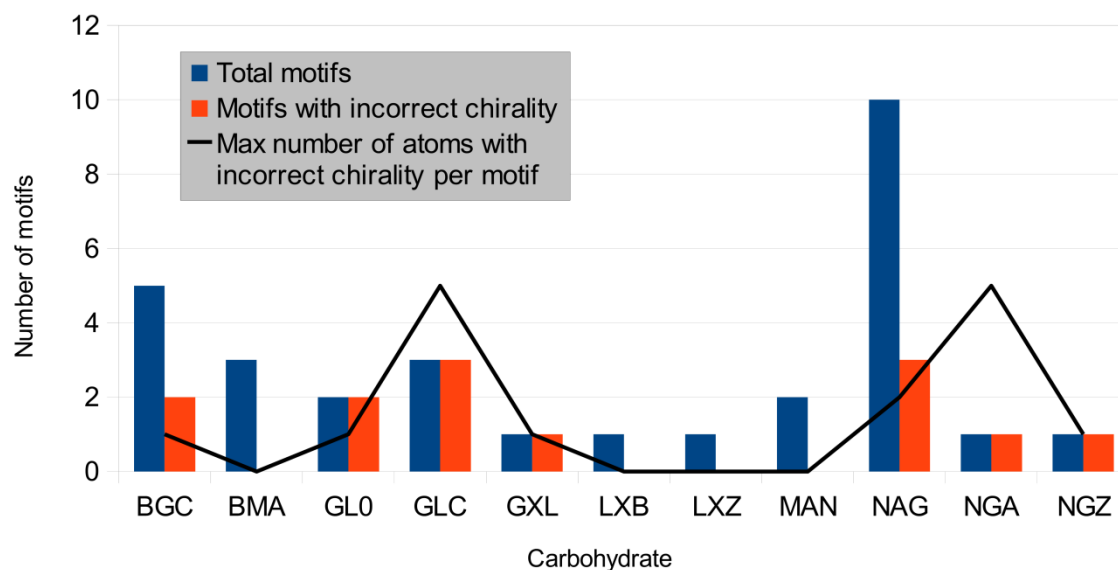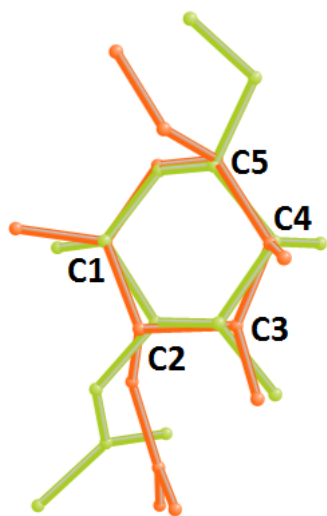
Can we still find wrong structures in PDB?
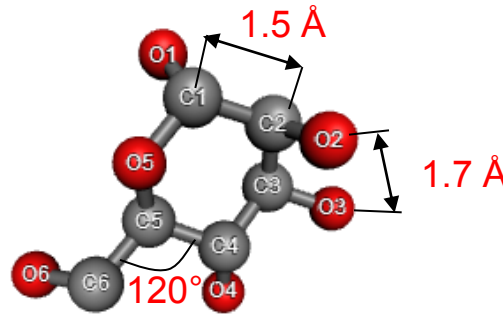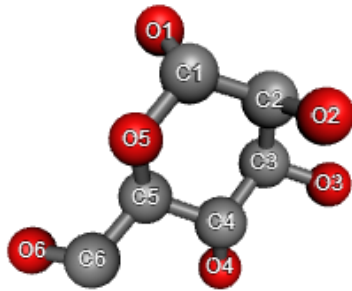
Unfortunately yes

CEITEC

Validation approaches

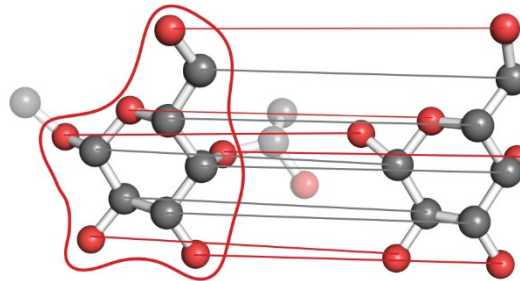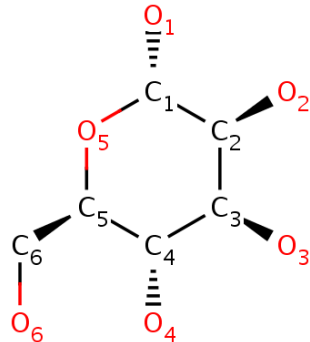| What to validate?	How?	Software
Geometry (3D)	Against tabular values	**Proteins and nucl. acids:** WHAT_CHECK, PROCHECK, PROCHECK-NMR, AQUA, MolProbity, OOPS **Ligands:** ValLigURL, Mogul, Coot, PHENIX
Topology (2D)	Against a template	**Proteins and nucl. acids:** -‖- **Ligands:** pdb-care, MotiveValidator, ValidatorDB

Compilation:
PDB validation reports

Improved data quality

- ## Cleaned up mmCIF data

  - ### Standard vocabularies

  - ### Experimental details, binding sites, secondary structure, antibiotic/inhibitor information, nucleic-acid parameters

- ## Clean mmCIF files are used in production
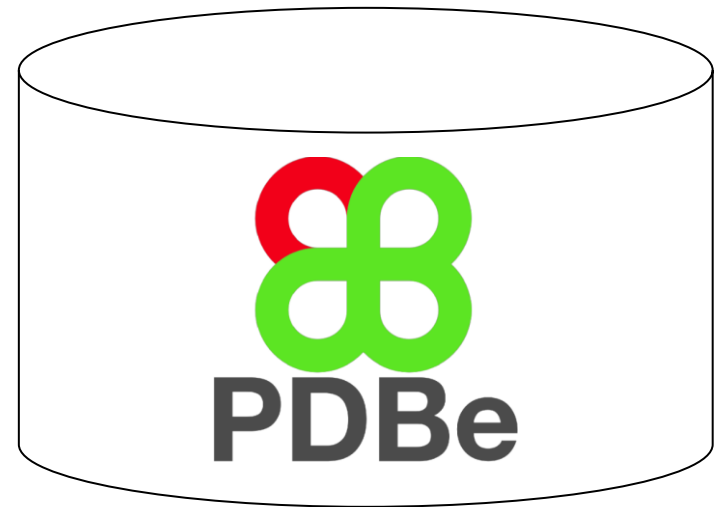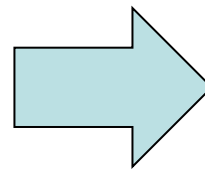
Validation reports

- ## Summary
  - ### Quality vs. all PDB
  - ### Quality vs. entries at similar resolution
  - ### Overview of residue-based quality for every polymer
  - ### Table of ligands that may need your attention

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.80 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.


■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution
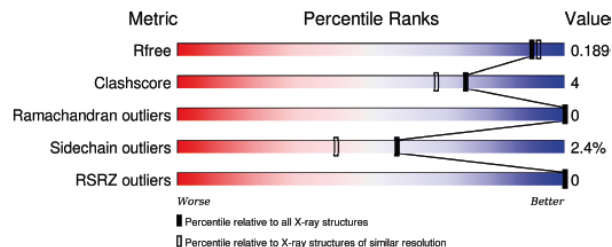
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	4533 (1.80-1.80)
Clashscore	102246	5383 (1.80-1.80)
Ramachandran outliers	100387	5320 (1.80-1.80)
Sidechain outliers	100360	5319 (1.80-1.80)
RSRZ outliers	91569	4547 (1.80-1.80)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	137	91%　　　　　9%

| Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
1	PAQ	A	471	X	-	-	-
1	PAQ	B	471	X	-	-	-
1	PAQ	C	471	X	-	-	-
1	PAQ	D	471	X	-	-	-
6	CU	A	1744	-	-	-	X
6	CU	B	1748	-	-	-	X

CEITEC
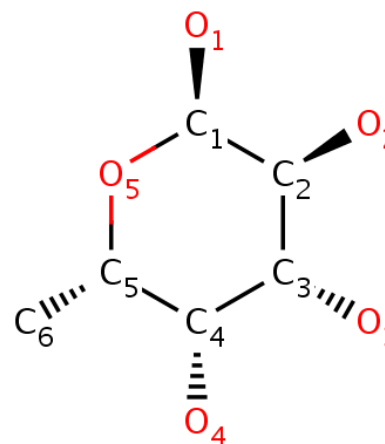
Topology validation (= validation of annotation)
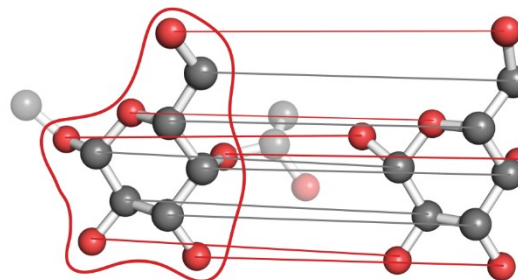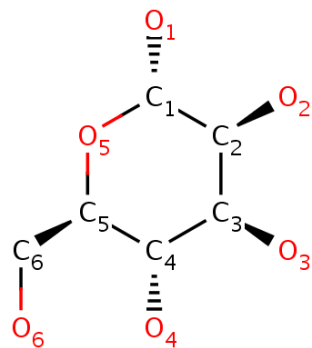
Topology validation – basic terms

Residue

- Any component of a biomacromolecule

- Examples: amino acids, nucleotides, saccharides, ions, …

- In PDB, residue is annotated via "Residue ID" – a unique 3-letter code
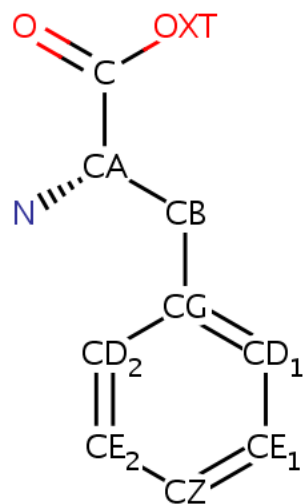
PHE, phenylalanine

FUC, *alpha-L-fucose*

Topology validation – basic terms

Types of residues:

Standard residues:
- amino acids
- nucleotides

Non-standard residues:
- modified amino acids and nucleotides

Ligands:
- Chemical compounds which form a complex with a biomacromolecule (e.g., sugar, drug, heme).
- Also ions are often referred as ligands

CEITEC

Topology validation – basic terms

Principles of topology validation:

- Subjects of topology validation are residues

- Validated residue is compared with a model residue, which has the same Residue ID

- The model residues are taken from a reference database

- Differences between the model residue and the validated residue are reported

CEITEC

Topology validation – approach

Complication with "shared atoms":

- When some residues bind together, one of them can lose an atom:

MAG

FUC

H_2O

PDB ID 1g1t

MAG-FUC
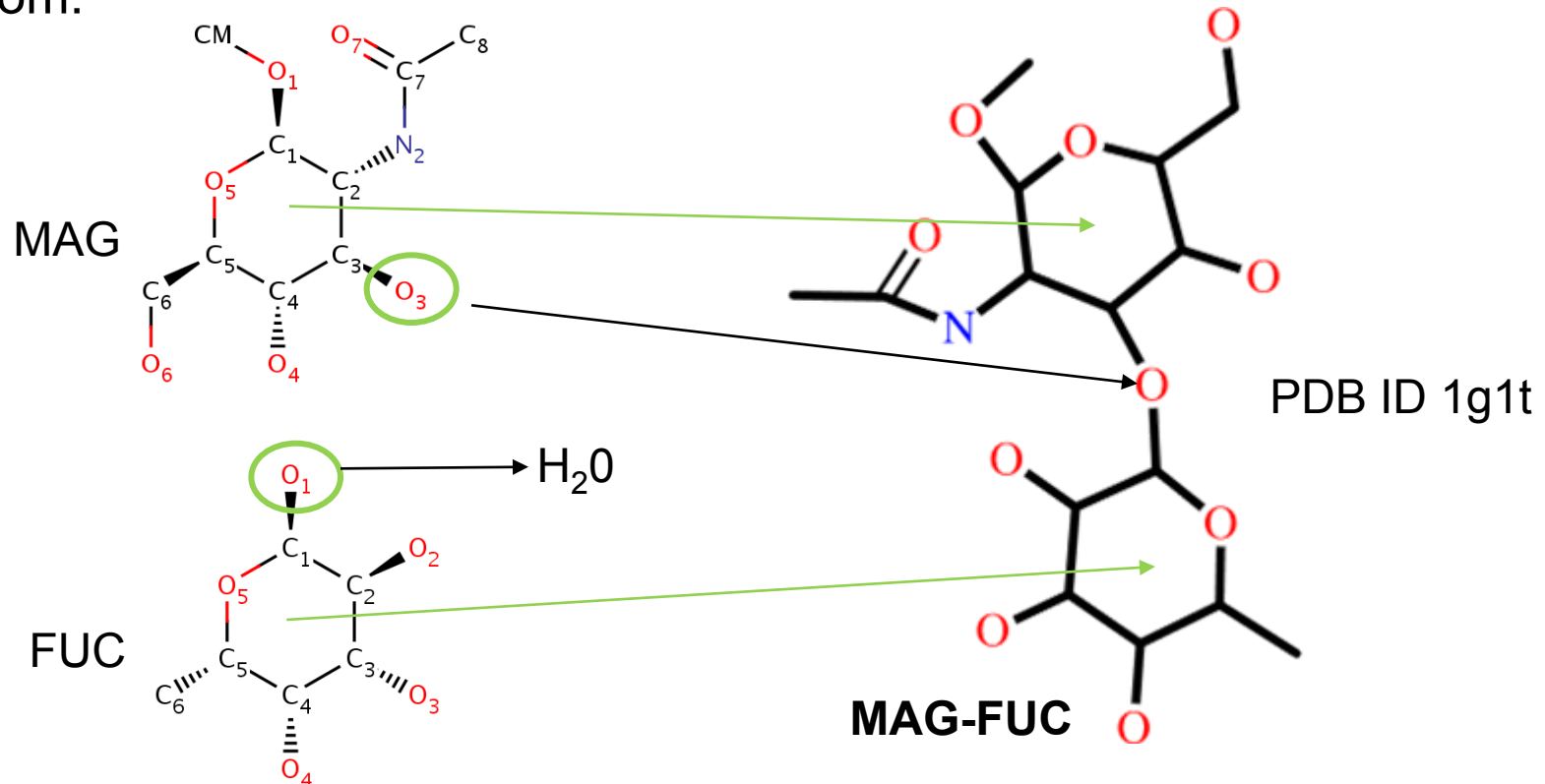
- **Solution:** When we validate a residue, we must include also its close surrounding

CEITEC

Topology validation - approach

Input PDB entry
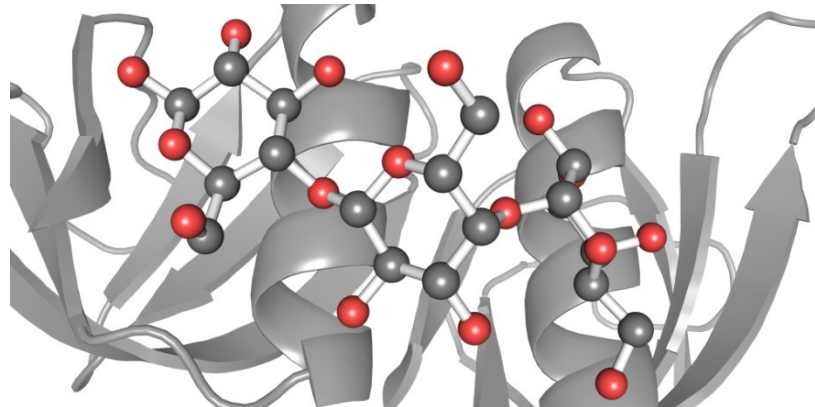
Selection of **validated residue** and its close surrounding

BGC   GAL   GLC
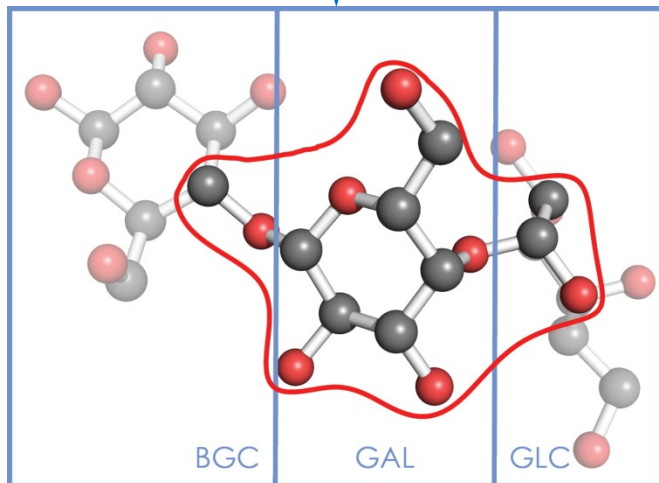
Mapping of **input motif** to the **model residue**
(via subgraph matching)
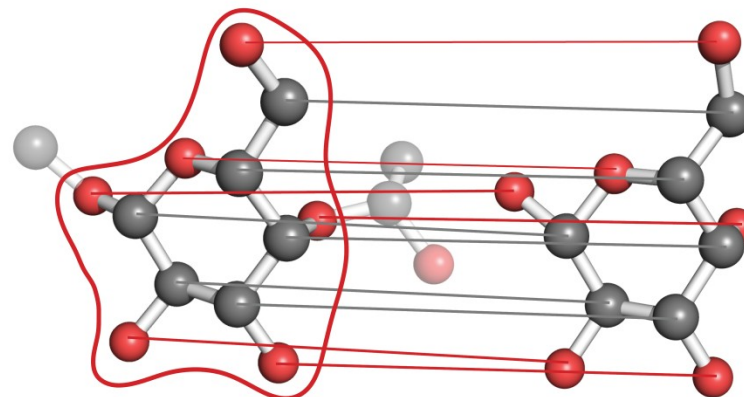
Validated motif mapped to the **Model Residue**

Input motif = validated
residue + surrounding

CEITEC

Topology validation – types of validation analyses

- **Completeness analyses**
  - Missing atoms
  - Missing rings

- **Chirality analyses**
  - Chirality on C atom, metal atom, high bond order atom, planarity

- **Advanced analyses**
  - Substitution
  - Different atom naming
  - Foreign atoms

CEITEC
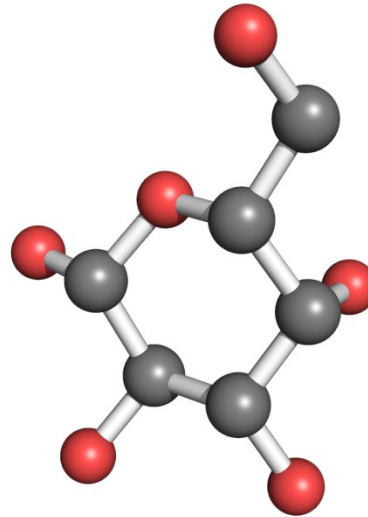
Correct residue

Degenerated structure

1IVG_17_7716 (MAN)

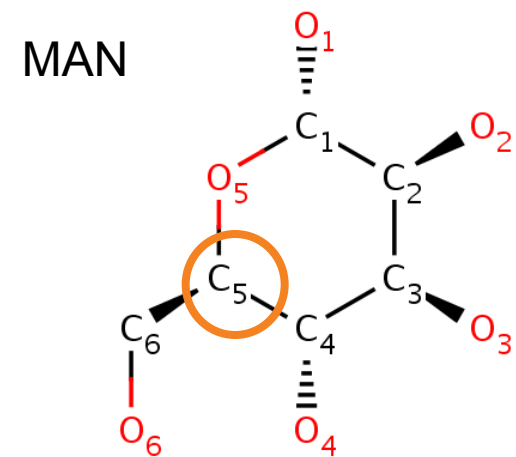## ERRORS – INCOMPLETE STRUCTURE

Correct residue
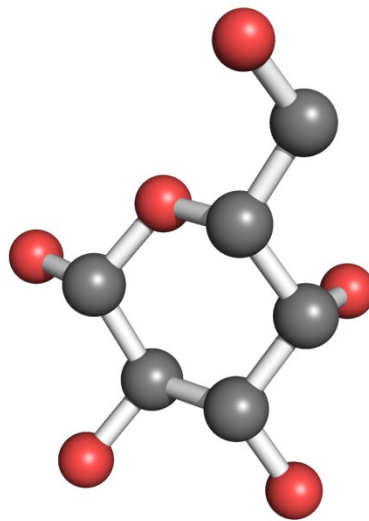
Missing atom
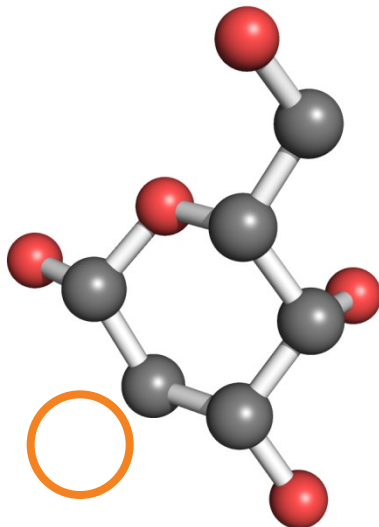
Missing ring
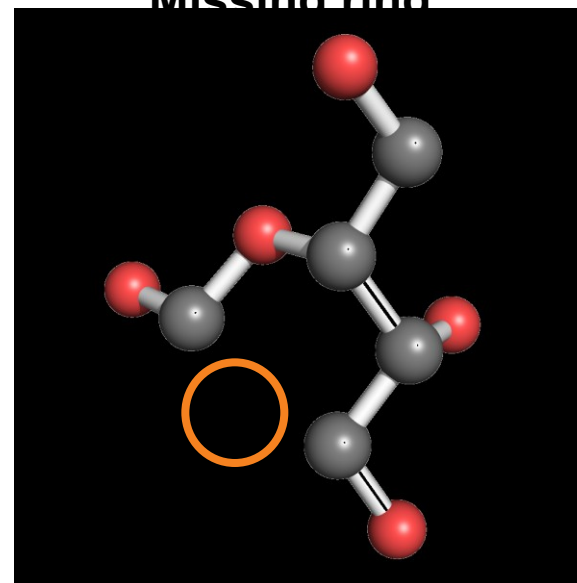
Topology validation – types of validation analyses
ERRORS – CHIRALITY
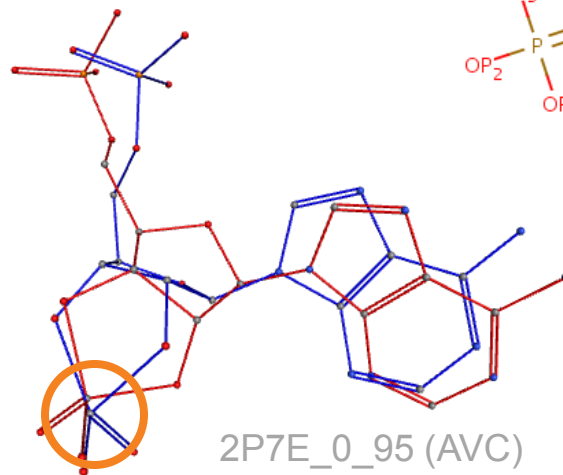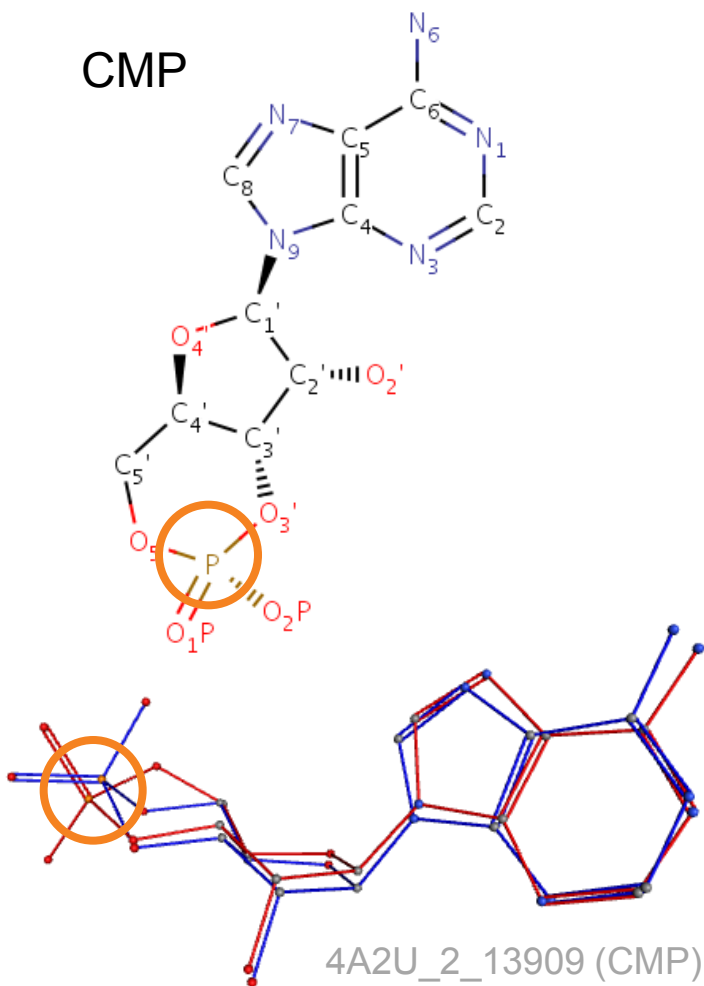
Wrong chirality on C atoms
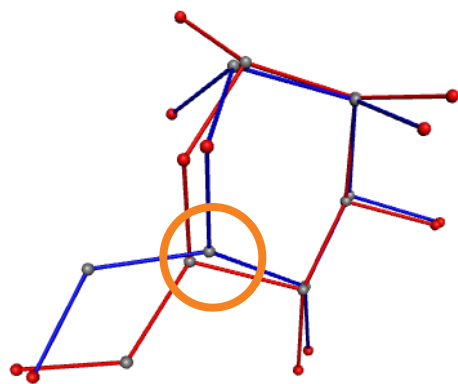
MAN

Wrong chirality on metal atom

AVC

1E4M_16_4280 (MAN)

2P7E_0_95 (AVC)

CEITEC

Wrong chirality on atom having high order bonds

CMP

4A2U_2_13909 (CMP)

Wrong chirality (planar)

HEM

1BZ0_1_4428 (HEM)
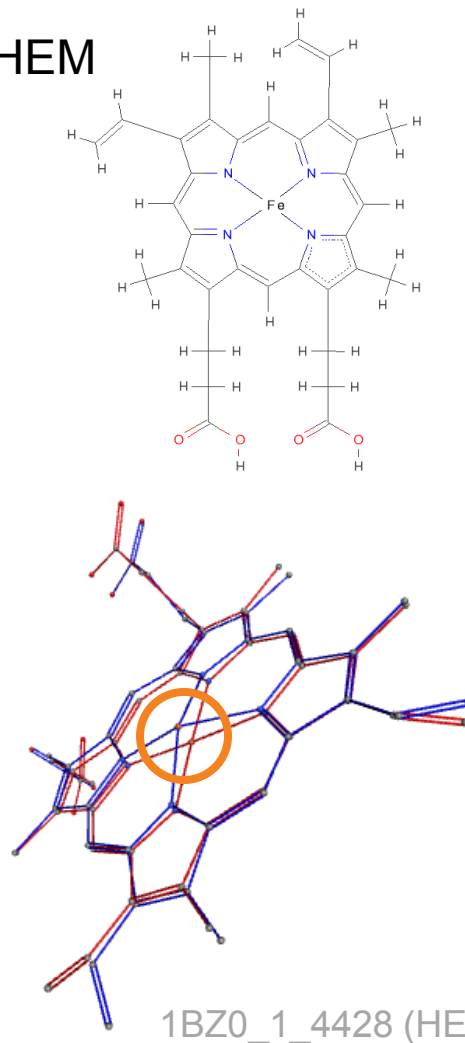
Correct chirality (Tolerant)

=

Correct residues
+
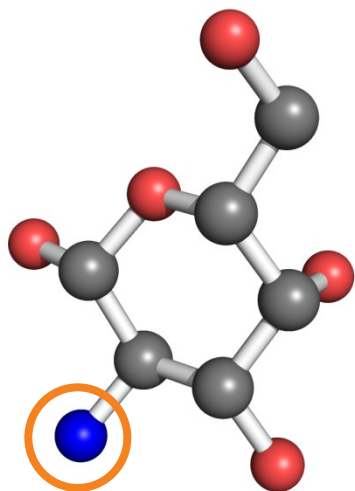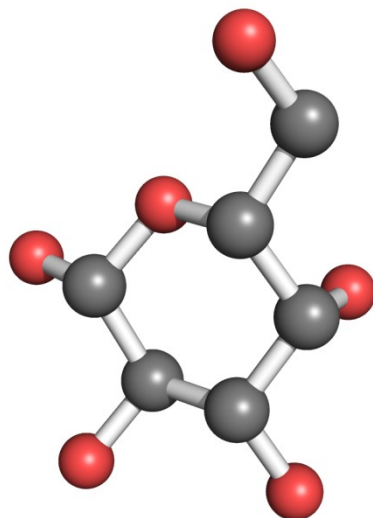**residues having only these issues:**
- **Wrong chirality (planar)**
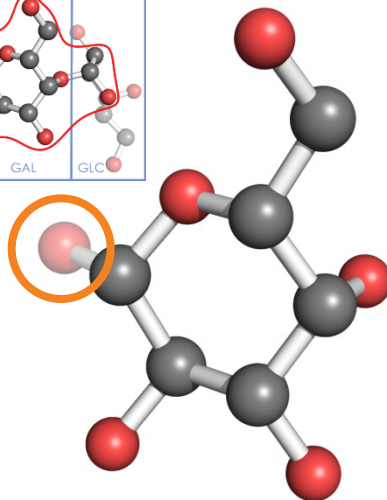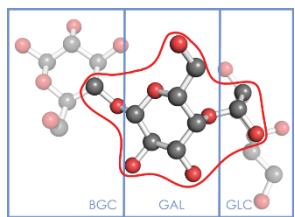- **Wrong chirality on atom having high order bonds**

WARNINGS

Correct residue

ADVANCED WARNINGS:

Alternate locations

Zero RMSD with model

1 instead O5

Substitution

Foreign atom
(= atom from neighboring residue)

Different atom name

CEITEC

Different atom names - example

4ACQ_37_39697 (NAG)
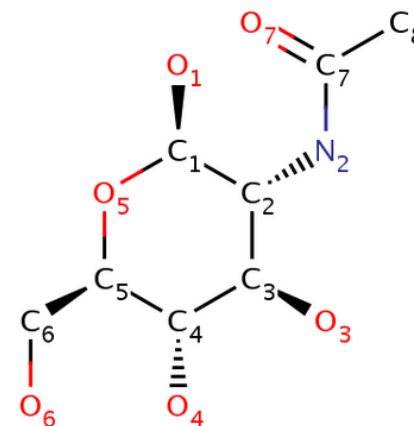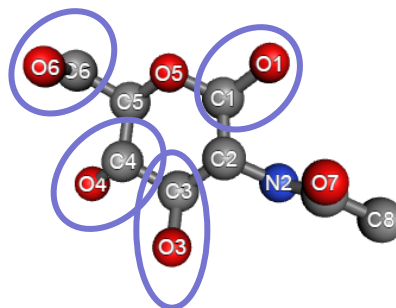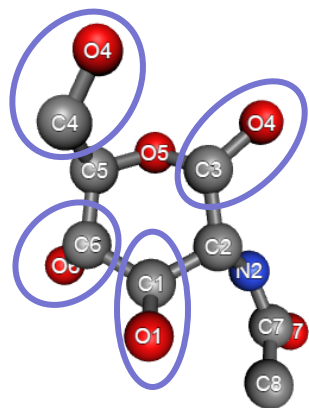
Different Atom Names 7

Model	Motif	Type
C1	C3 C 39699	N N'
C3	C1 C 39697	N N'
C4	C6 C 39702	N N'
C6	C4 C 39700	N N'
O3	O1 O 39706	N N'
O4	O6 O 39709	N E'
O6	O4 O 39707	N E'

CEITEC

Topology validation
Software tools

PDB care (Lütteke et al., 2006):

- Tool focused on carbohydrates validation

- First application, which implements topology validation

- Performs basic validation analyses (missing atoms, missing rings, wrong chirality)

MotiveValidator (http://ncbr.muni.cz/MotiveValidator) :

- Tool, which allows validation of all residues

- Performs basic validation analyses + reports basic warnings (substitutions, foreign atoms, different naming)

ValidatorDB (http://ncbr.muni.cz/ValidatorDB):

- Database, containing validation results for all* ligands and non-standard residues in PDB (weekly updated)

- Performs basic validation analyses + advanced validation analyses (report degenerated residue, distinguish type of chirality error)

- Reports basic warnings + next warnings (Alternate locations, Zero RMSD with model)

* Except amino acids, nucleotides, and small residues (<7 heavy atoms)

Thank you for your attention