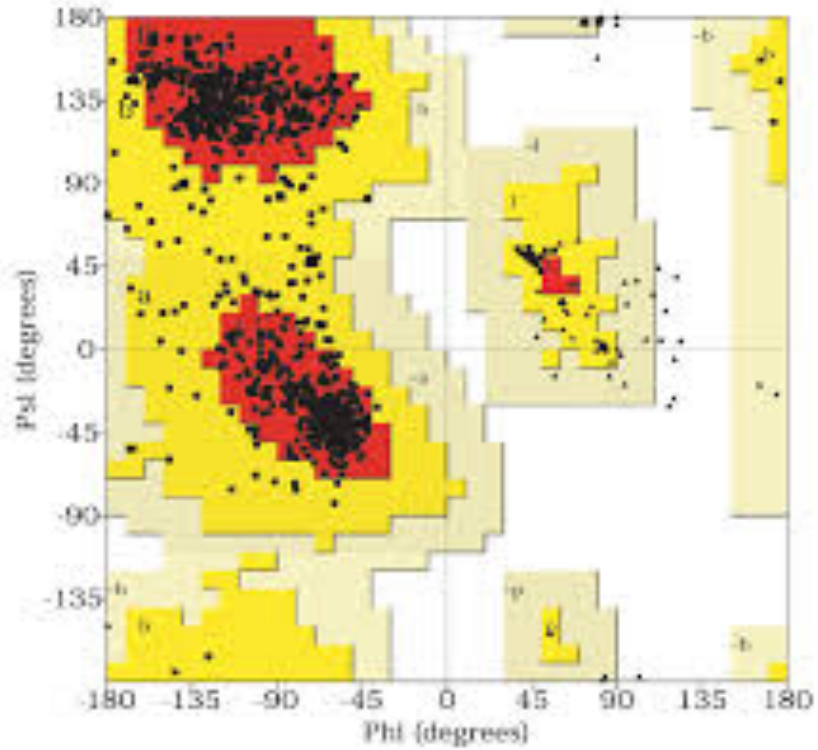
MACROMOLECULAR STRUCTURE VALIDATION

Krisztián Fodor

Eötvös Loránd University, Department of Biochemistry



WHY DOING VALIDATION?

- To check on the validity of the thousands to millions of measurements in the experiment
- To check on how consistent the model is with the experimental data
- To check on the consistency of the model with known physical and chemical properties of the protein

Global vs. Local criteria

- Global evaluation criteria apply to the entire structure
- R-factor, completeness, anisotropy, etc.

- Local evaluation criteria apply to local regions or individual residues
- Bond lenghts and angles, rotamer conformations, steric clashes etc.

What makes a good Model

- Chemistry

- Bond lenghts and angles are close to ideal, flat aromatic rings, good chirality (no D-amino acids unless verified)

- Physics

- Crystal packing is OK, B-factor distribution is good, atoms are bonded, occupancies are 1

- Statistics

- The model well explains experimental data

- Structure

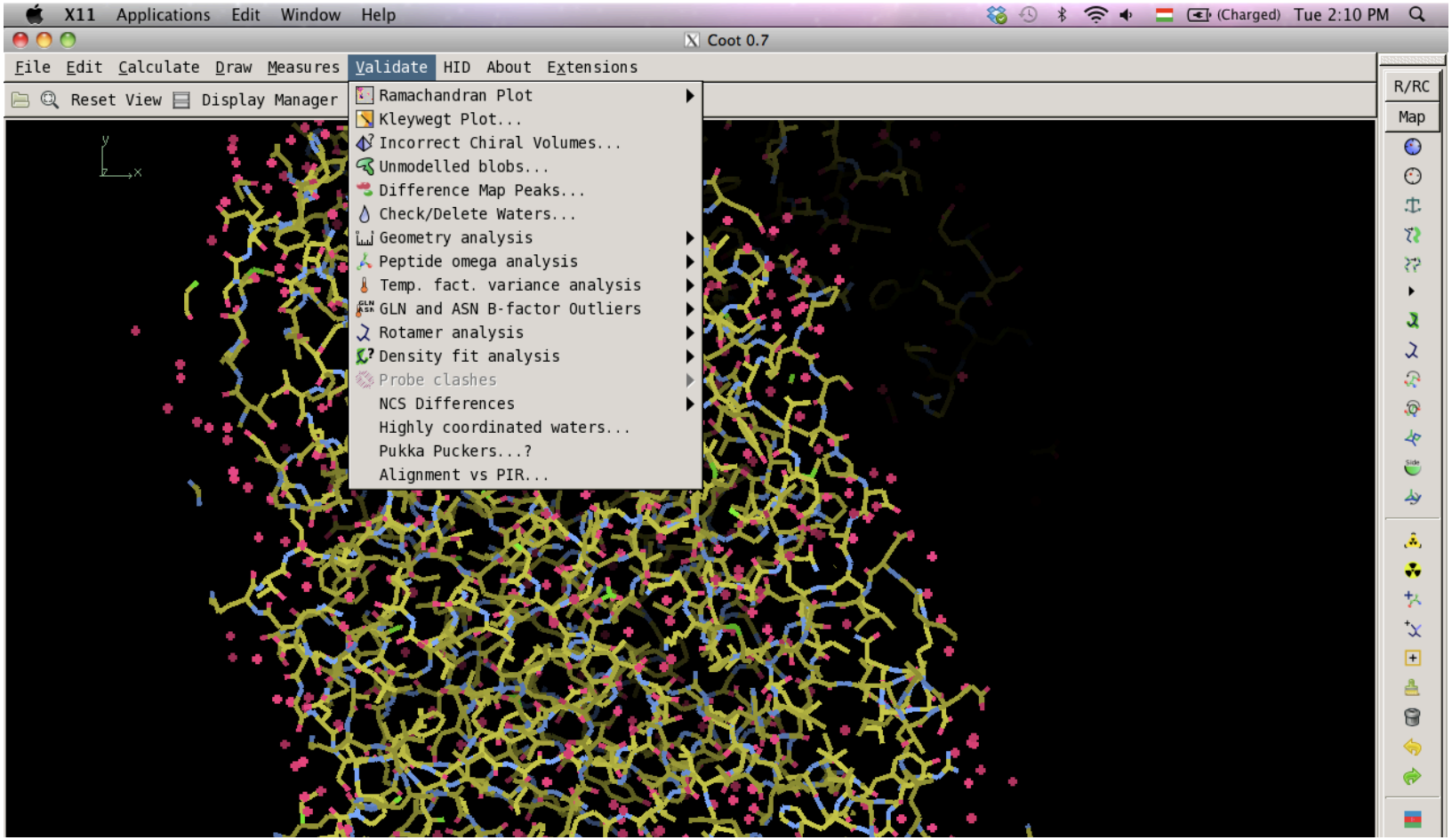
-Reasonable Ramachandran plot, side chain conformation are OK

- Biology

Validation tools

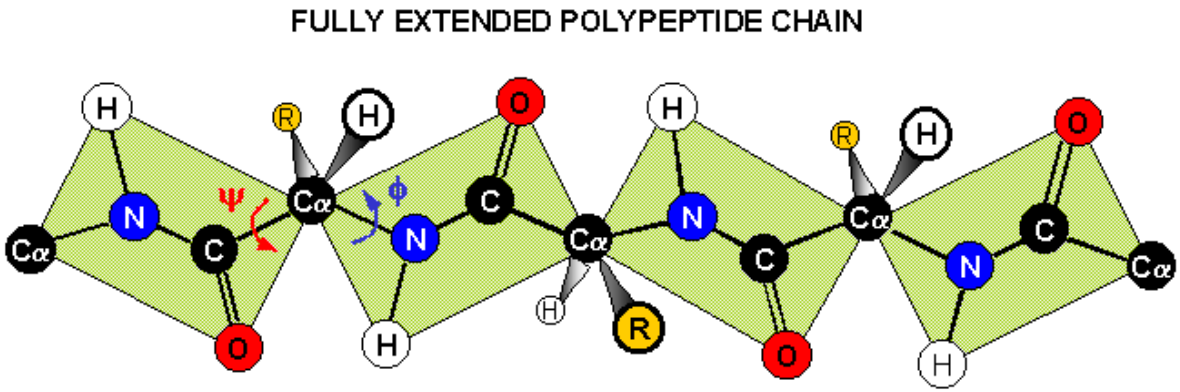
- Coot
- MolProbity
- PROCHECK
- Etc.

Coot


X11 Applications Edit Window Help
(Charged) Tue 2:10 PM
Coot 0.7
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View
Display Manager
Ramachandran Plot
Kleywegt Plot...
Incorrect Chiral Volumes...
Unmodelled blobs...
Difference Map Peaks...
Check/Delete Waters...
Geometry analysis
Peptide omega analysis
Temp. fact. variance analysis
GLN and ASN B-factor Outliers
Rotamer analysis
Density fit analysis
Probe clashes
NCS Differences
Highly coordinated waters...
Pukka Puckers...?
Alignment vs PIR...
R/RC
Map

RAMACHANDRAN PLOT

- The conformation of the protein backbone is defined by three torsion angles
- Phi (Φ)
- Psi (Ψ)
- Omega (ω)

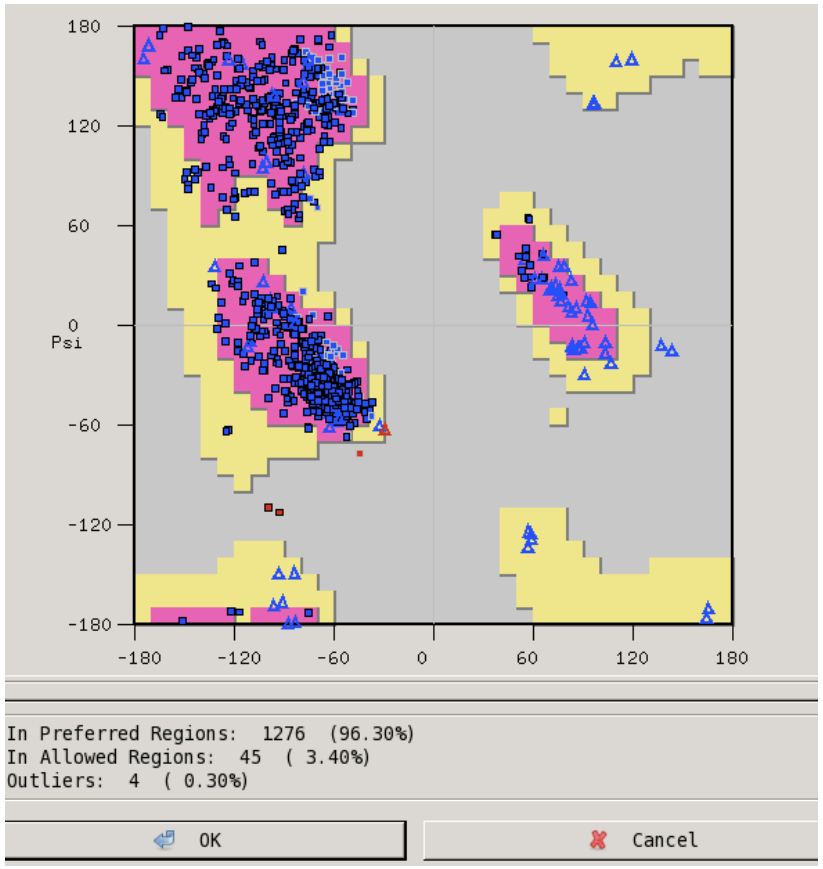


RAMACHANDRAN PLOT

- A scatter plot for psi, phi angles for all residues

Phi (Φ): limited to the range between -60° and 150°

Psi (Ψ): limited to regions centered about -60° and 120°



RAMACHANDRAN PLOT

- **Good models** have most of the residues clustered tightly in the most-favoured regions with very few outliers

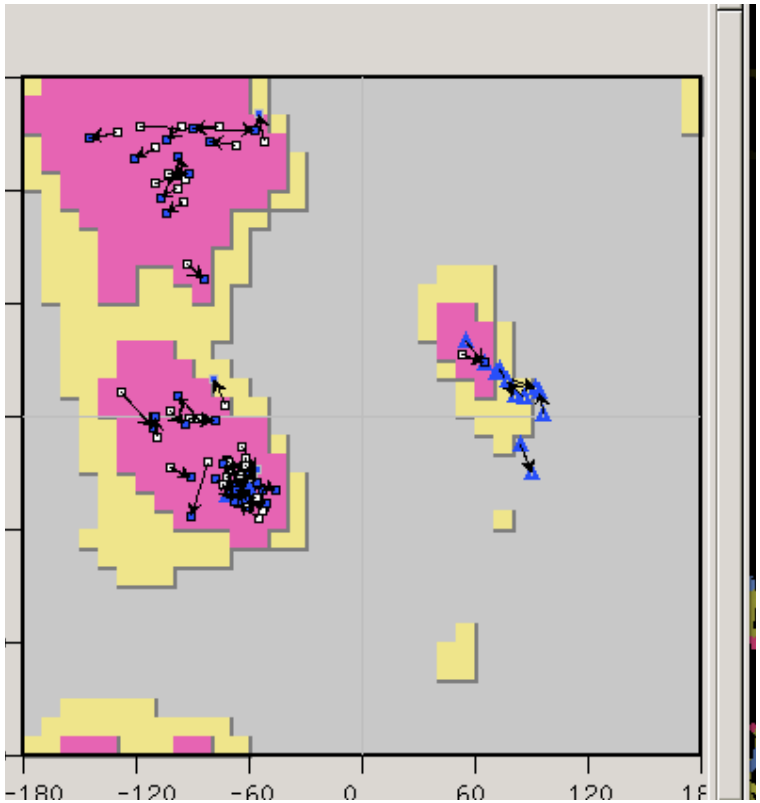
- **Good, but low-resolution models**, may have less pronounced clustering, but still have few outliers

- **Poor models** have no clustering and there are many outliers

From the lecture of Ranu Kadirvelraj, CCRC-UGA, 2008

KLEYWEGT PLOT

- Examine differences between the torsions of NCS-related chains



Other validation tools

- Incorrect chiral volumes
  - Check for chiral centres with wrong handedness

- Geometry analysis
  - Check for improbable bond lengths, angles etc.

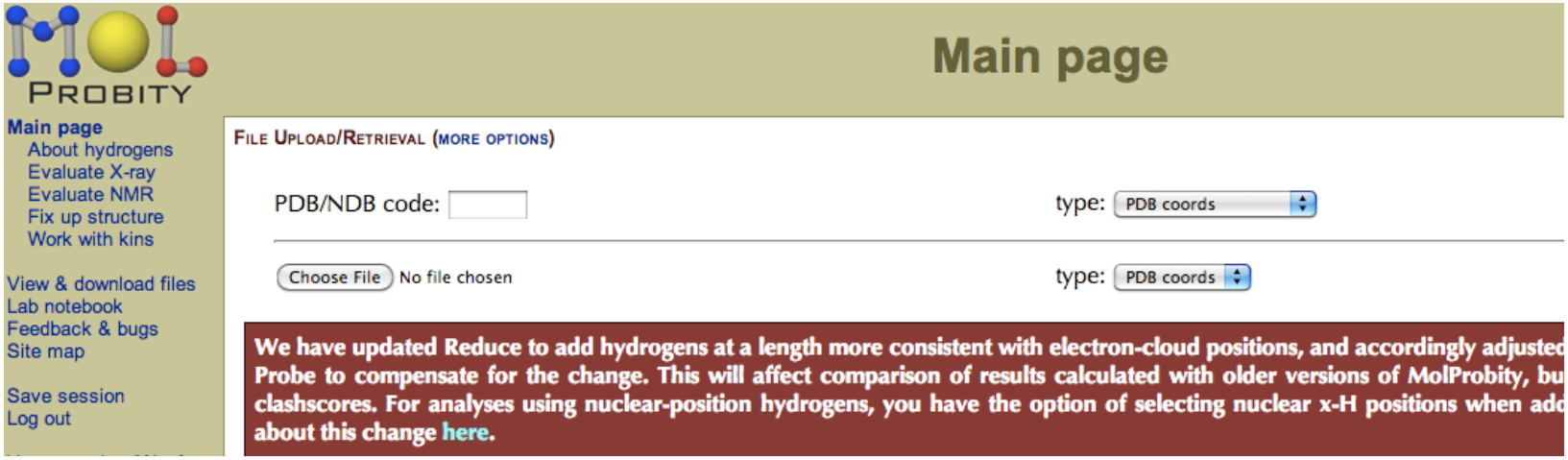
- Rotamer analysis
  - Check for unusual protein side chain conformations

- … some more

MolProbity

- Structure validation web-service
- Based on:

1. Optimized hydrogen placement and all-atom contact analysis
2. Updated versions of covalent-geometry and torsion angle criteria

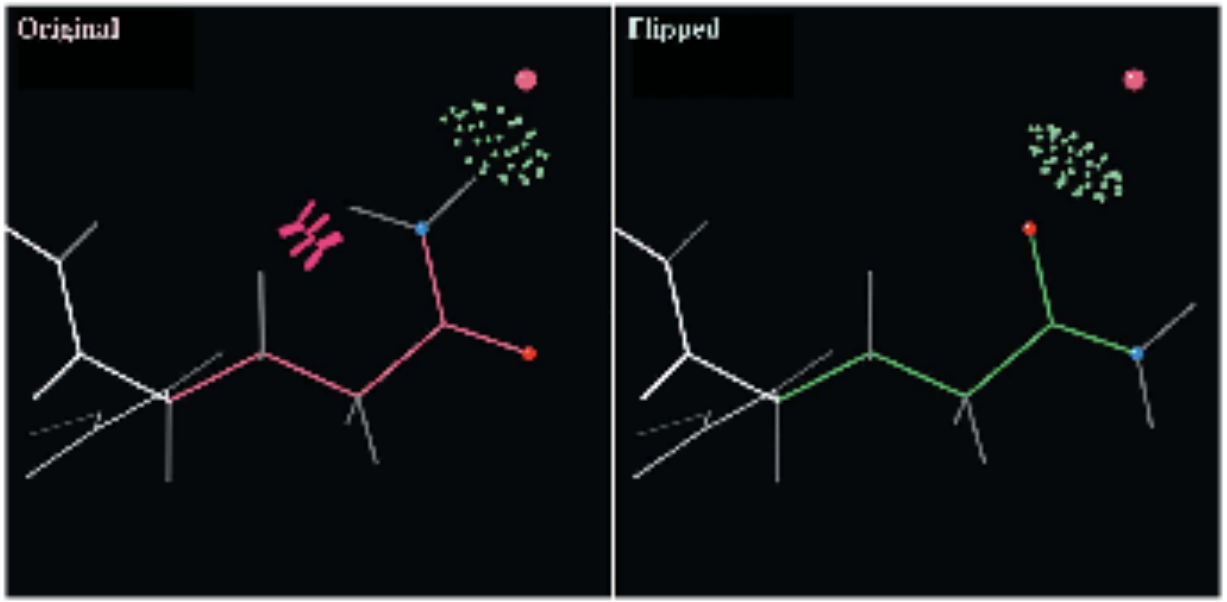


Addition of hydrogen atoms

- The presence of H atoms is a prerequisite for all-atom contacts analysis
- Addition and optimization with the software REDUCE – onto everything except waters

Fitting ASN, GLN and HIS side chains

- A common problem: side chain ends of Asn, Gln and His are often placed 180° backwards
- Automatic correction is the default option of MolProbity (don't forget to download the new file…)



All atom contact analysis

- Once hydrogens are added: all-atom contact and geometry analysis option is enabled
- All-atom contacts:

- Performed by the program called PROBE

A 0.5Å diameter ball is rolling around the van der Waals surfaces of atoms to meaure the amount of overlaps between pairs of non-bonded atoms

All atom contact analysis

- Overlap by more than 0.4Å: serious clash
- Included in the reported clashscore
- One of the atoms is modeled incorrectly

- Clashscore: the number of serious clashes per 1000 atoms (percentile rank within the relevant resolution range is also given)

RAMACHANDRAN AND ROTAMER ANALYSES

- Reference distributions: 100 000 residues in 500 files, quality filtered
- Absolute goals:
  - >98% for Ramachandran favored
  - <0.2% for Ramachandra outliers
  - <1% for poor rotamers
  - 0 for C^{β} deviation outliers

Protein Geometry	Poor rotamers	20	1.81%	Goal: <1%
	Ramachandran outliers	2	0.15%	Goal: <0.05%
	Ramachandran favored	1292	97.14%	Goal: >98%
	MolProbity score^	1.51		98th percentile* (N=10167, 2.20Å ± 0.25Å)
	Cβ deviations >0.25Å	2	0.16%	Goal: 0
	Bad backbone bonds:	0 / 5376	0.00%	Goal: 0%
	Bad backbone angles:	0 / 6706	0.00%	Goal: <0.1%

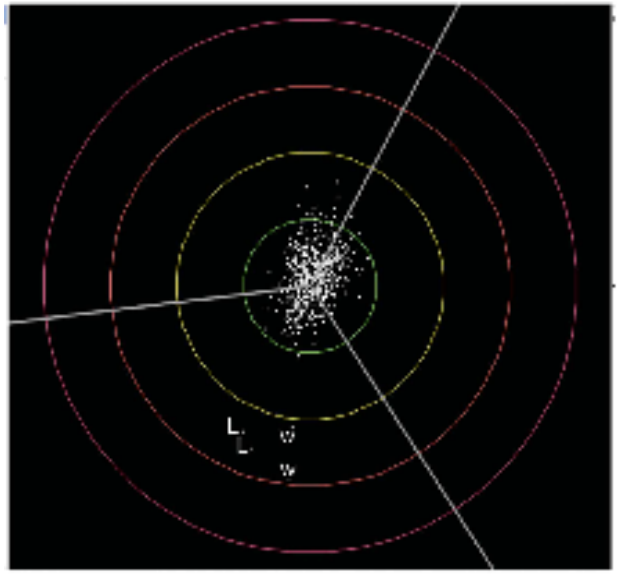
In the two column results, the left column gives the raw count, right column gives the percentage

- Poor rotamers: not outliers, since they are disfavored, but can be stabilized by tight packing or hydrogen bonds

Covalent-geometry analyses

- Local geometry outliers are often the result of misfitting (useful diagnostics, especially for bond angles)

- C^{β} deviation: overall distortion from ideality



Reasons of geometric distortions

- Low resolution (alternative conformations)
- Radiation damage
- Building problems (human factor)

The overall MolProbity score

- People like to describe things with one single number…
- Log-weighted combination of :
  - Clashscore
  - Percentage of Ramachandran not favored
  - Percentage of bad side-chain rotamers

MPscore = 0.426 *ln(1+clashscore) + 0.33 *ln(1+max(0, rota_out|-1)) + 0.25 *ln(1+max(0, rama_iffy|-2)) + 0.5

If smaller than the actual resolution: better than the average structure at this resolution