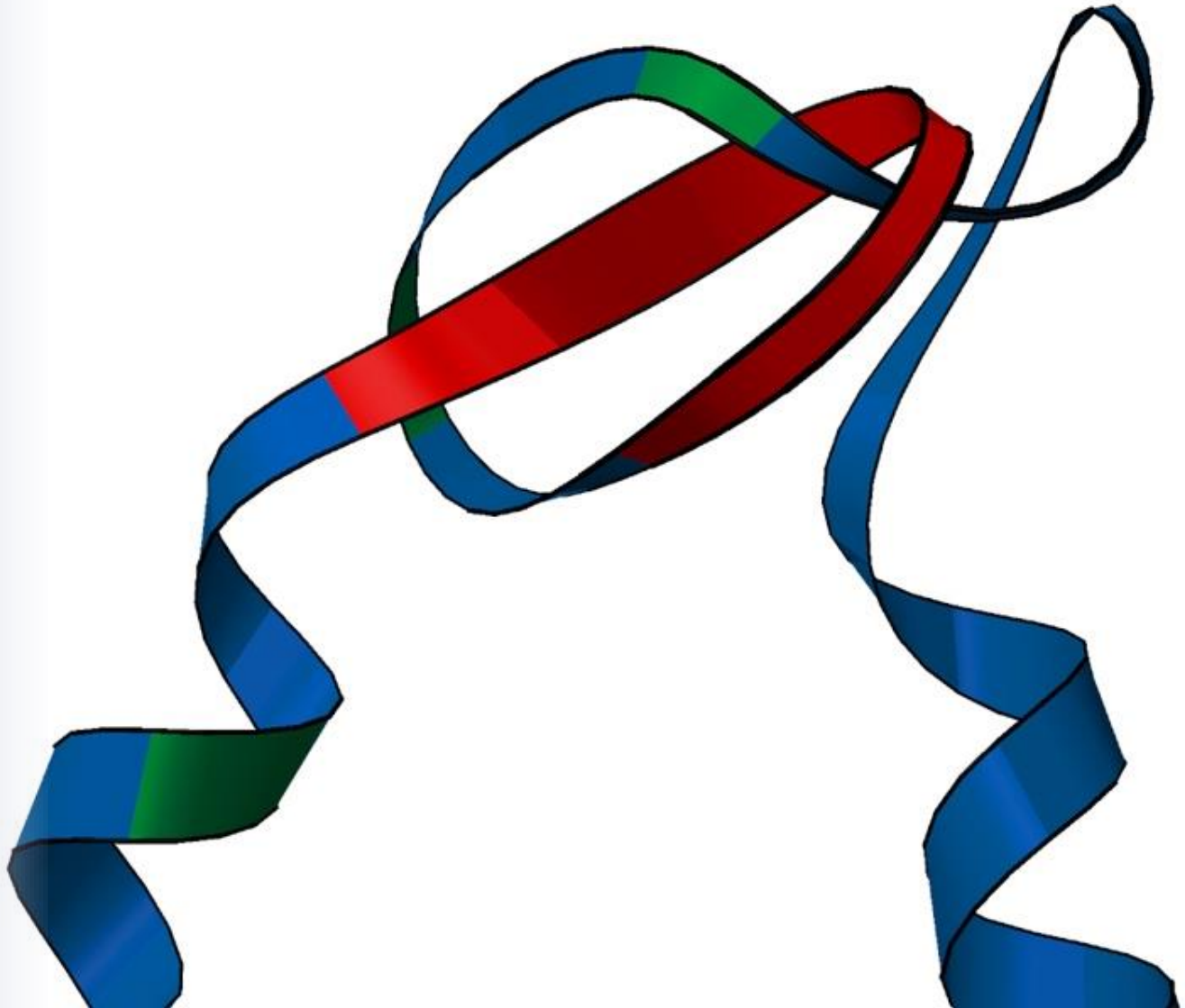
Structural Validation with MolProbity and Phenix

Phenix Users Meeting 2026

Vincent Chen

Duke University



Learning Outcomes

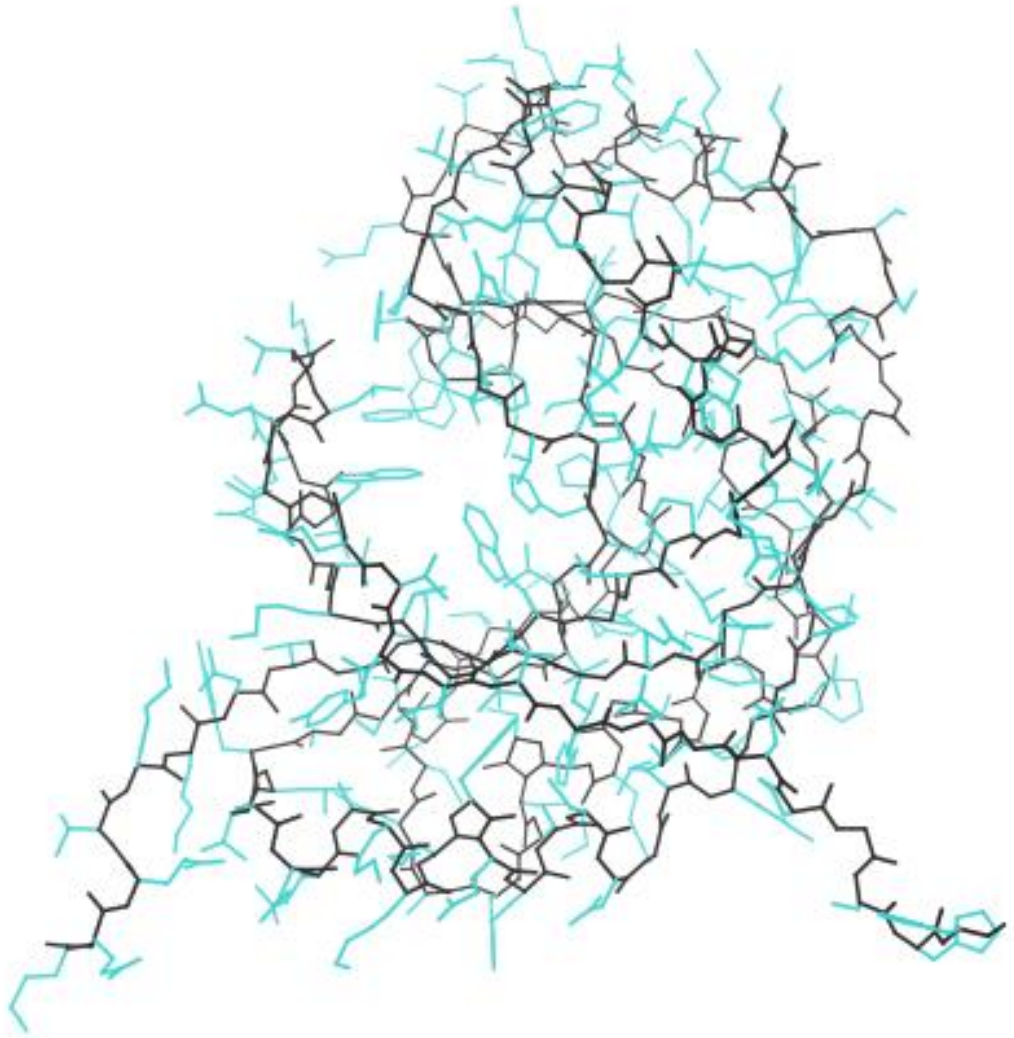
By the end of this session, you should be able to:

- Describe the purpose of structure validation
- Recall a few software tools and commands to diagnose errors in structural models
- Locate and identify the different types of errors you may encounter in structural models
- Apply different strategies to correct those errors

INTRODUCTION & OVERVIEW

Part 1

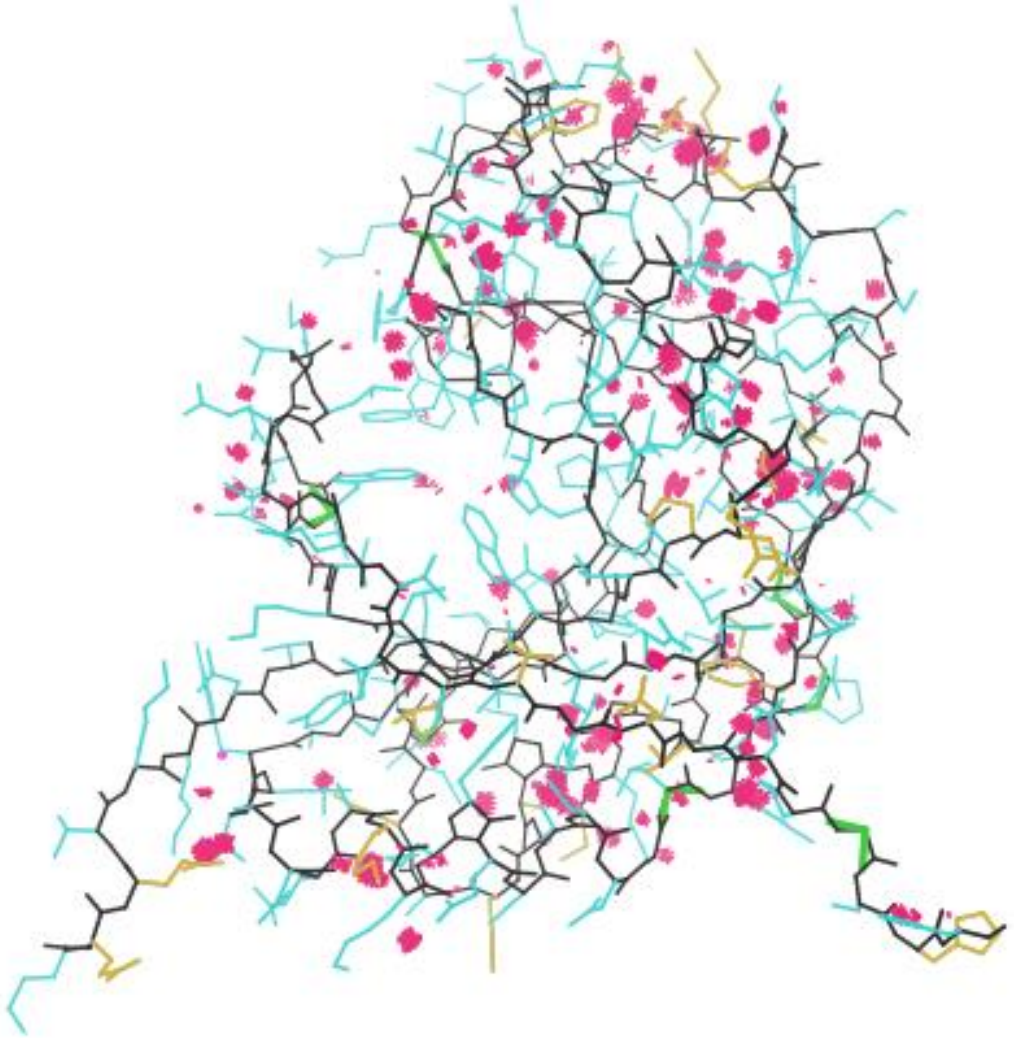
What Is Structural Validation?



Cyclic Nucleotide Phosphodiesterase

Can you tell where this model has errors?

What Is Structural Validation?



Cyclic Nucleotide Phosphodiesterase

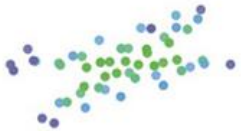
How about with errors shown?

“Proofreading” a model for what is possible and what is probable - checking model quality using independent criteria

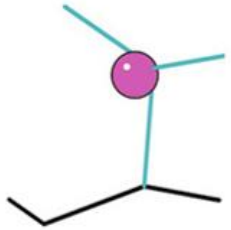
Types of errors

MolProbity's visual markup lets you **see** outliers in their local context.

There are other errors that we don't specifically check for (e.g. register shifts, underpacking).

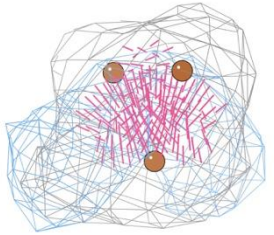


All-atom contacts/clashes

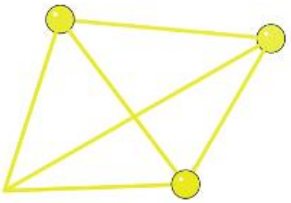


C-beta deviations

Peptide cis/twisted

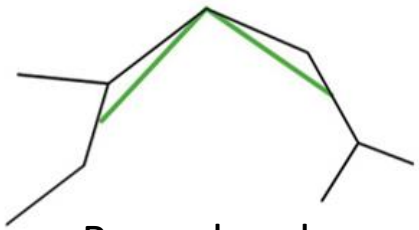


water analysis

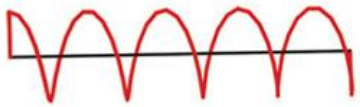


chiral centers

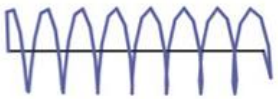
Sidechain rotamers

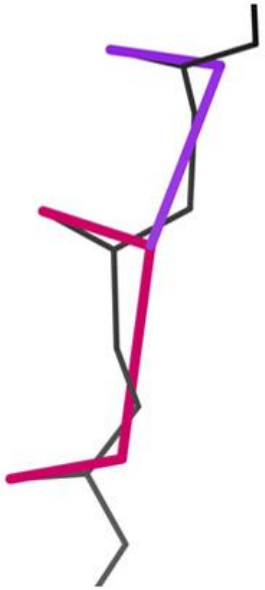


Ramachandran



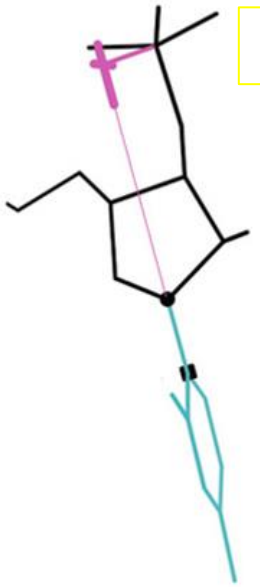
bond angles/lengths



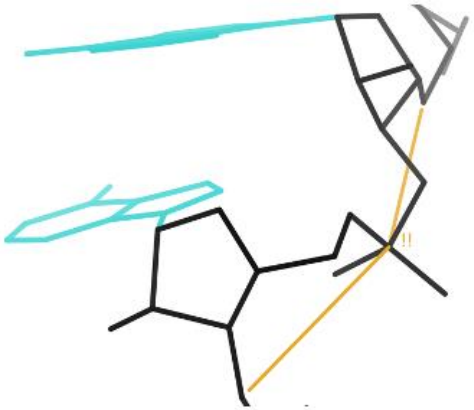


CaBLAM

RNA



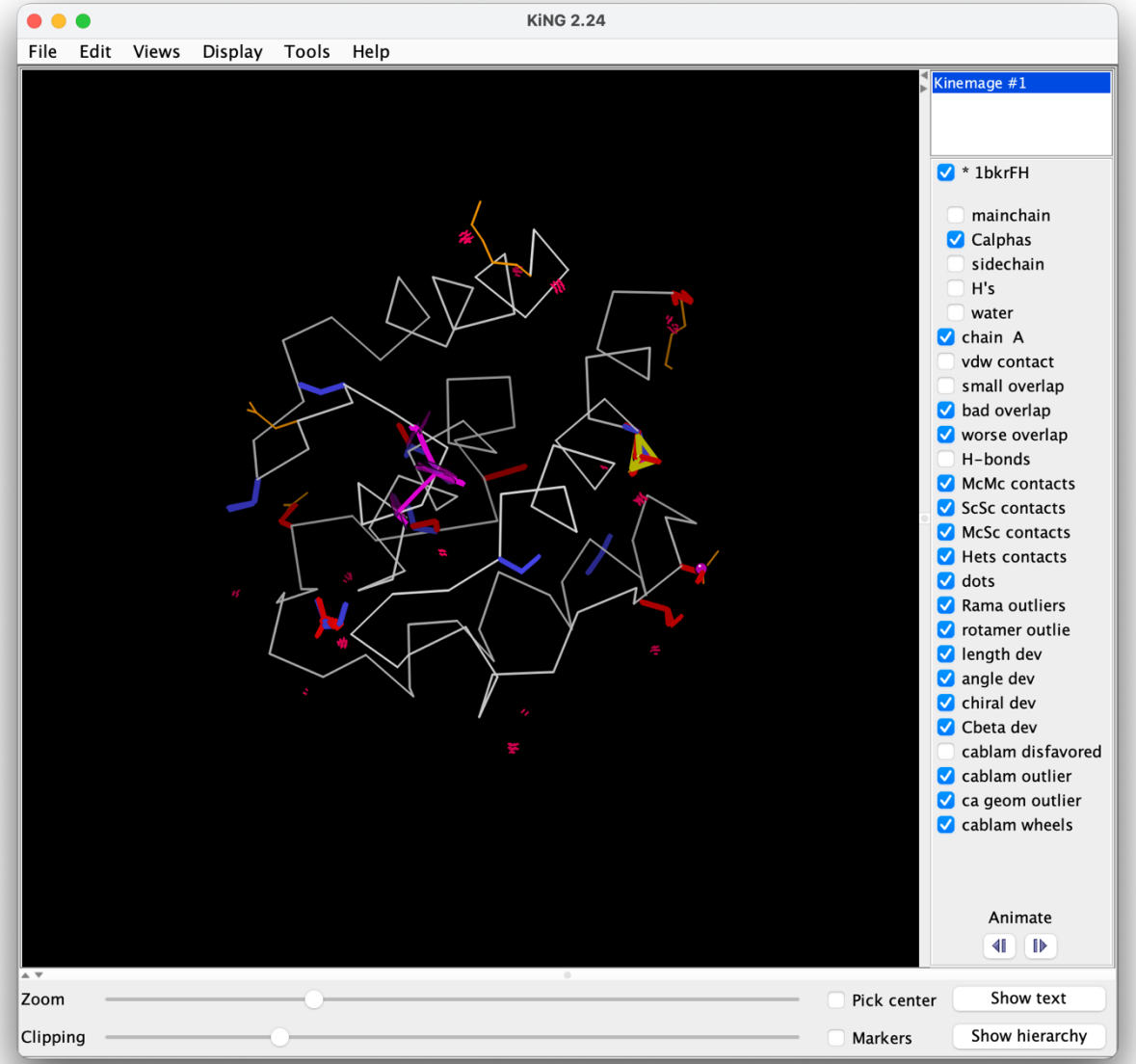
RNA puckers



Suite analysis

KiNG and kinemages (phenix.king)

- Yet another visualization software package.
- “Kinematic image” - interactive graphics scientific illustration format.
- Relatively simple text file, human-readable, that represents display objects (dots, lines, triangles, etc.) and their properties.
- KiNG was the onlyway to see all of our validation metrics, but we are adding it to other viewers.



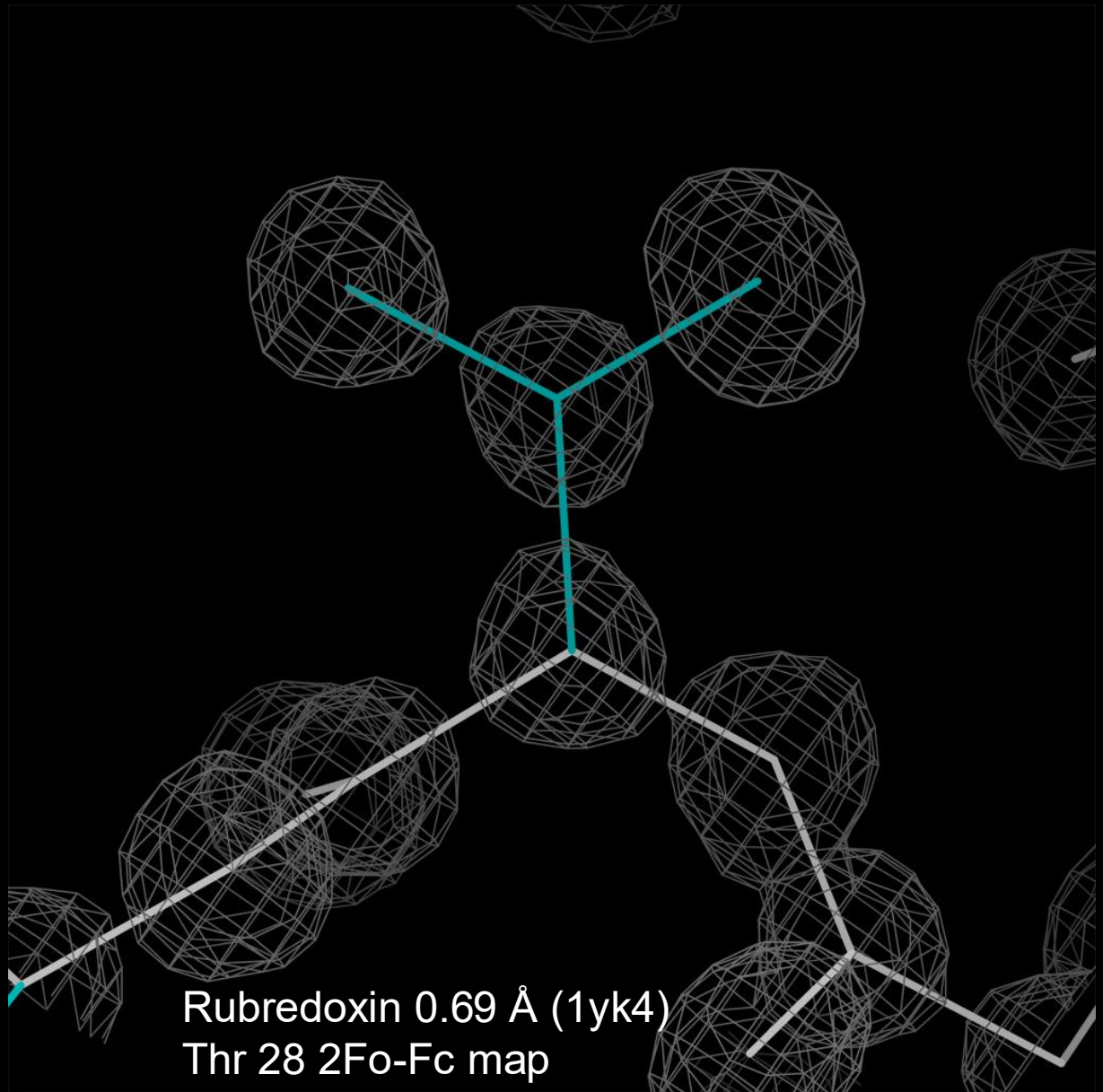
Why We Need Structure Validation

No way to directly see atoms: structure determination methods provide *models* of structures

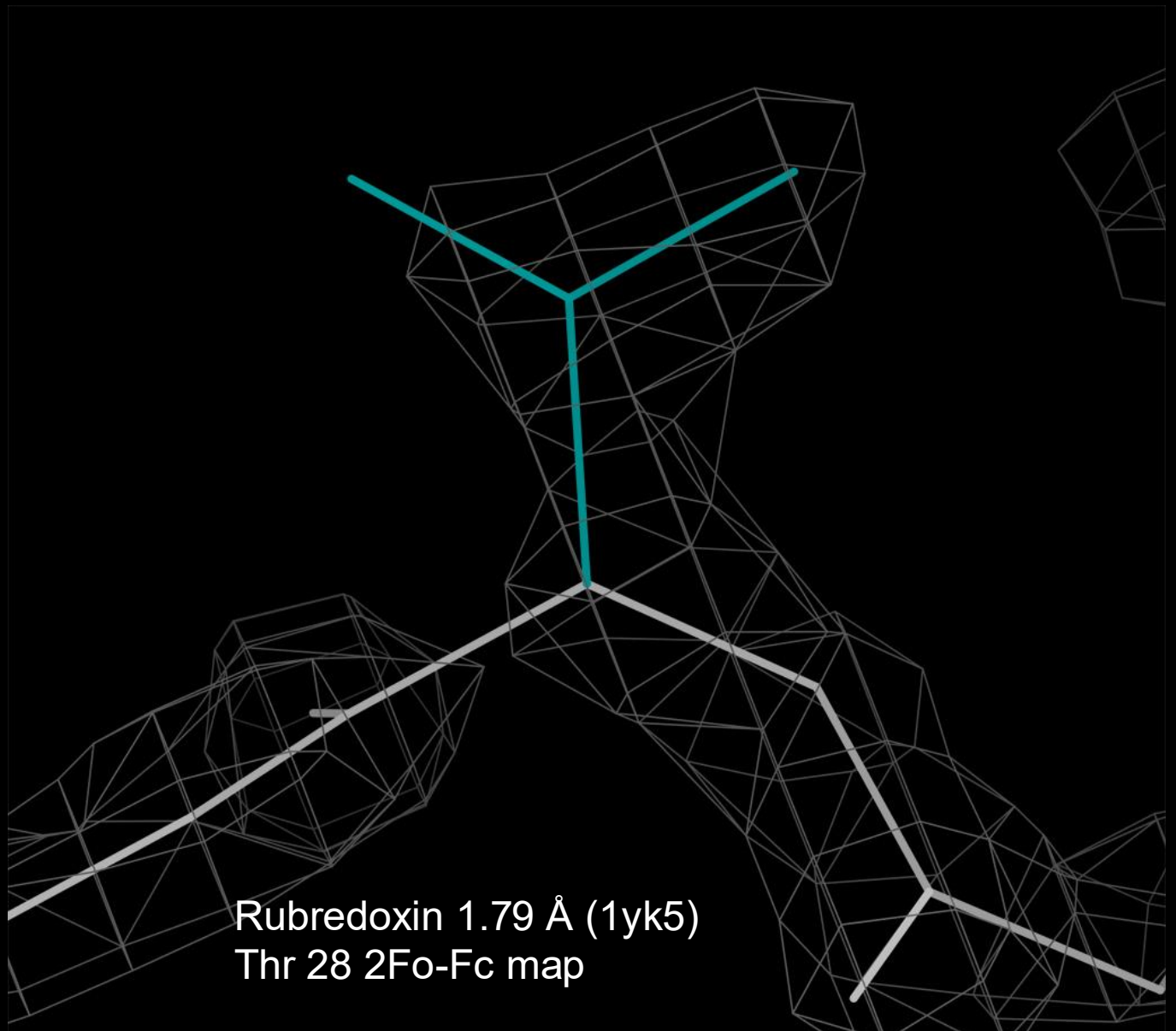
Data is limited (resolution, crystal issues/damage, etc) and data interpretation can be subjective.

Errors propagate to subsequent structures and to downstream biology

Solving structures can be easy…

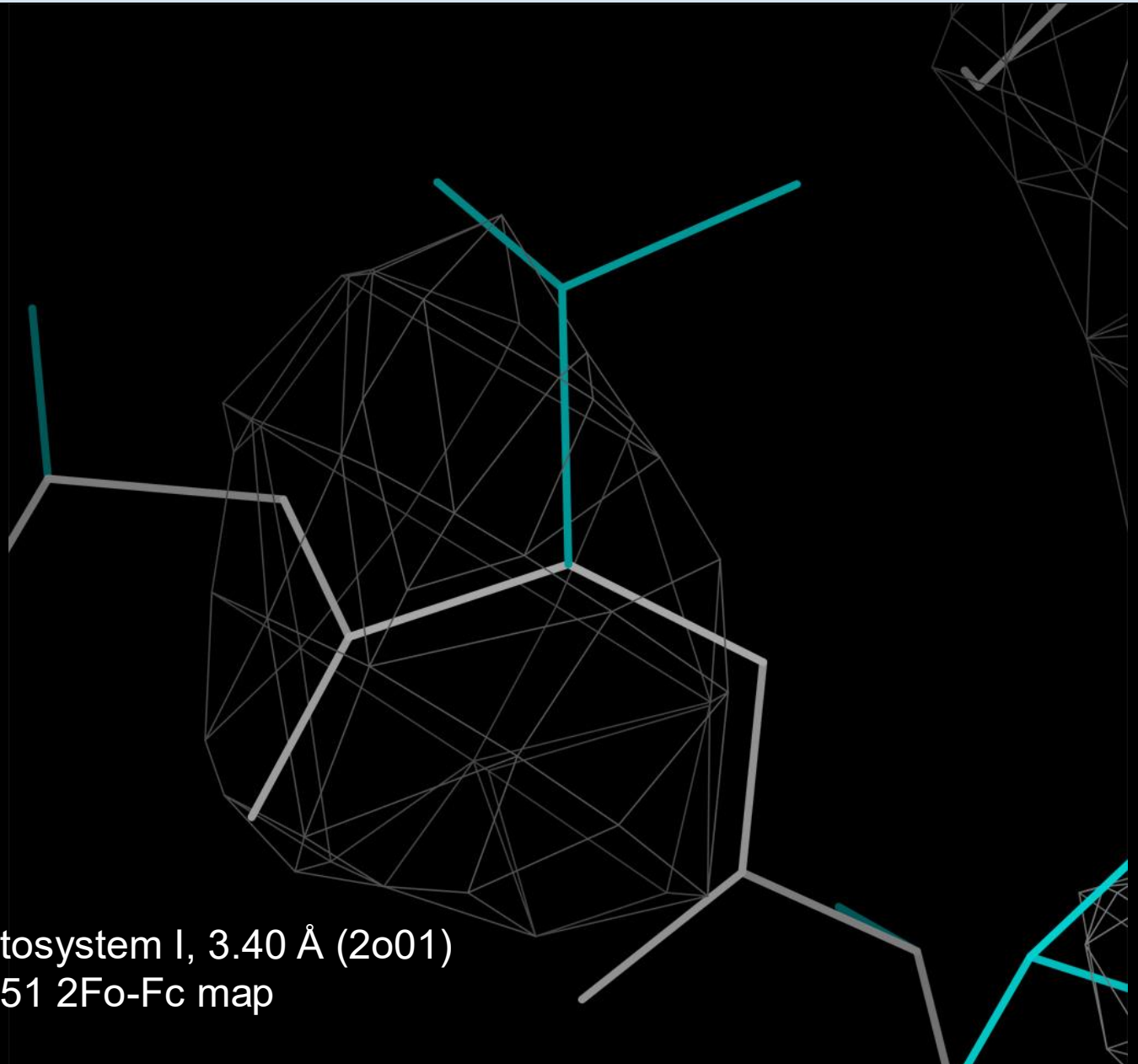


…but is usually harder…



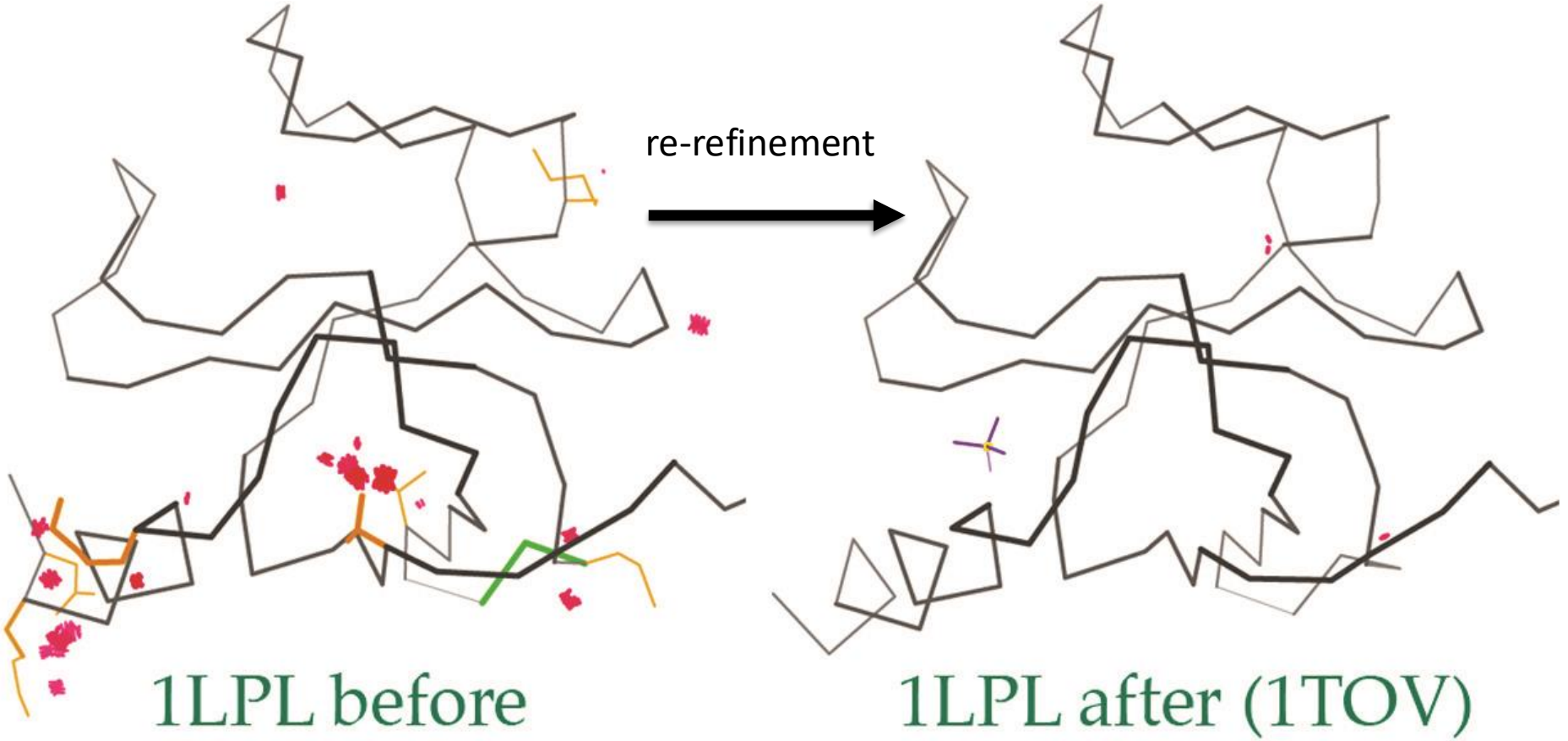
Rubredoxin 1.79 Å (1yk5)
Thr 28 2Fo-Fc map

…and in the worst case…



Photosystem I, 3.40 Å (2o01)
Thr 51 2Fo-Fc map

What's the goal?



Guiding human intervention into models

Richardson Lab Validation Philosophy

Your eyes can be a very powerful tool for finding errors

Global quality scores often obscure local problems in structures

Zero outliers is **NOT** the goal: outliers can be real and interesting!

Try to keep independent measures for validations, if you use it as a restraint, the validation becomes less helpful

Where Do Validation Criteria Come From?

Physical/chemical first principles

Contacts, steric clashes

Bond geometry (lengths and angles)

cis/trans peptide bonds

Cβ deviation

original Ramachandran

Empirical surveys of high-quality structures

Updated Ramachandran (Top8000)

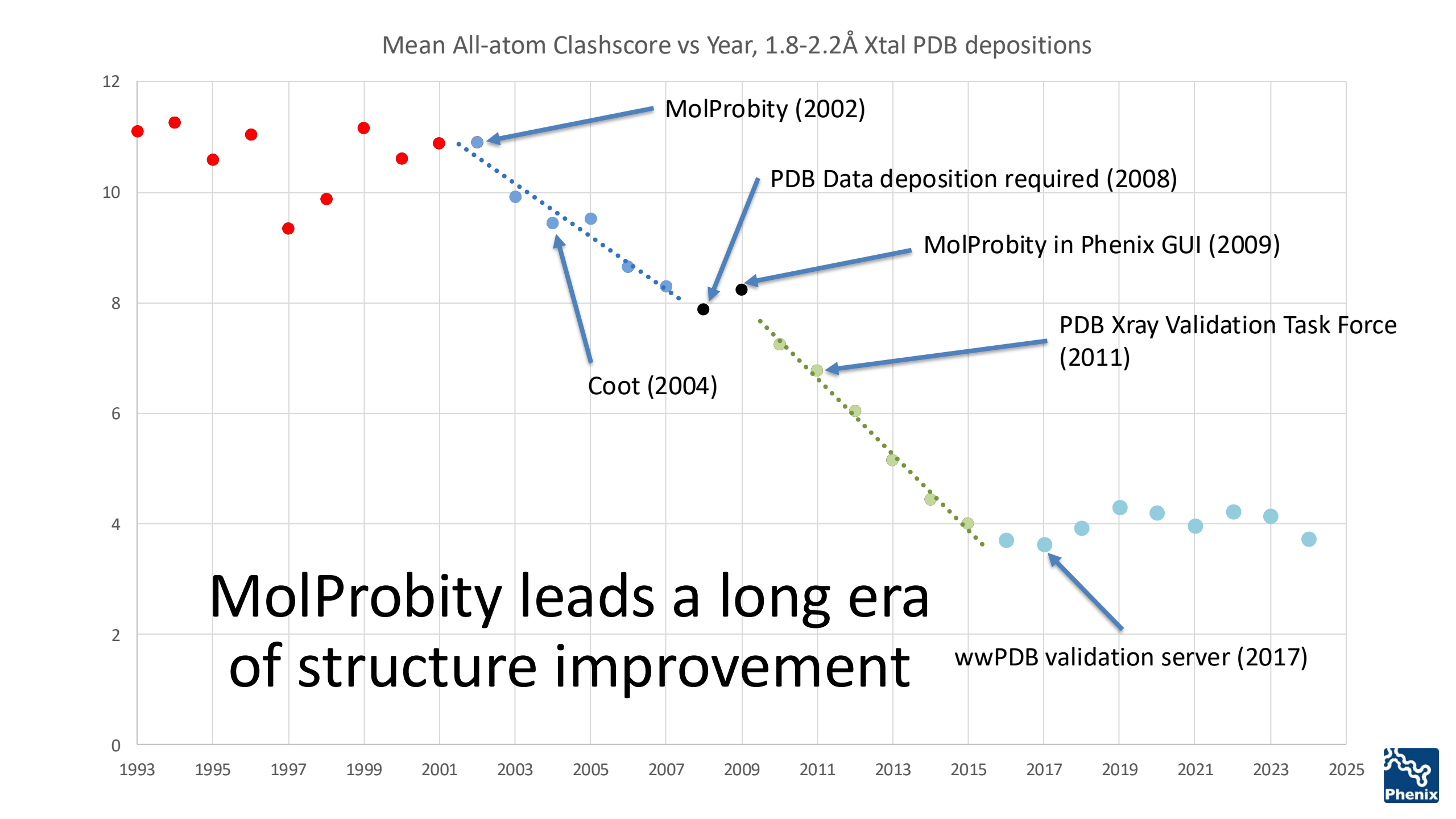
Rotamer libraries

CaBLAM distributions

RNA backbone suites

Key insight: update validation metrics using high-quality structures, and using only high-quality parts of structures

Phenix


Mean All-atom Clashscore vs Year, 1.8-2.2Å Xtal PDB depositions
12
10
8
6
4
2
0
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015
2017
2019
2021
2023
2025
MolProbity (2002)
PDB Data deposition required (2008)
MolProbity in Phenix GUI (2009)
PDB Xray Validation Task Force (2011)
Coot (2004)
wwPDB validation server (2017)
MolProbity leads a long era of structure improvement
Phenix

TOOLS & VALIDATION METRICS

Part 2

For each validation:

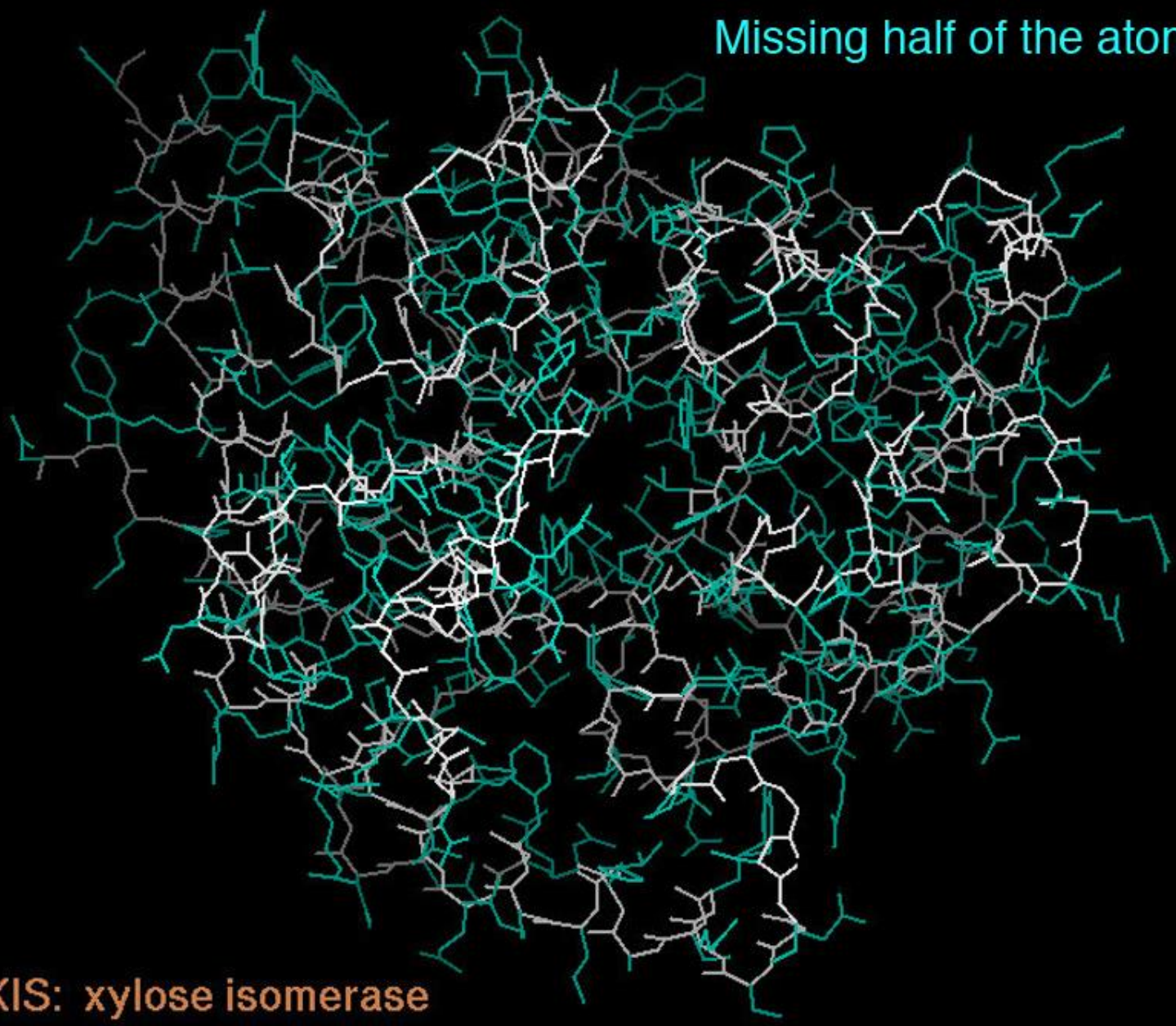
Method

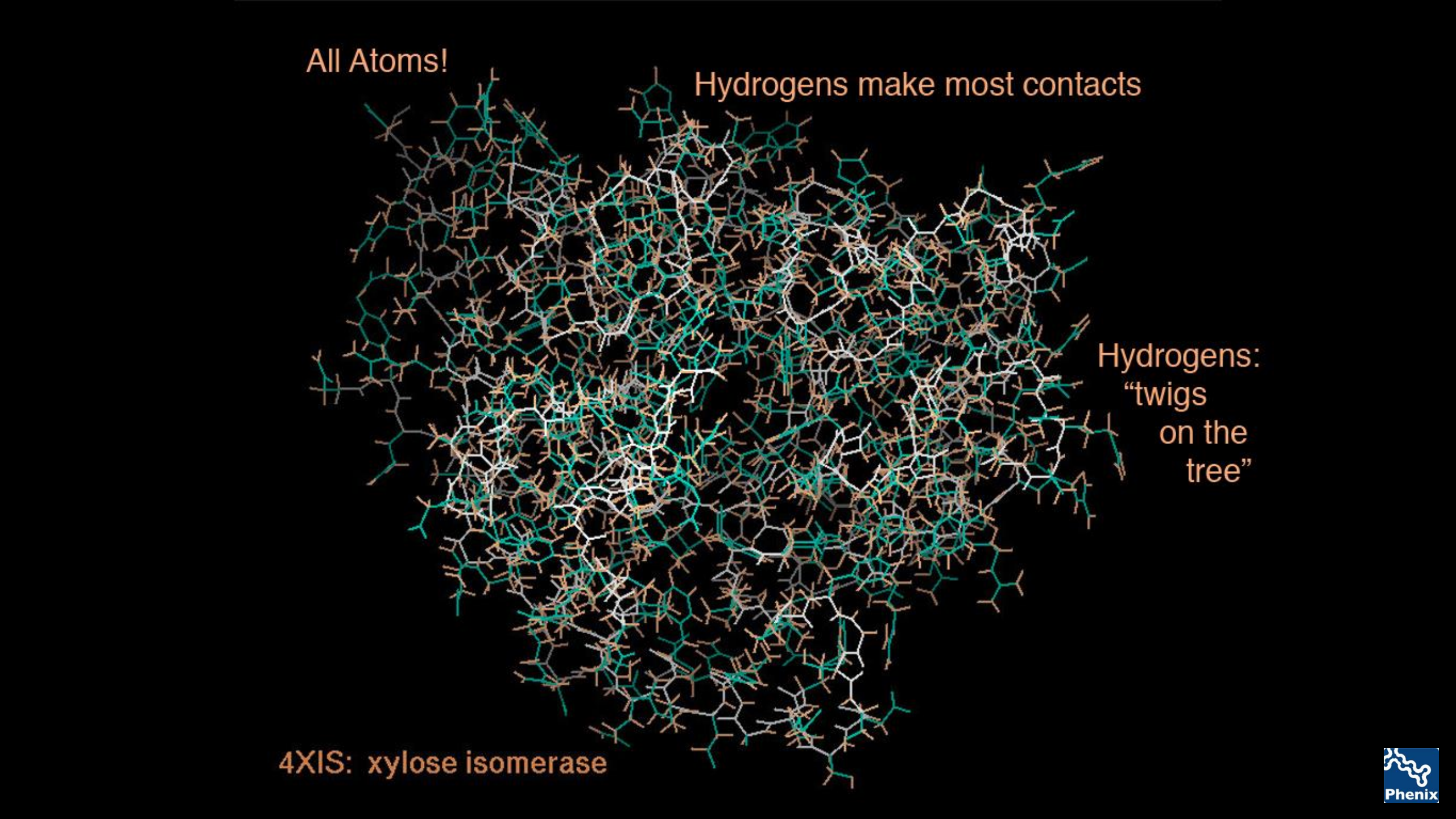
Visualization

Potential causes

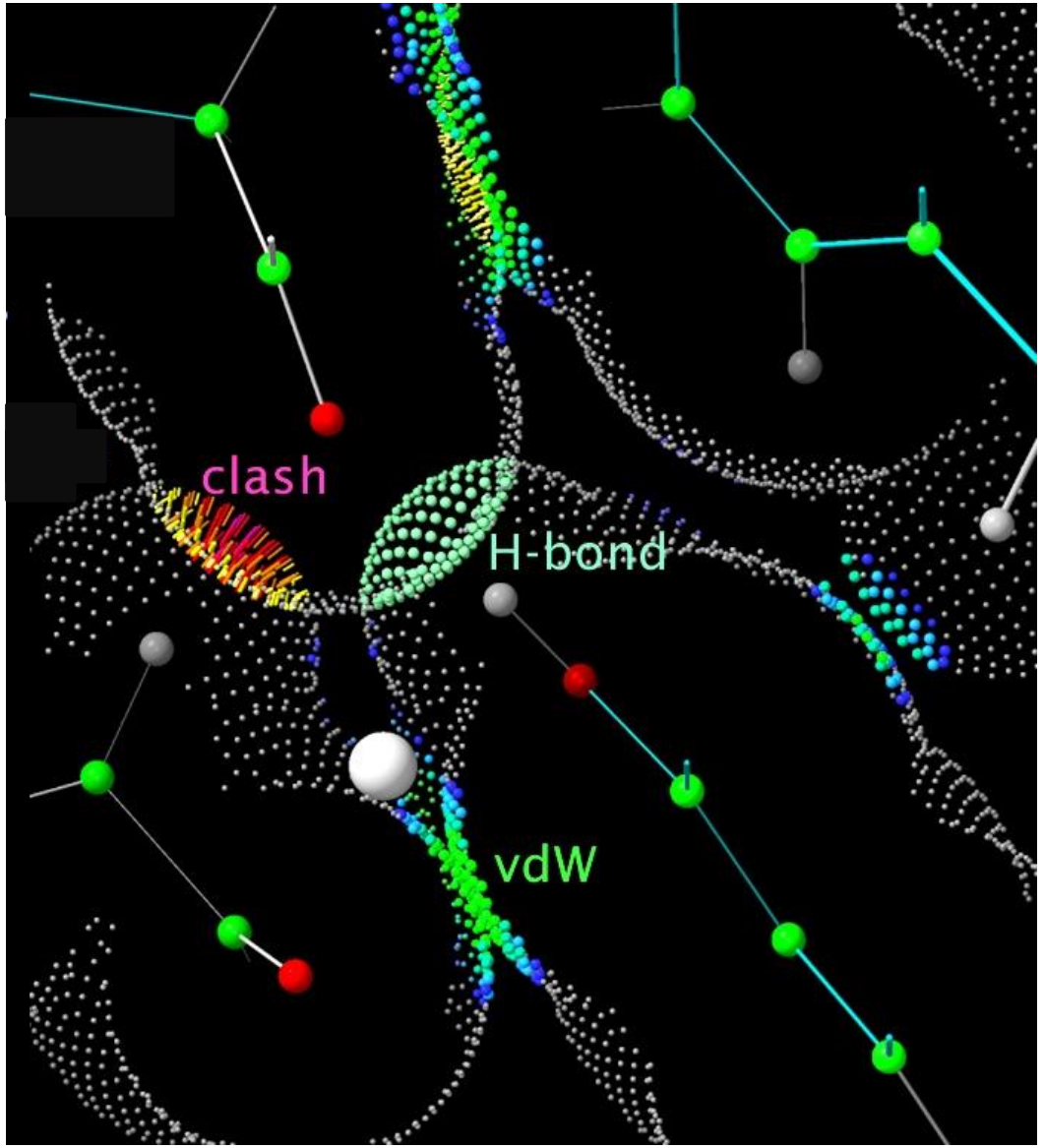
All-Atom Clashes and Contacts

Add hydrogens
(phenix.reduce or MolProbity website)


Missing half of the atoms!
4XIS: xylose isomerase


All Atoms!
Hydrogens make most contacts
Hydrogens:
"twigs
on the
tree"
4XIS: xylose isomerase

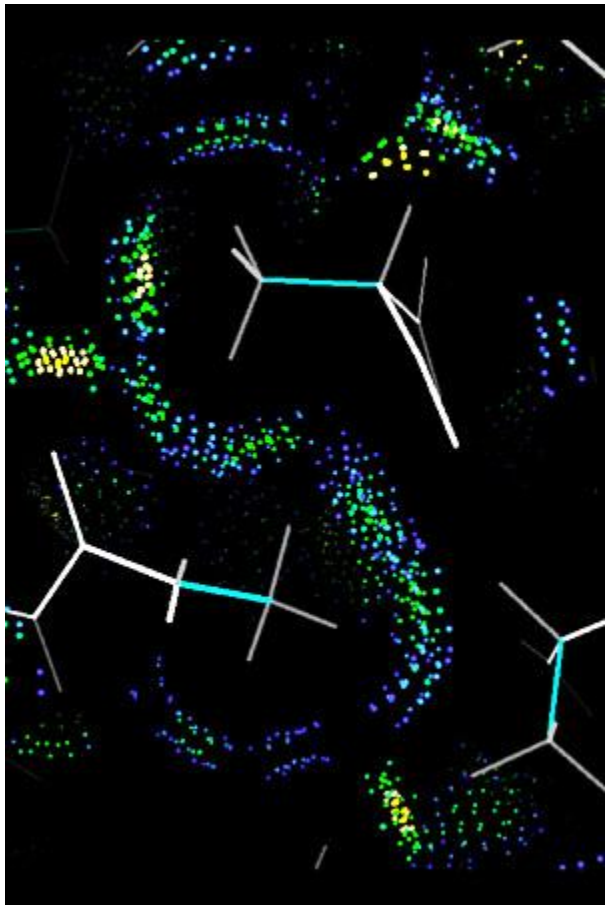
All-Atom Contacts and Clashes: Method



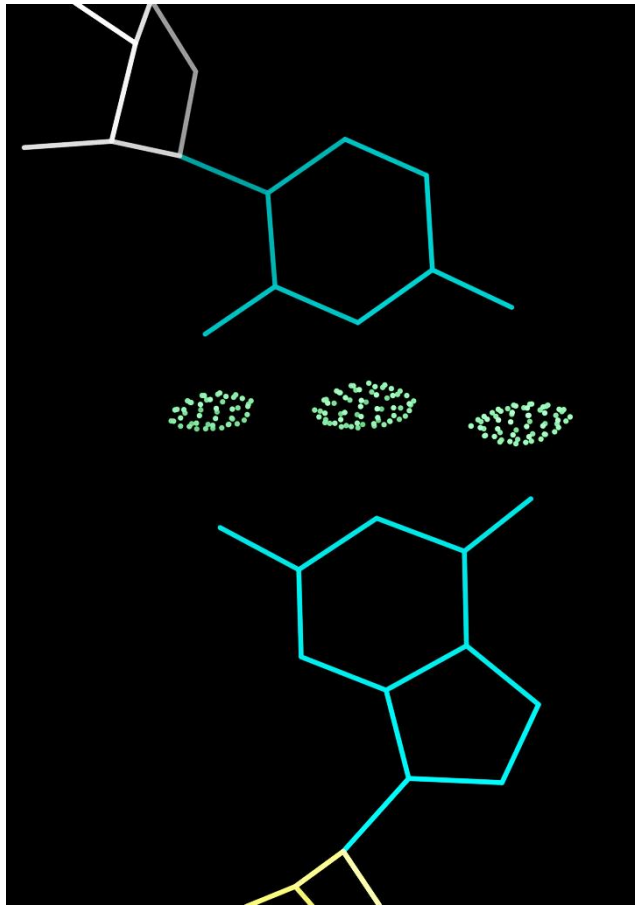
- Roll a 0.25Å radius “Probe” sphere over the van der Waals surface of each atom
- Mark where the probe touches or overlaps with another van der Waals surface
- Note that hydrogen atom surfaces can shield heavy atom surfaces

molprobity.clashscore

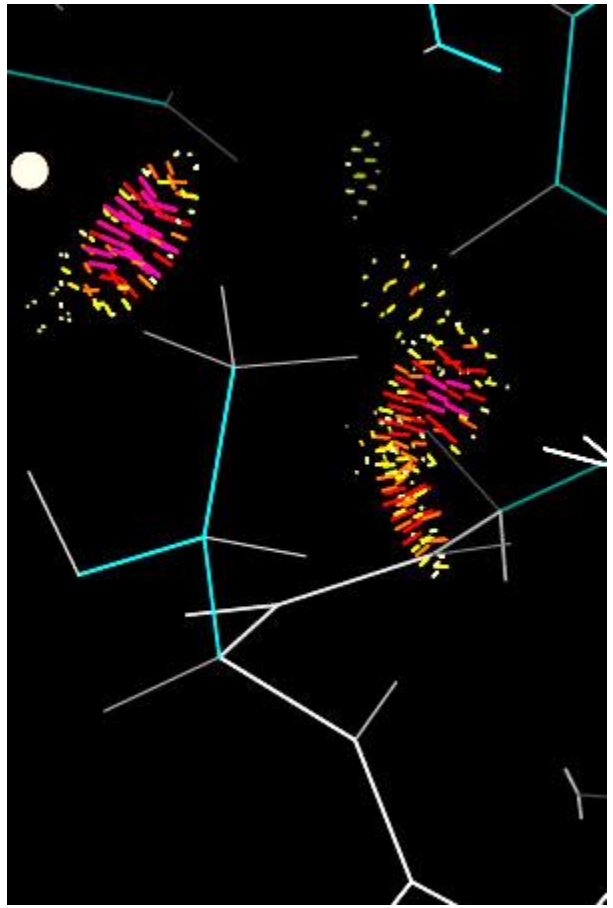
All-Atom Contacts and Clashes: Visualization



Favorable vdW packing in greens and blues

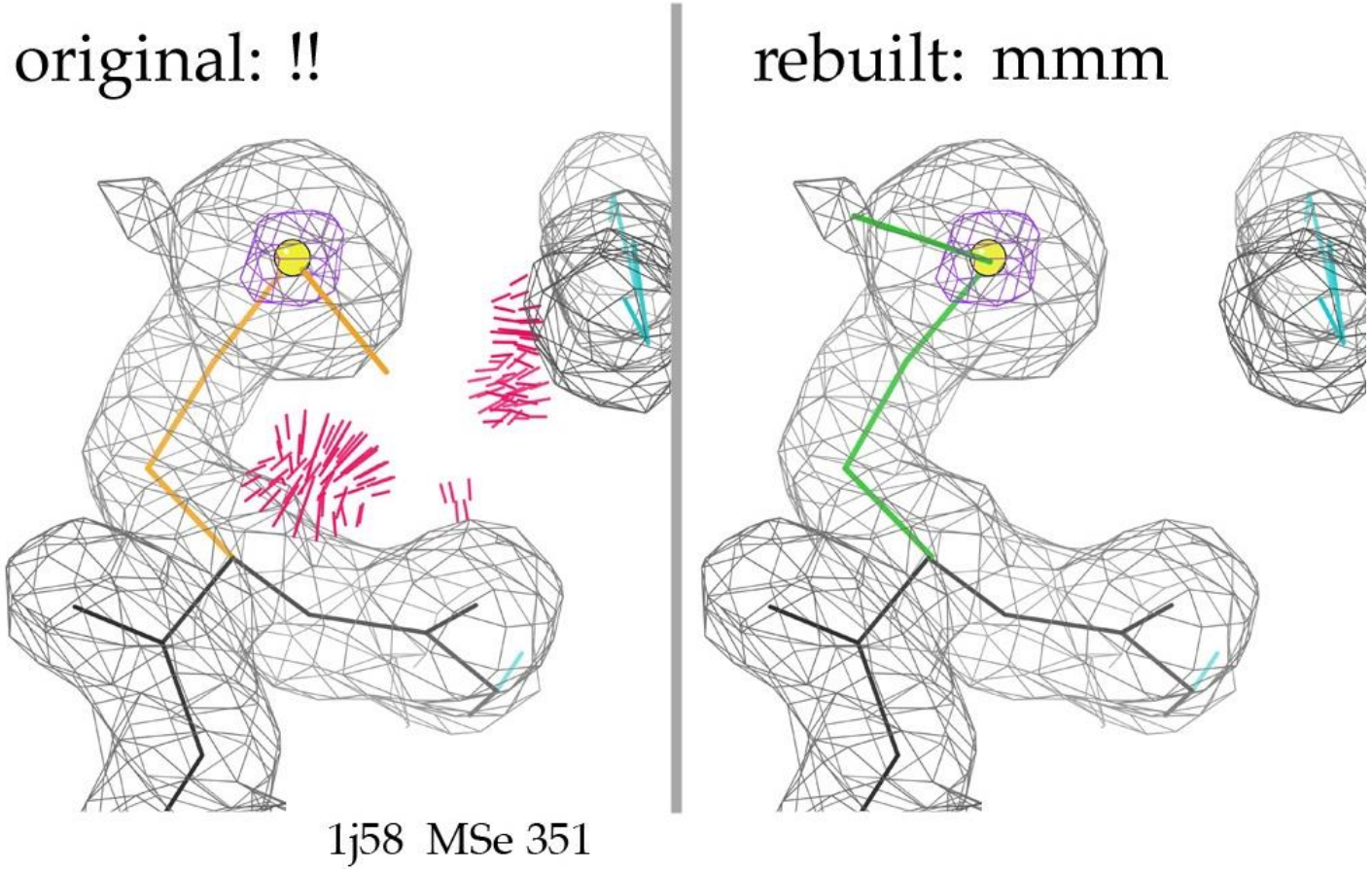


Favorable hydrogen bonding as light green pillows



Steric overlaps, aka “clashes”, as hot pink spikes

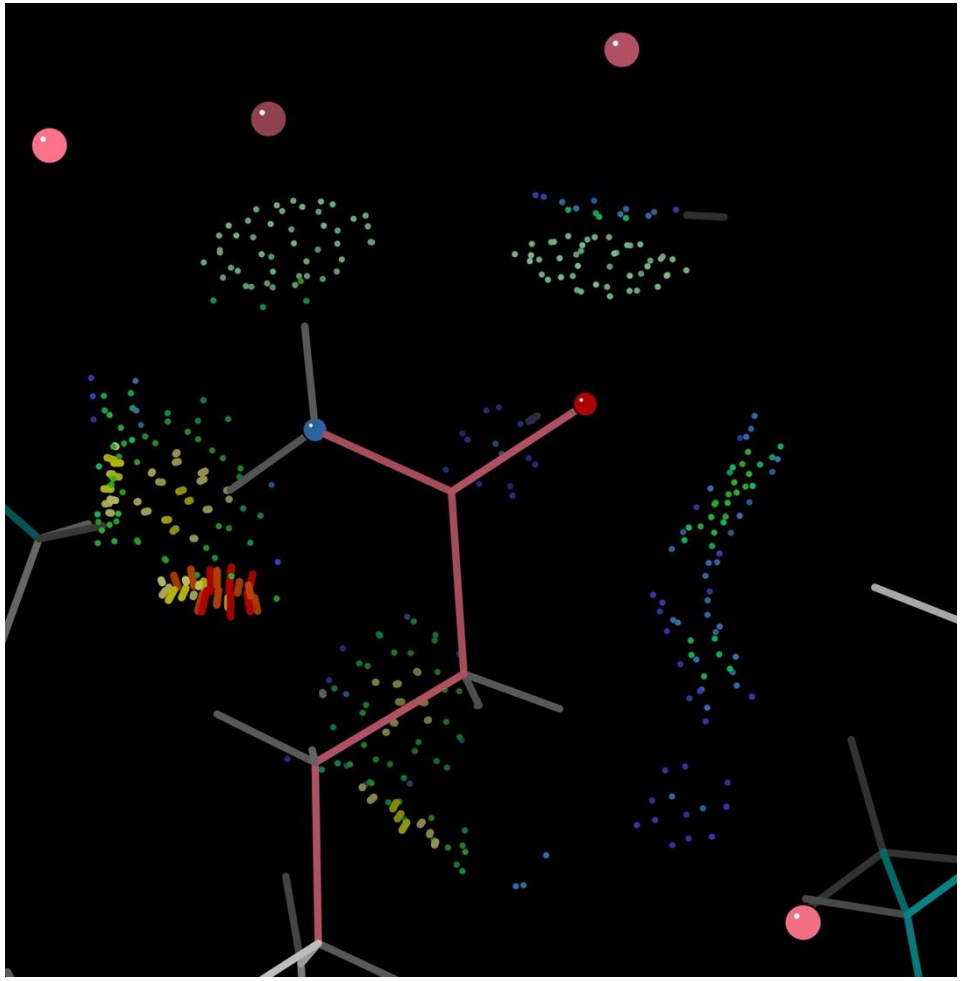
All-Atom Contacts and Clashes: Potential causes

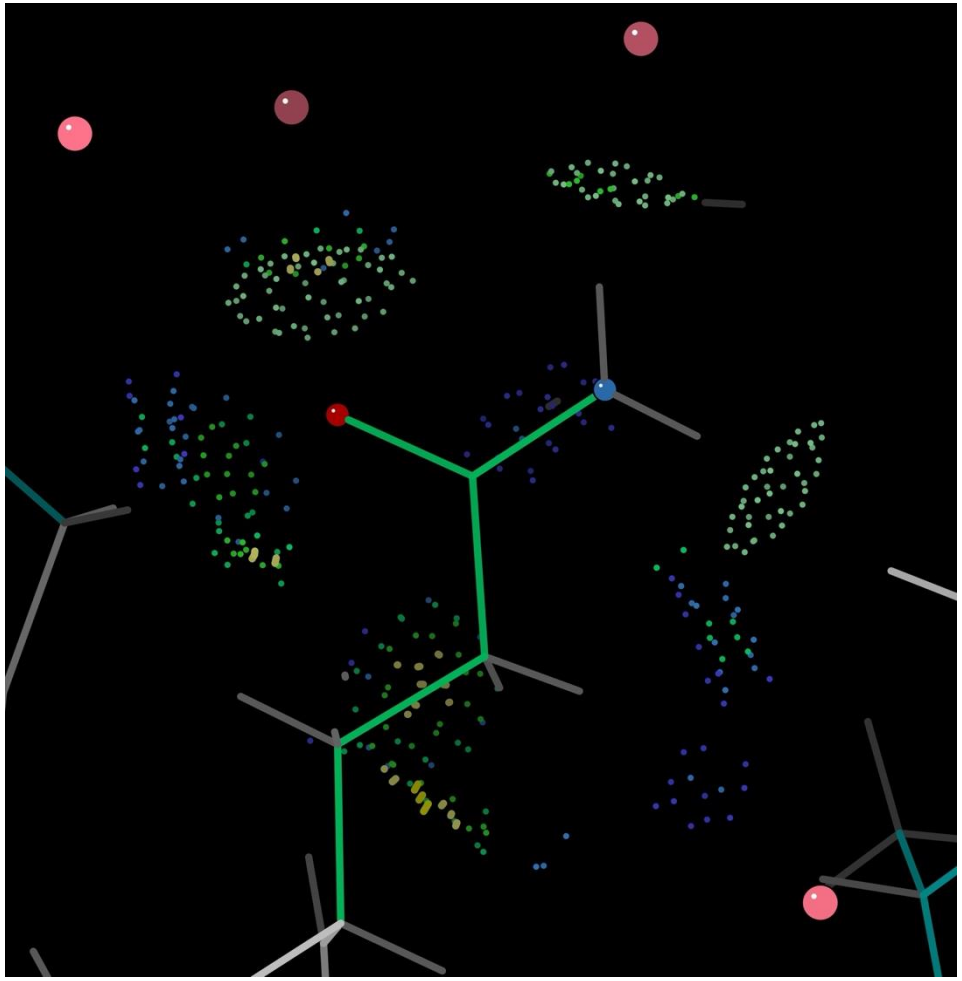


Other outliers

- Clashes usually occur alongside other outliers
- Emphasize modeling errors
  - *Real* rare features are less likely to have clashes
- Can imply direction for fixups

All-Atom Contacts and Clashes: Asn/Gln/His Flip corrections

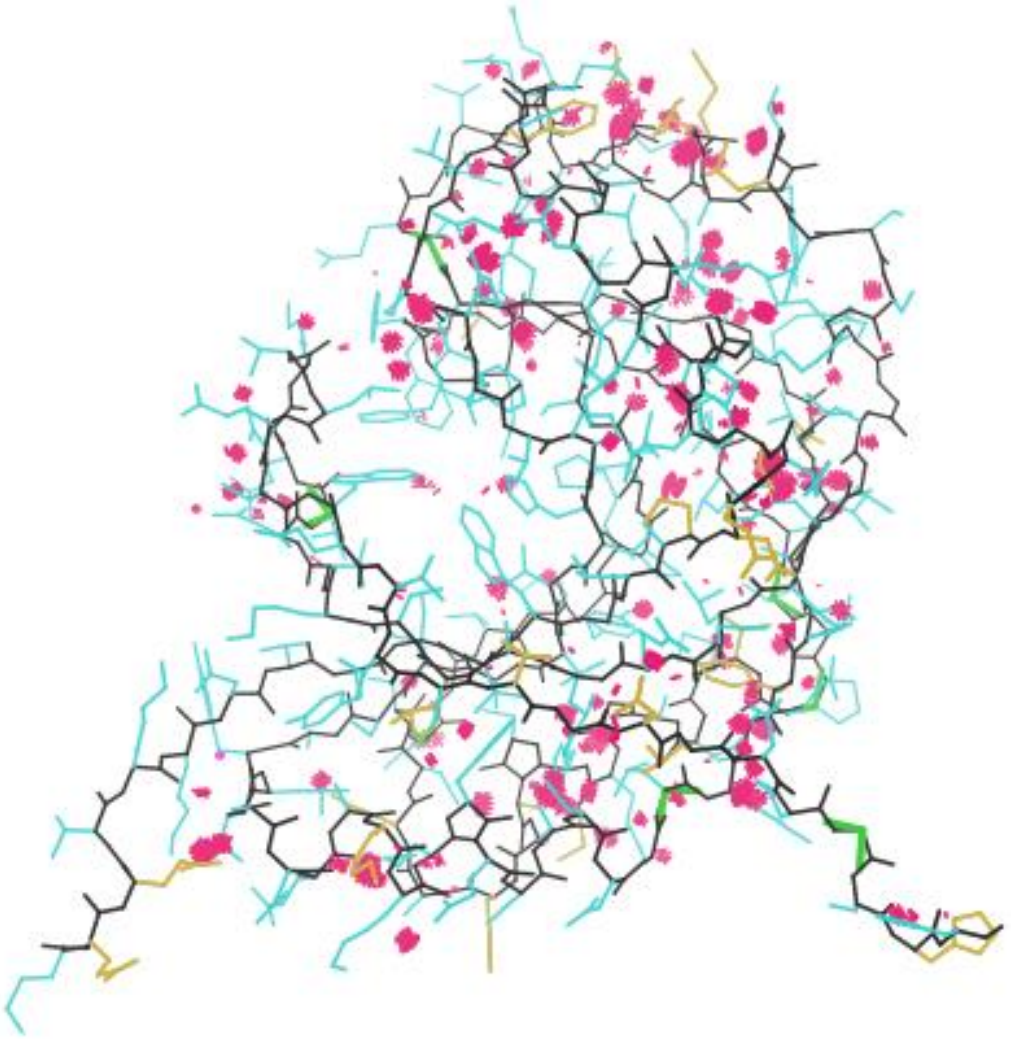




Which Gln is correct?

Clashscore (phenix.clashscore)

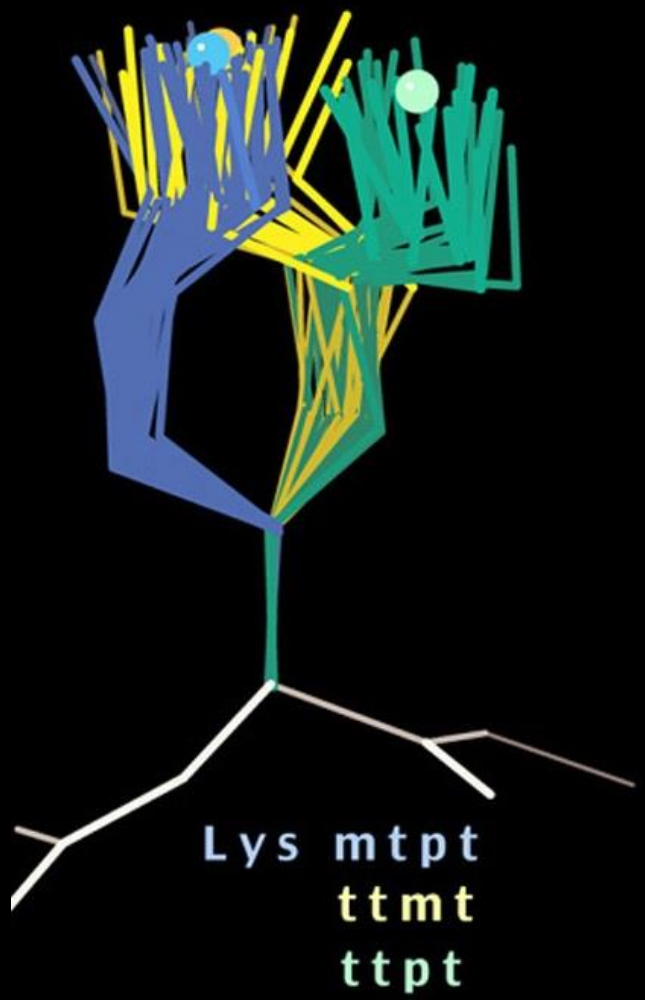
- Global score for a structure, normalized by size
- Clashes ≥0.4 Å overlap per 1000 atoms
- Useful for giving an overall sense about quality, but again, need to look at local parts of structures



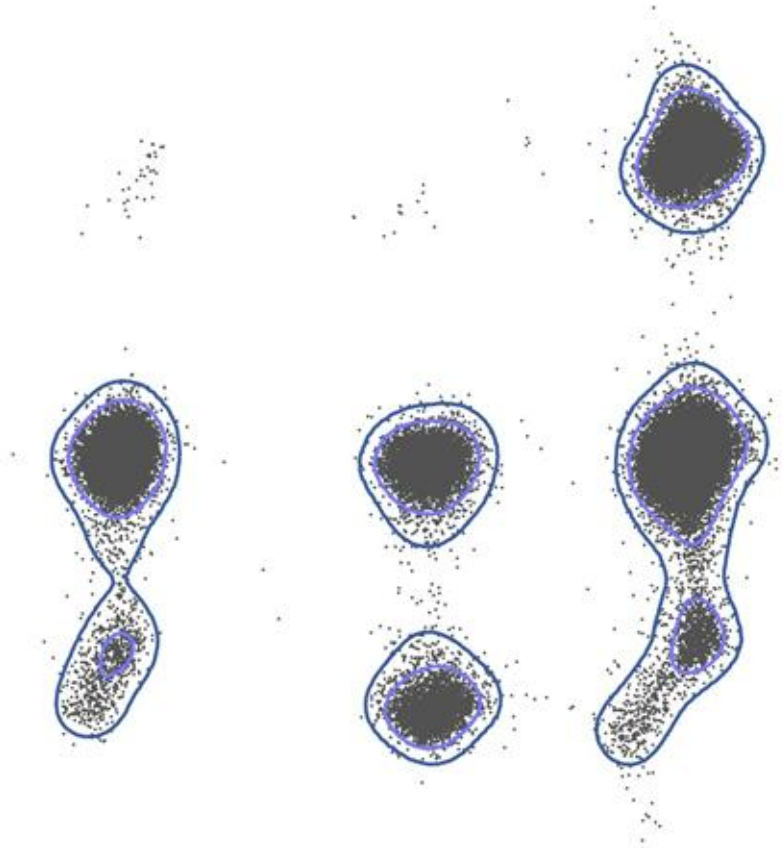
Sidechain Rotamers (molprobity.rotalyze)

Sidechain Rotamers: Method

Protein sidechains have specific conformations that are tight and distinct



Sidechain Rotamers: Method

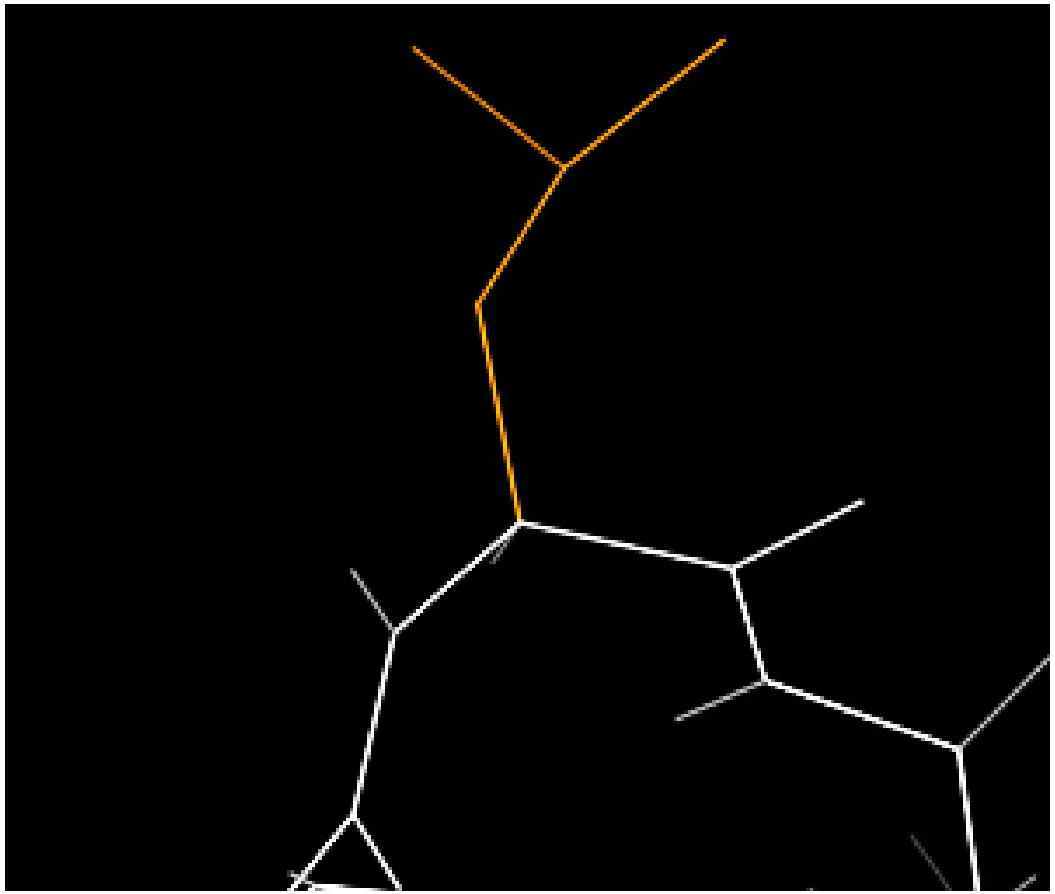


Rotamer distribution for
Isoleucine in χ1/ χ2 space

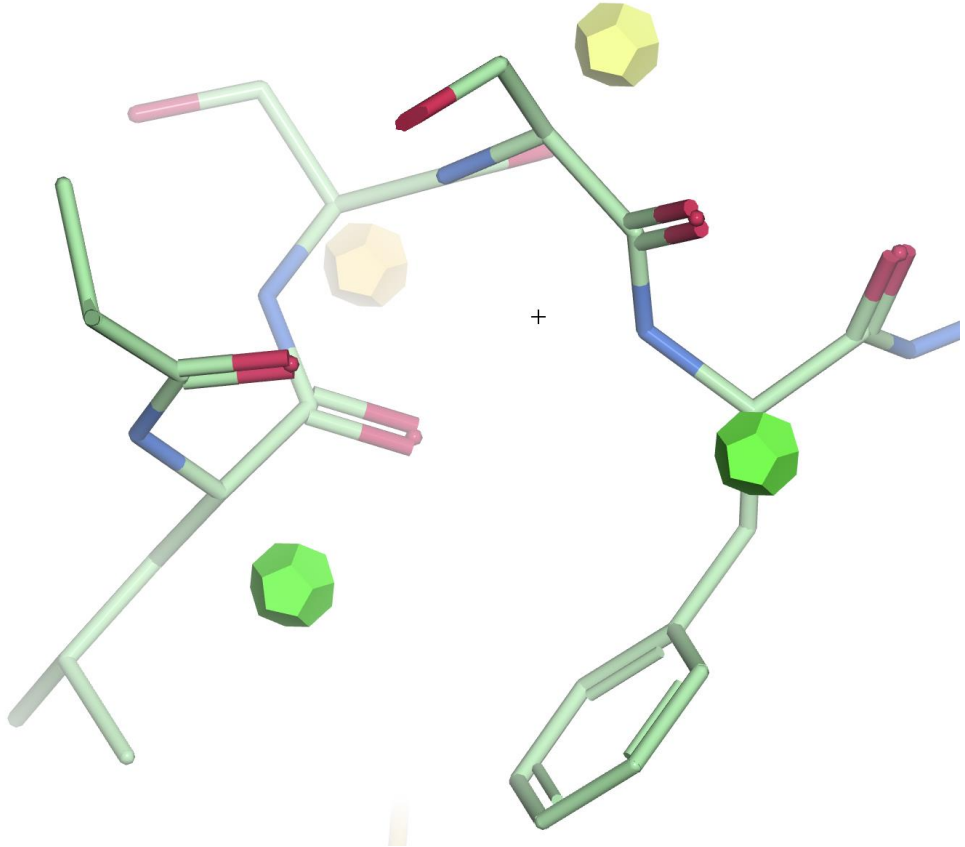
- Sidechain conformations are described by a series of χ (Chi) torsions
- Rotamers are statistically expected combinations of χ values
- For tetrahedral atom centers, this means staggered
  - p +60°
  - t 180°
  - m -60°
- For planar atom centers, rotamers are much more continuous
  - Rotamers are named with a central value
  - e.g m90 or p-80 for Histidine
- Updated in 2016:
  - Favored (98% of data) Allowed (99.7% of data)
- Outliers = sidechain in a conformation not seen in good parts of good structures

molprobity.rotalyze

Sidechain Rotamers: Visualization

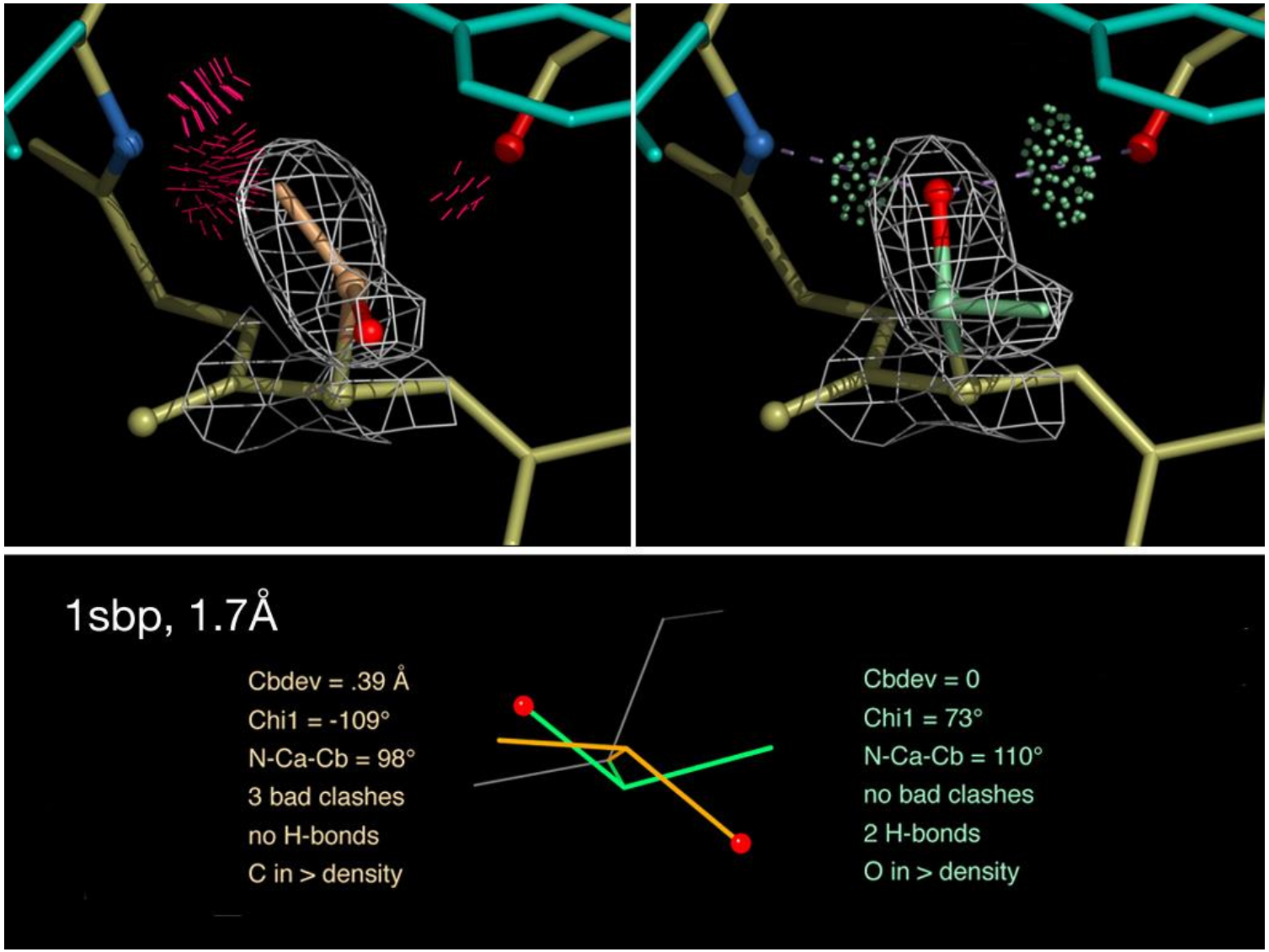


In KiNG, Rotamer outliers are traced in gold over the modeled sidechain



In Coot/Moorhen, Rotamers are marked with a colored dodecahedron

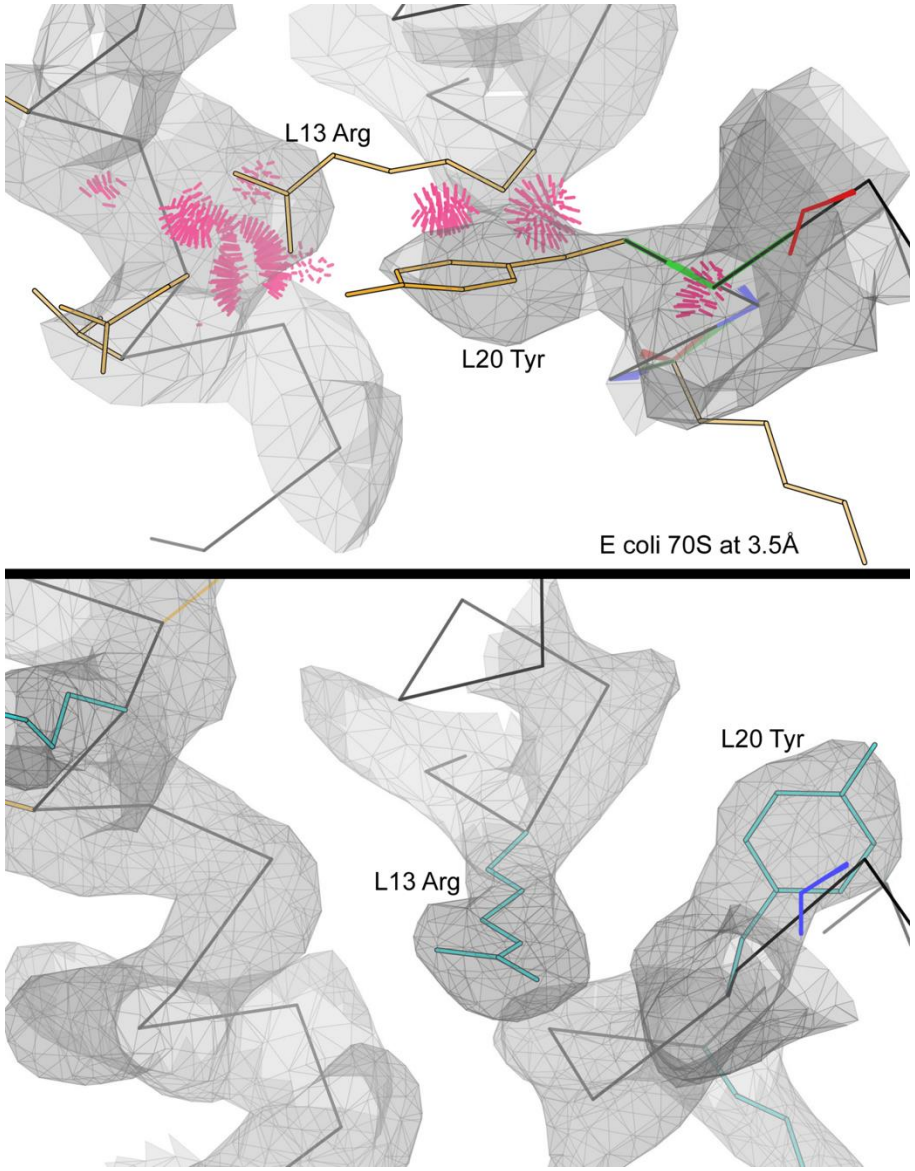
Sidechain Rotamers: possible causes



<u>Backwards Valine, Leucine, Threonine</u>

- May find terminal atoms fit into density at the expense of the branch atom
- Simple to fix with a flip (then re-refinement)

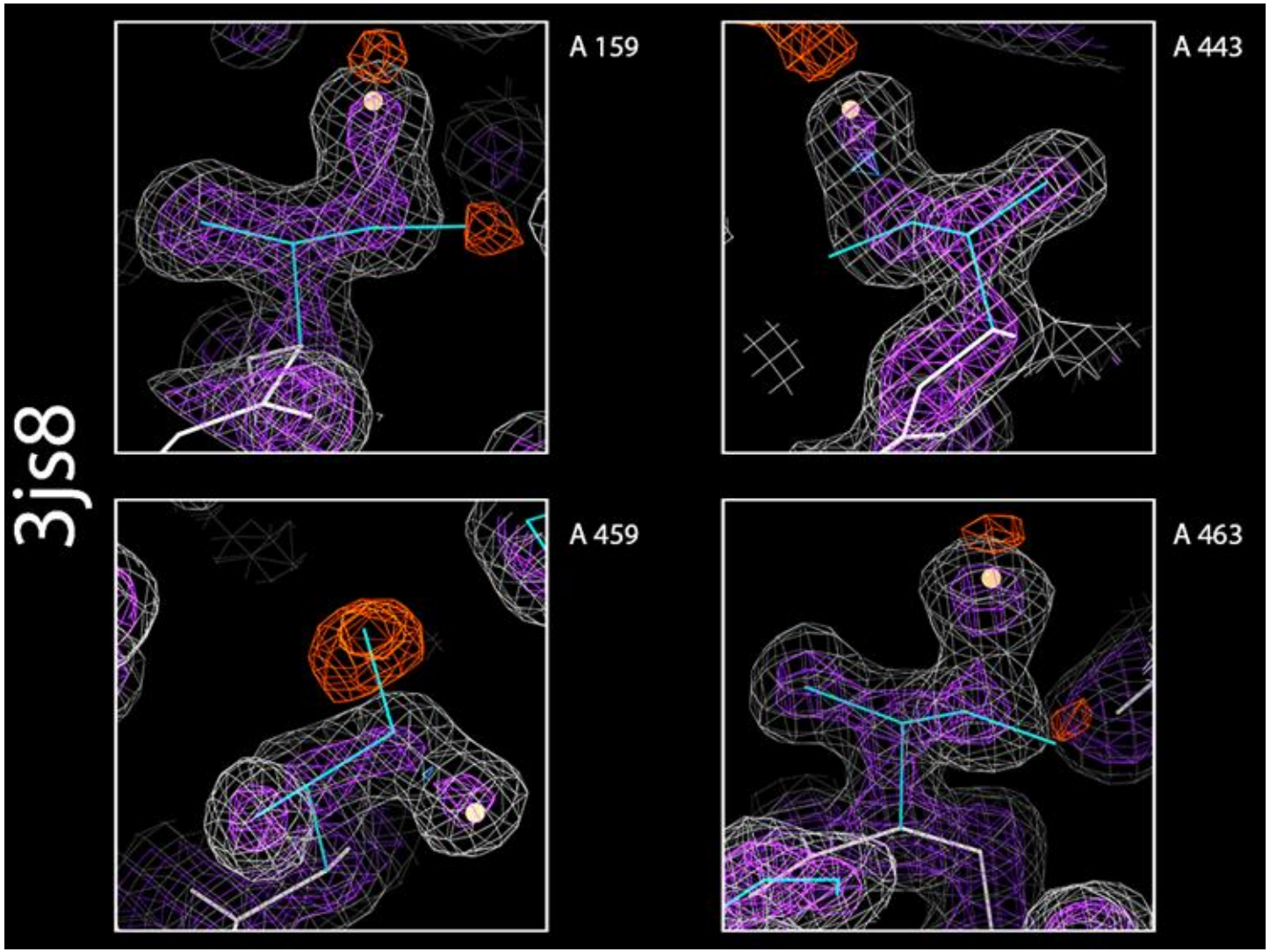
Sidechain Rotamers: possible causes



Sidechains in wrong density

- Sidechains can get stuck in the density for other features
  - Other sidechains
  - Ligands
  - Backbone in ~3Å maps
- Have to fix the whole network of misplacements
- Low resolution map may not distinguish rotamers

Sidechain Rotamers: possible causes

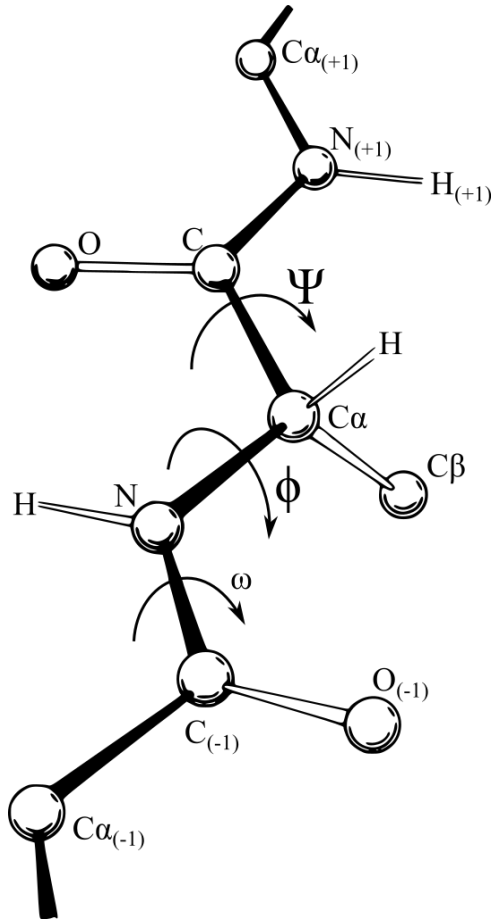


Water problems

- Modeled water may co-opt sidechain density and create a rotamer outlier
- Isoleucine CD1 is especially vulnerable
- Delete water, rebuild sidechain

Ramachandran (molprobity.ramalyze)

Ramachandran: Method

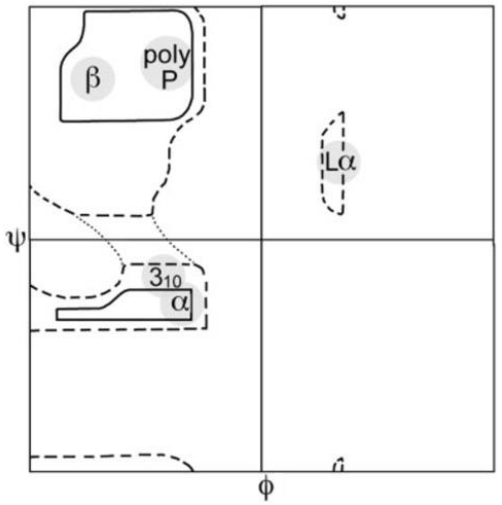


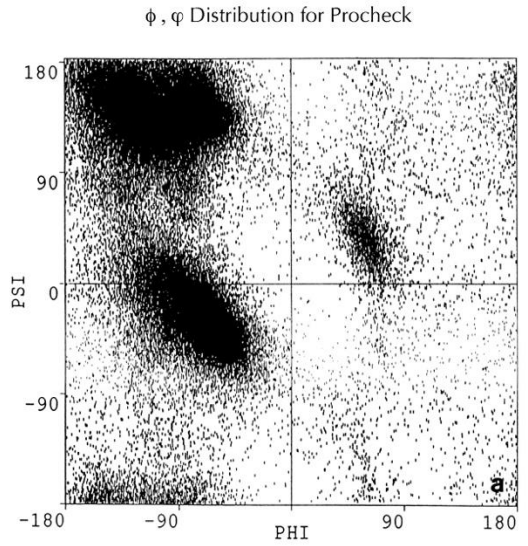
- Phi and Psi torsions describe local protein backbone conformation
- Phi **ϕ** = C_{i-1}-N-CA-C
- Psi **ψ** = N-CA-C-N_{i+1}
- Each residue's ϕ/ψ pair is converted into cartesian coordinates and checked against contours of expected behavior (based on best parts of good structures).

molprobity.ramalyze

Updating validation metrics

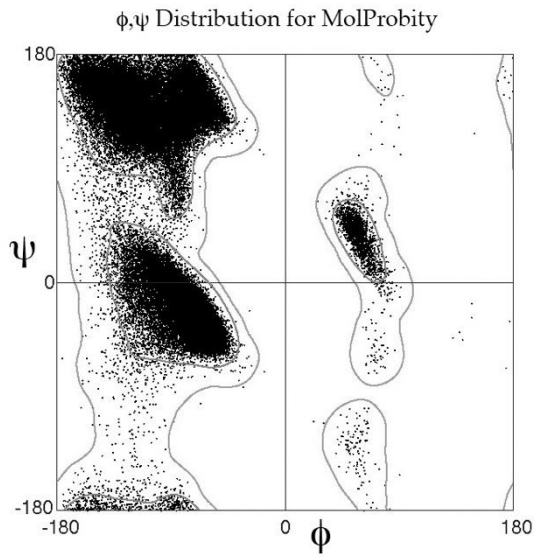
Original Ramachandran diagram from theory (1963)



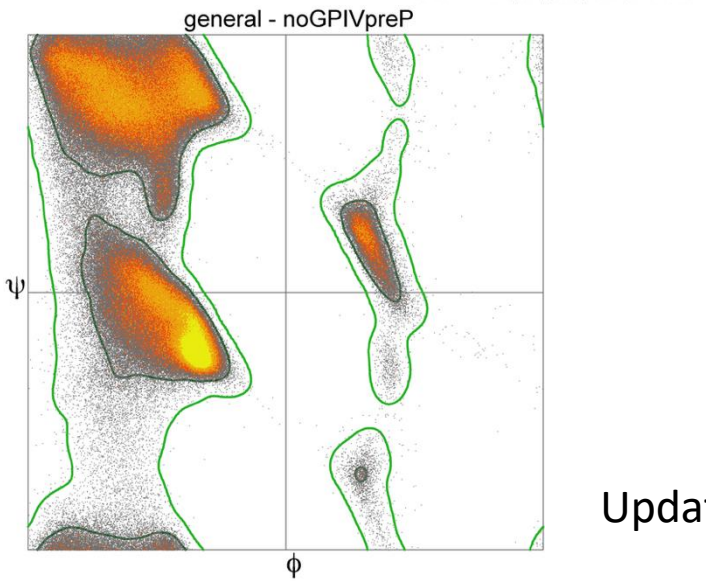


~ 100,000 residues, entire PDB as of 1992

General: All except pre-pro, pro, gly



~ 100,000 residues, Top500 structures in 2003



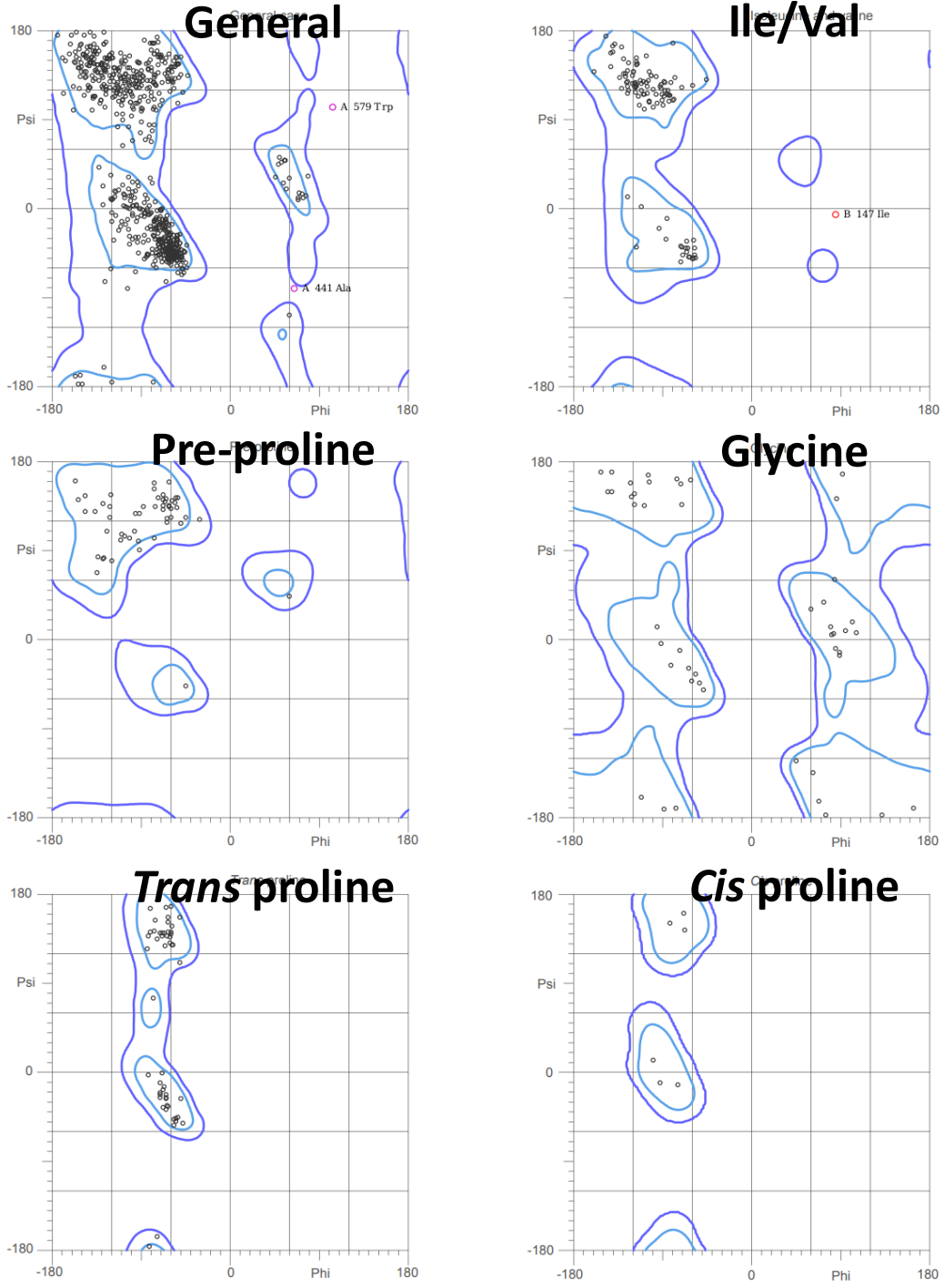
Updated 2012

Ramachandran: Visualization

Ramachandran plots shows location of each residue relative to contours of expected behavior

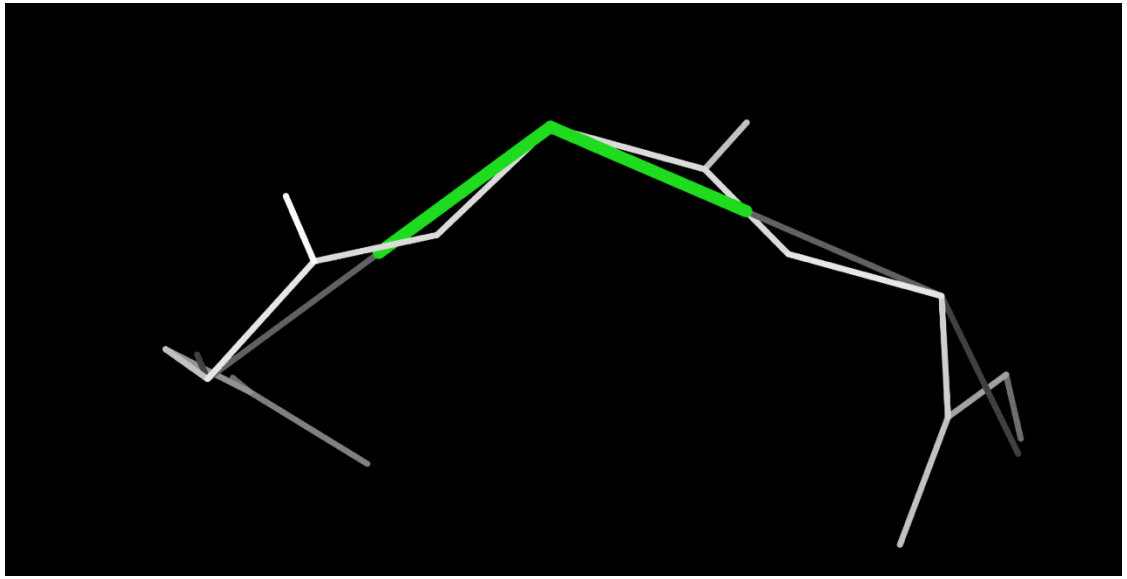
Different residue categories have very different expectations!

Glycine is permissive and symmetrical
Proline is restrictive
Branched C-Beta sidechain (Ile,Val) affect distribution

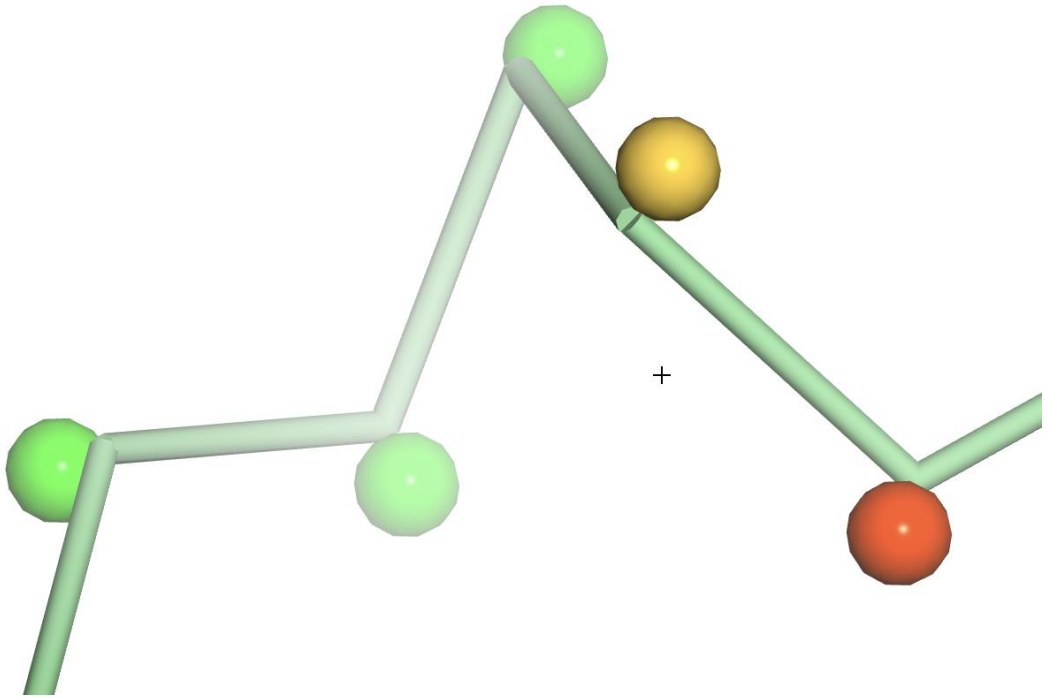


Favored (98% of data) Allowed (99.5% of data)

Ramachandran: Visualization



KiNG markup highlights an outlier residue's CA in green, and extends to the peptide bonds on either side, along the CA-CA-trace

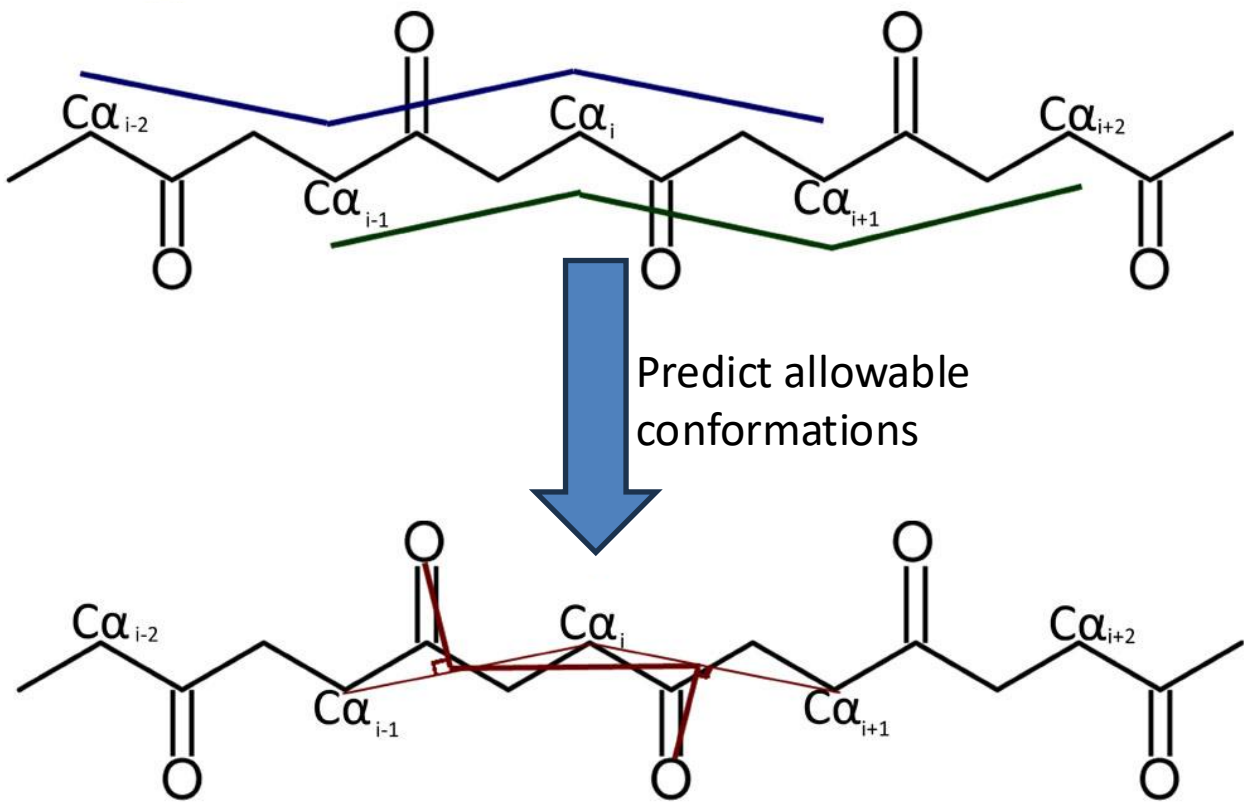


Coot/Moorhen markup places a ball at each CA, color-coded by Ramachandran favorability.

CaBLAM
(molprobity.cablam)

CaBLAM: Method

CA-pseudodihedrals capture model “intent”

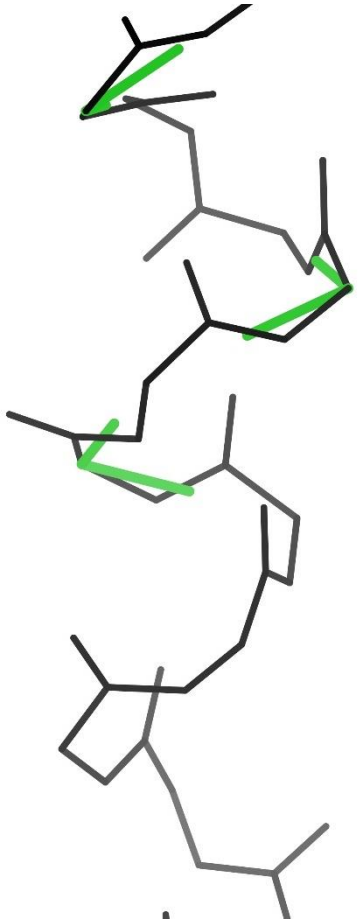


Peptide-peptide-pseudodihedral captures common model errors

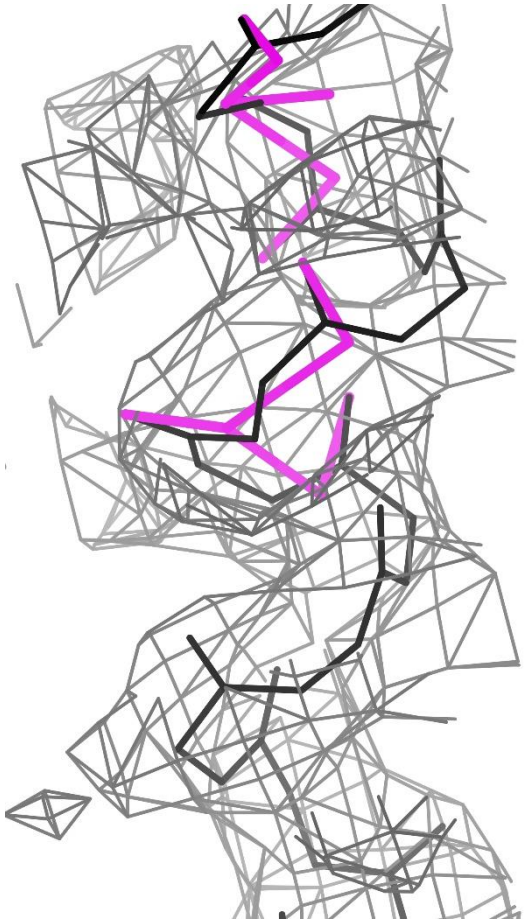
- At low resolution, the backbone CA trace is modeled better than the backbone details
- Common model errors involve wrong peptide plane orientation
- CaBLAM uses modeled CA trace geometry to predict likely peptide plane orientation, and marks the discrepancies

molprobity.cablam

Rama/CaBLAM: possible causes



Rama markup



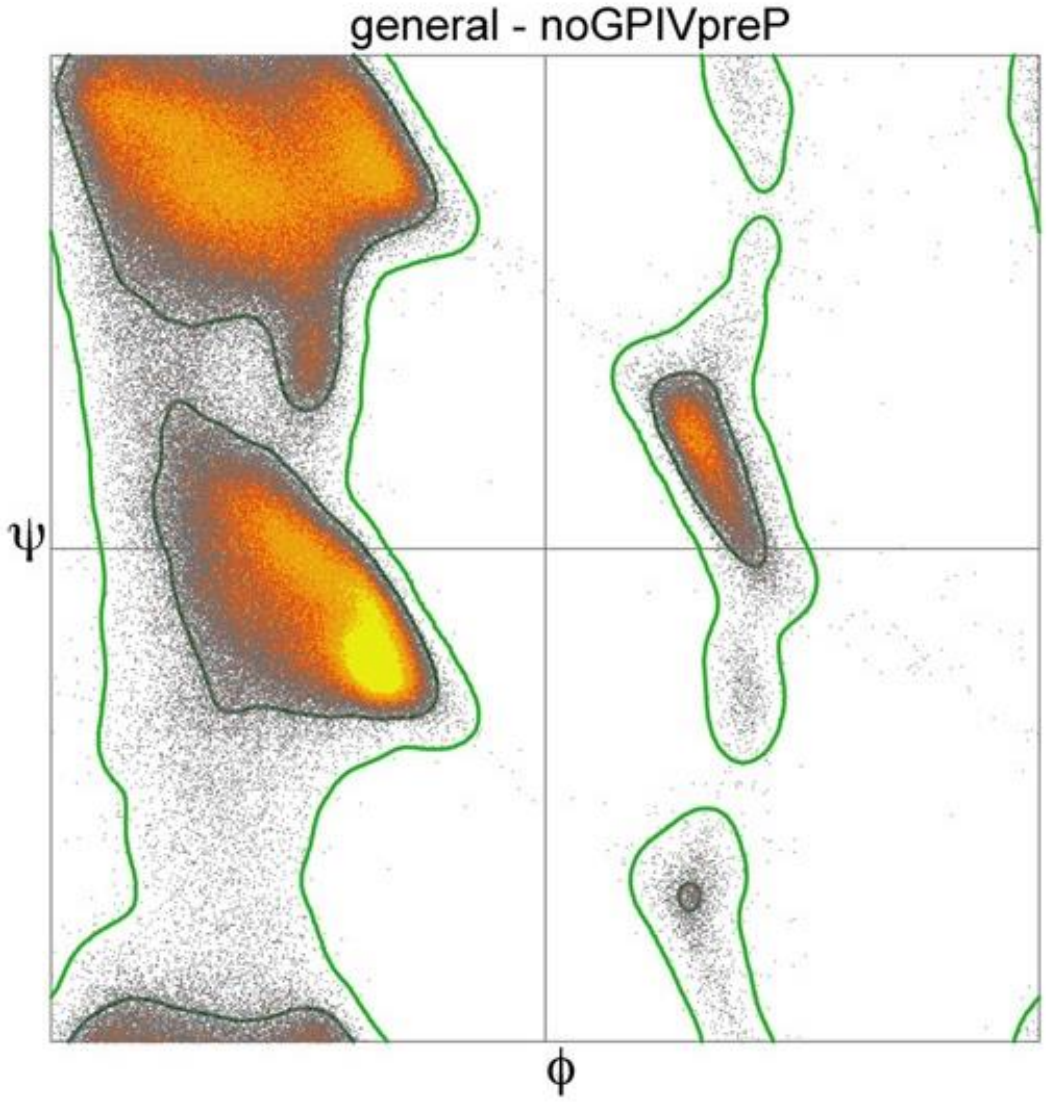
CaBLAM markup
(magenta bars)

Misplaced carbonyl oxygens

- At resolutions worse than ~2.5Å, carbonyl oxygen density disappears
  - O may be fit in arbitrary orientation
- Low-resolution density envelope allows multiple models
  - Not everything that fits is protein-like
  - Data doesn't have enough information to choose among models

Ramachandran Z-score (phenix.rama_z)

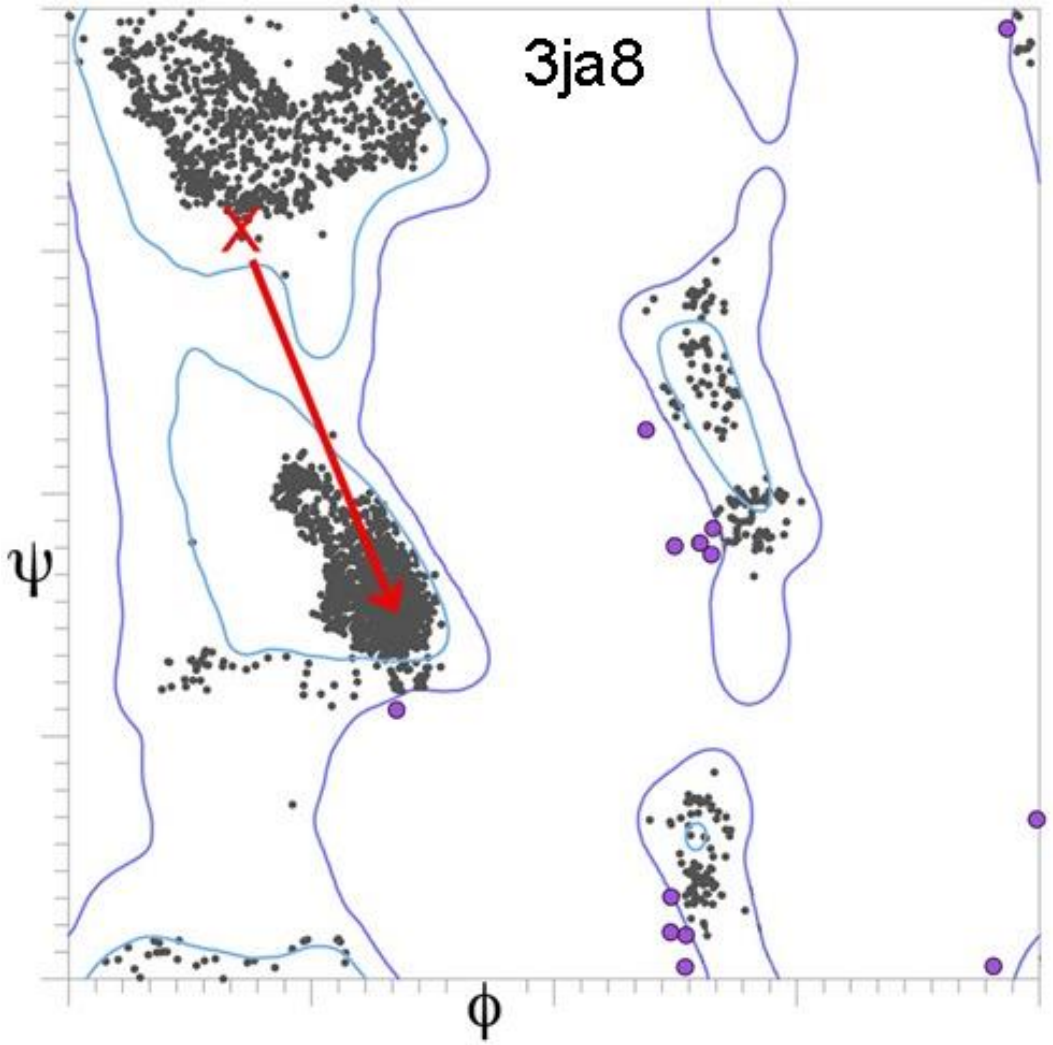
Ramachandran Z-score: Method



- Compare observed Ramachandran distribution against expected distribution (shown)
- Assign statistical Z-score based on distance from expectation
- |Z-score| <= 2 indicates a realistic distribution
- |Z-score| > 3 indicates a highly unrealistic distribution

phenix.rama_z

Ramachandran: possible causes



Rama Z-score -4.26 ± 0.10

Overfitting to Rama criteria

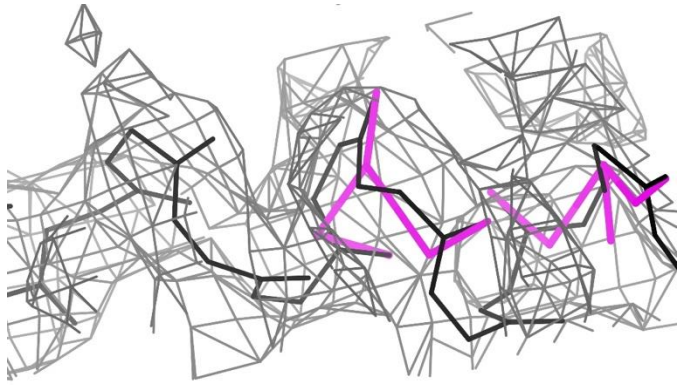
- Some programs allow refinement of the Ramachandran plot
  - Hides/compounds rather than fixes errors, if used carelessly
  - Artificially improves Ramachandran and MolProbity scores
  - Distributions ”too perfect”
- Over-idealized distribution may be detectible by Rama Z-Score
- Use other methods to fix model errors
- Then (maybe) Rama restraints to hold good structure in place

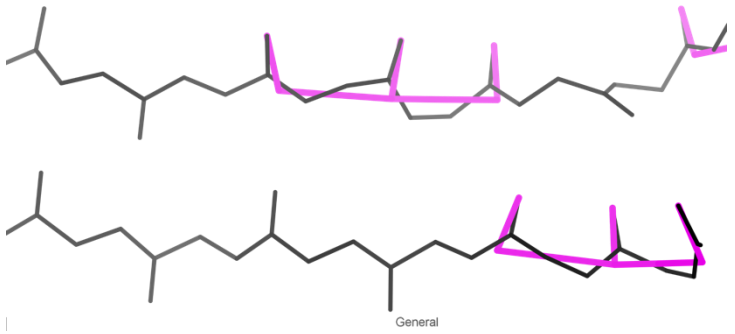
Rama/CaBLAM: possible causes

Current Rama position

does not predict

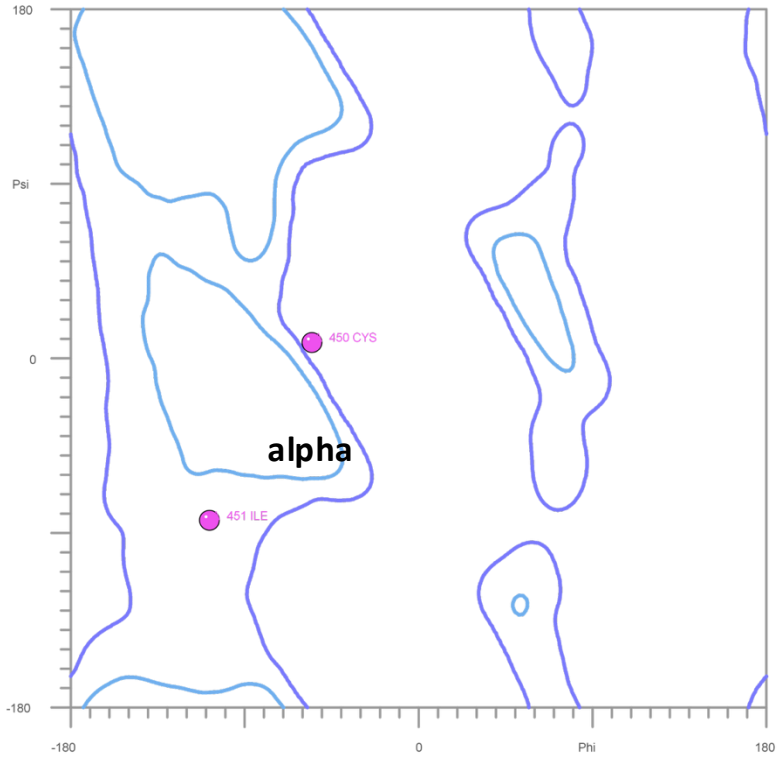
Correct Rama position

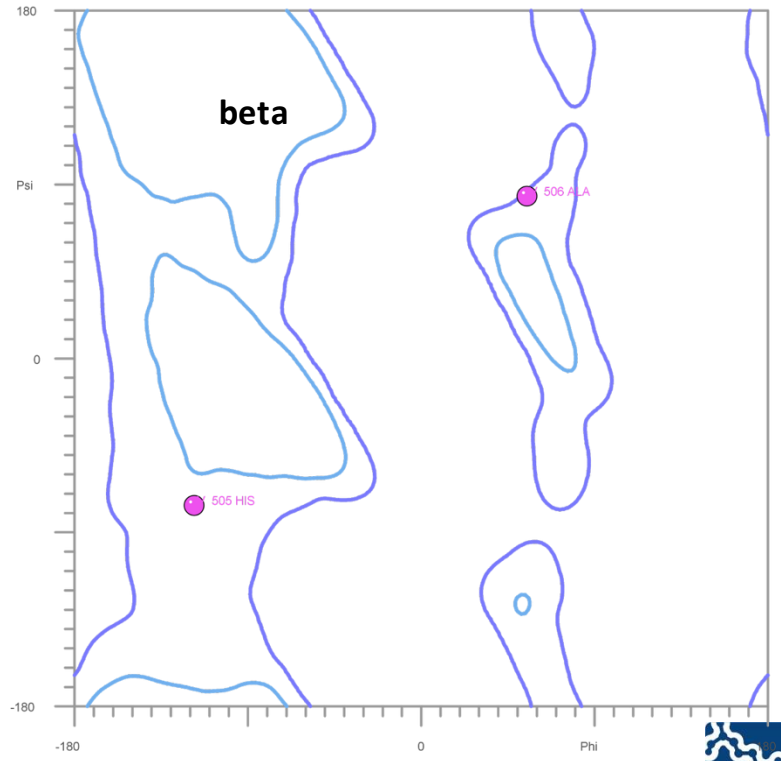




- If model errors are large, points in Rama space are displaced far from their intended regions

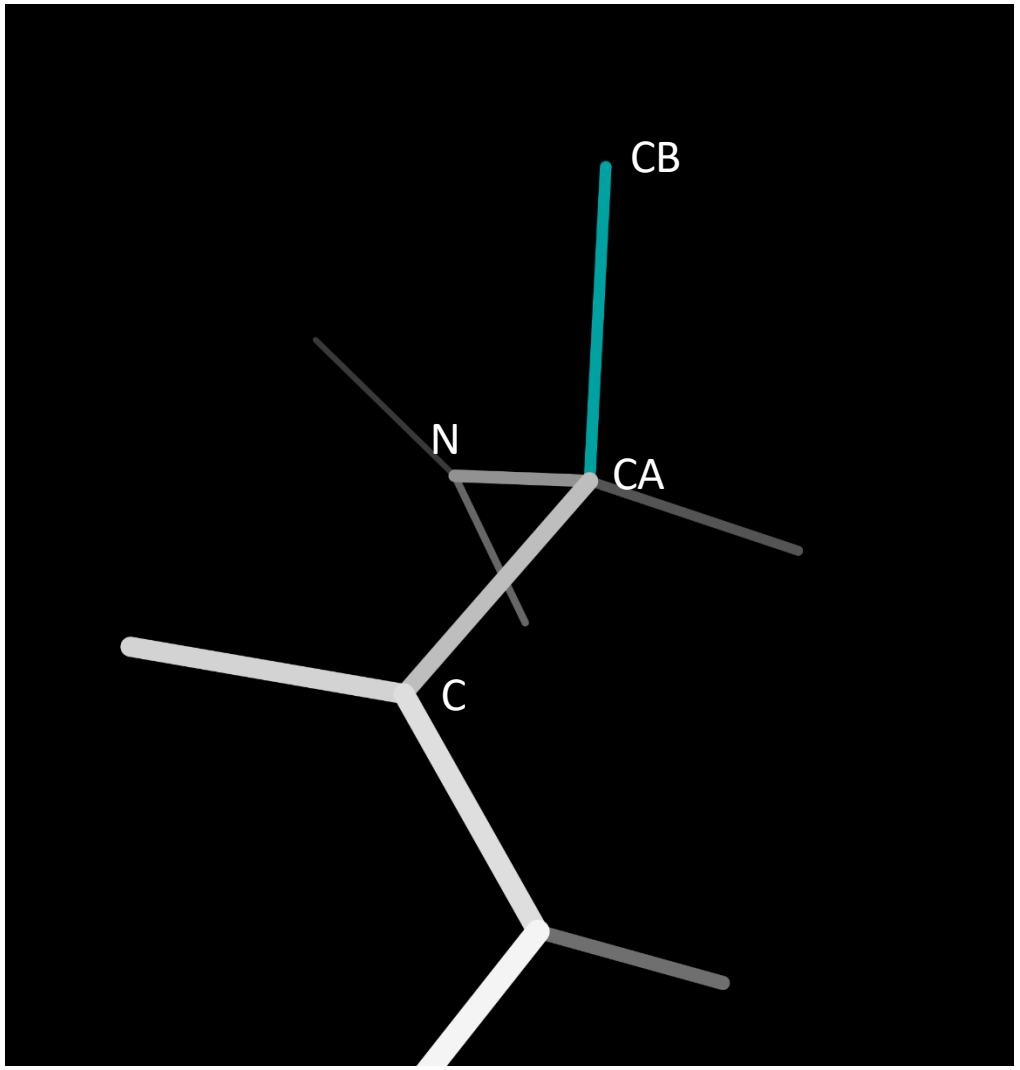
- 90° or even 180° peptide orientation errors are possible in low-resolution maps!





C-Beta Deviation (molprobity.cbetadev)

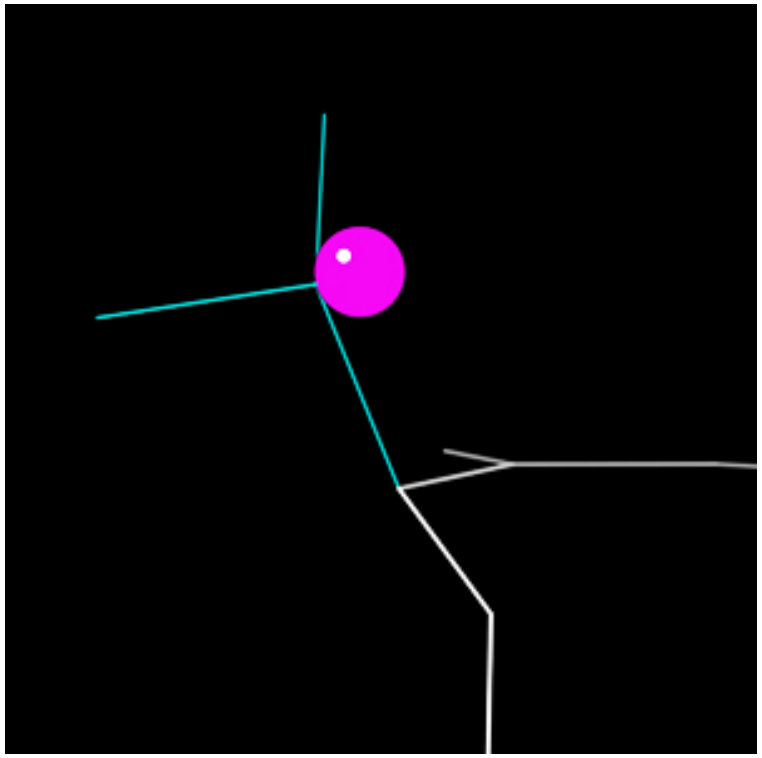
C-Beta Deviation: Method

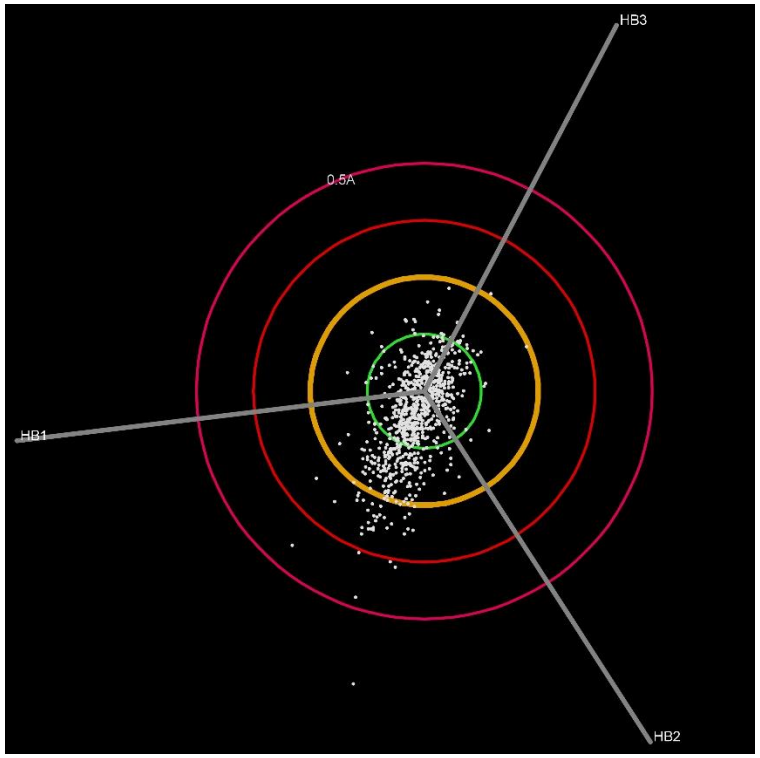


- Ideal CB position is defined by backbone geometry
- Calculate ideal position using average of two torsions
  - N-C-CA-CB
  - C-N-CA-CB
- CBs modeled >0.25Å from ideal position are outliers

molprobity.cbetadev

C-Beta Deviation: Visualization

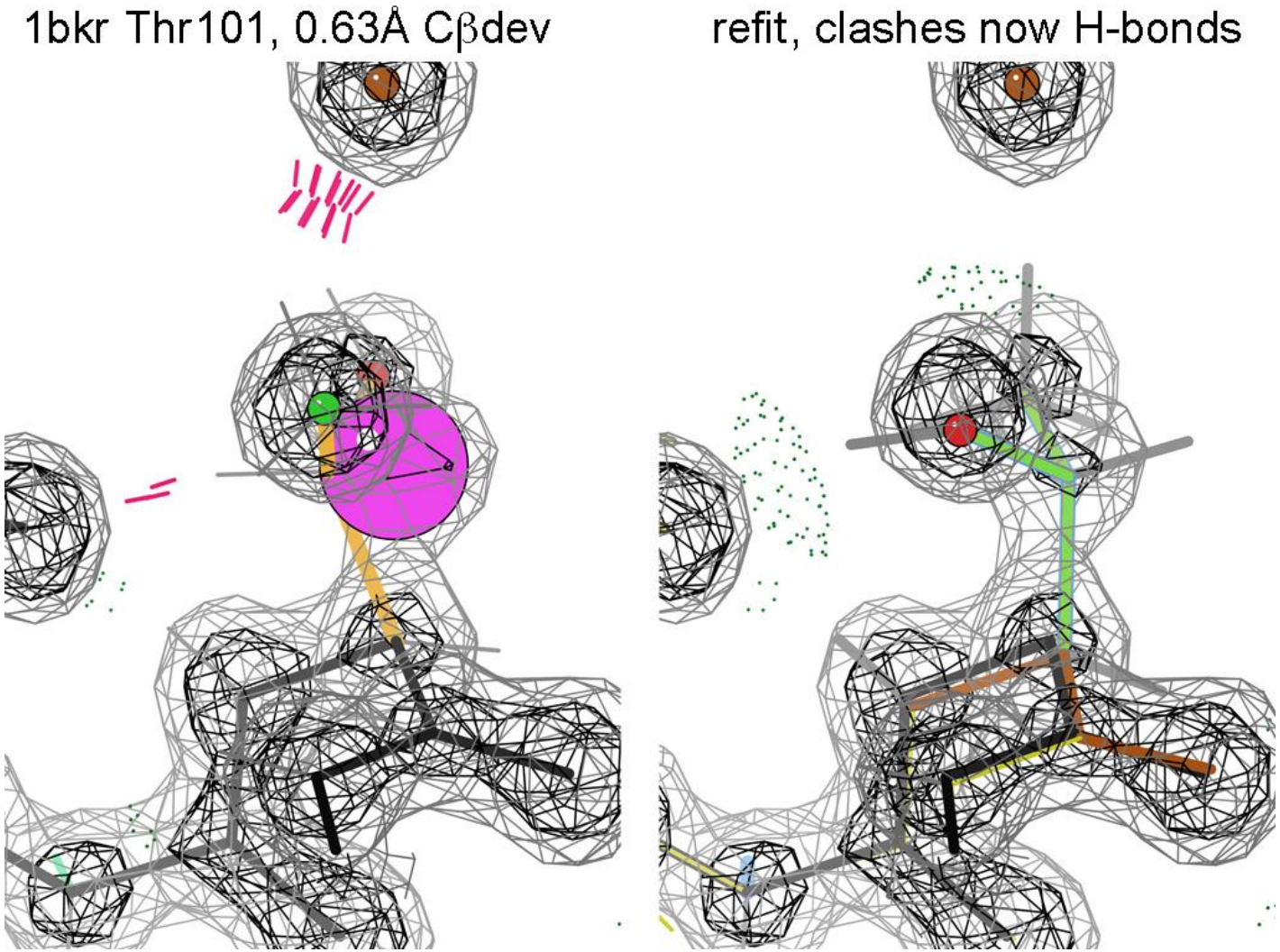




- In KiNG, a magenta sphere is drawn
  - Center at ideal CB position
  - Edge tangent to modeled position
  - Size of sphere proportional to severity of outlier

- Bullseye kinemage shows distribution and direction of all CB positions.
- Yellow circle is 0.25Å outlier cutoff

C-Beta Deviation: Possible causes



Misplaced sidechains

- CB deviation outliers are usually caused by misplaced sidechains
  - Especially branched sidechains fit backwards, like this Thr

Chirality errors

- If D amino acids are misnamed as L amino acids (e.g. ALA for DAL), or vice versa, very large Cbdevs result

Covalent Bond Geometry (molprobity.mp_validate_bonds)

Bond Geometry: Method

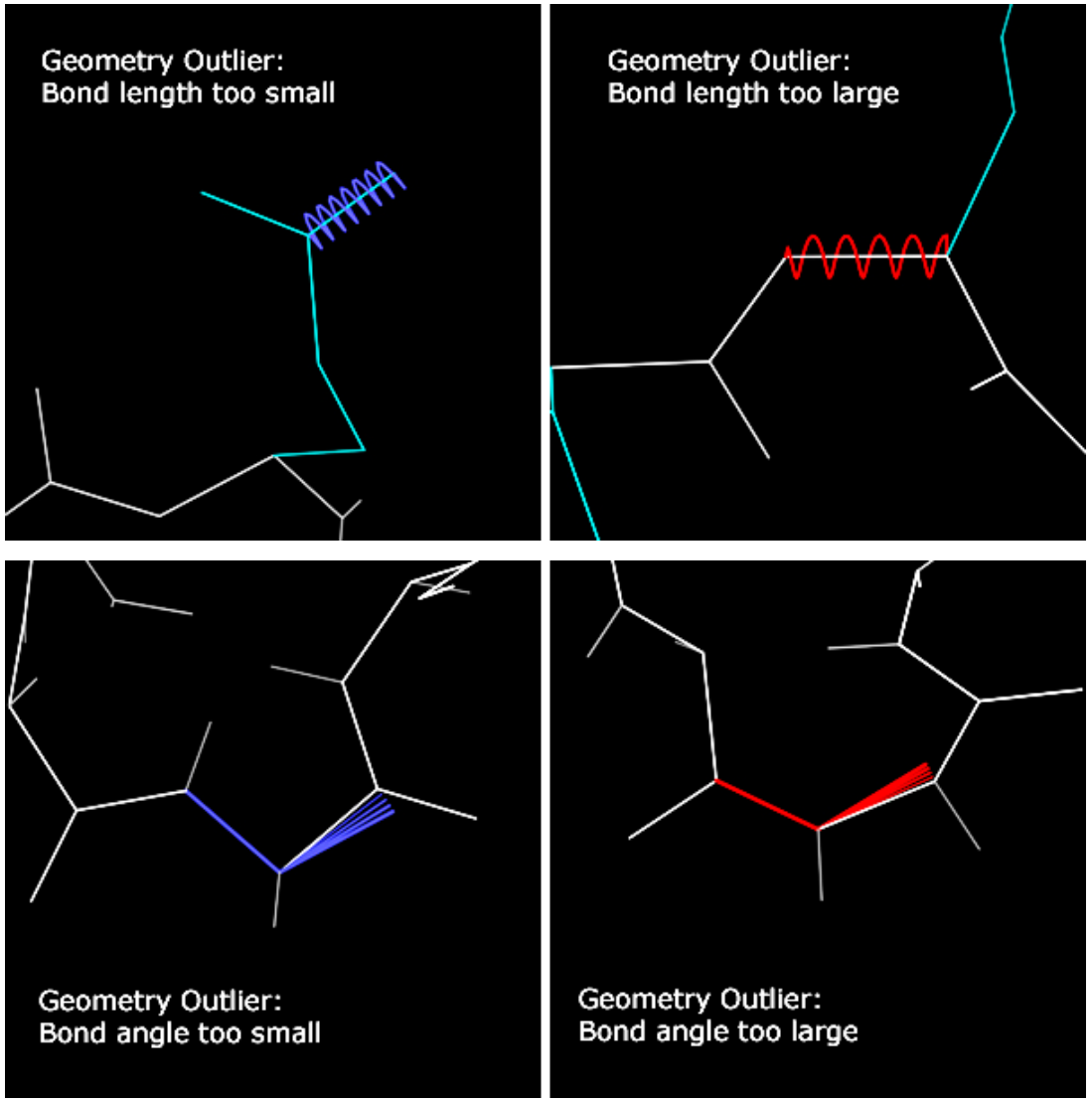
- Measure bond lengths and angles
- Check against a library of expected values
  - >4σ deviation from expected = outlier

- Standard reference library has 1 value per bond or angle
- Derived from Engh and Huber
  - https://doi.org/10.1107/S0108767391001071

- Conformation-Dependent Library (CDL) has values that depend on local Ramachandran conformation
- Phenix default
- Derived from Karplus et al.
  - https://doi.org/10.1107/S2059798315022408

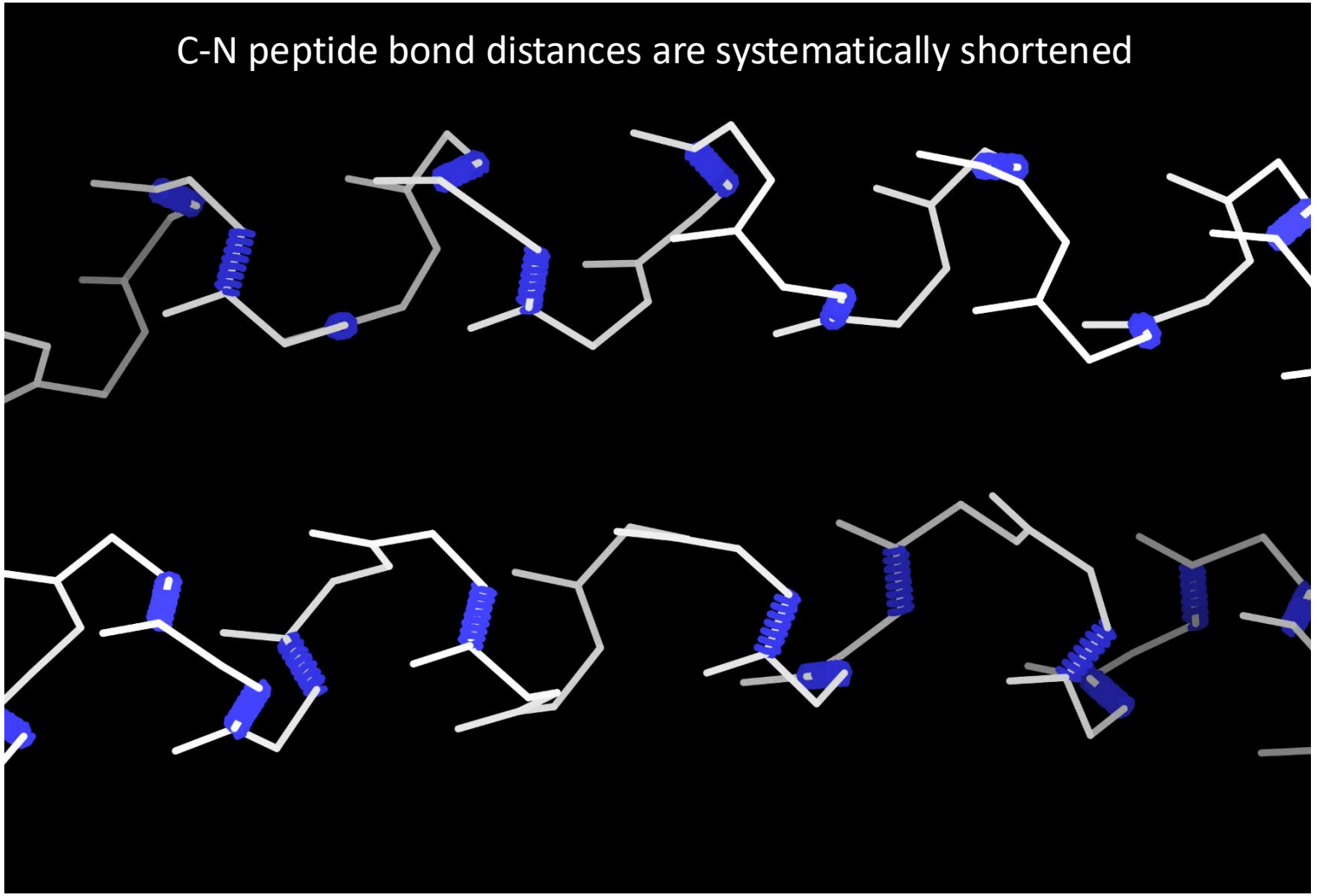
molprobity.mp_validate_bonds

Bond Geometry: Visualization



- Bond length outliers are drawn as springs
- Bond angle outliers are drawn as fans
- Color-coded
  - Red-shift = too far
  - Blue-shift = too close

Bond Geometry: possible causes

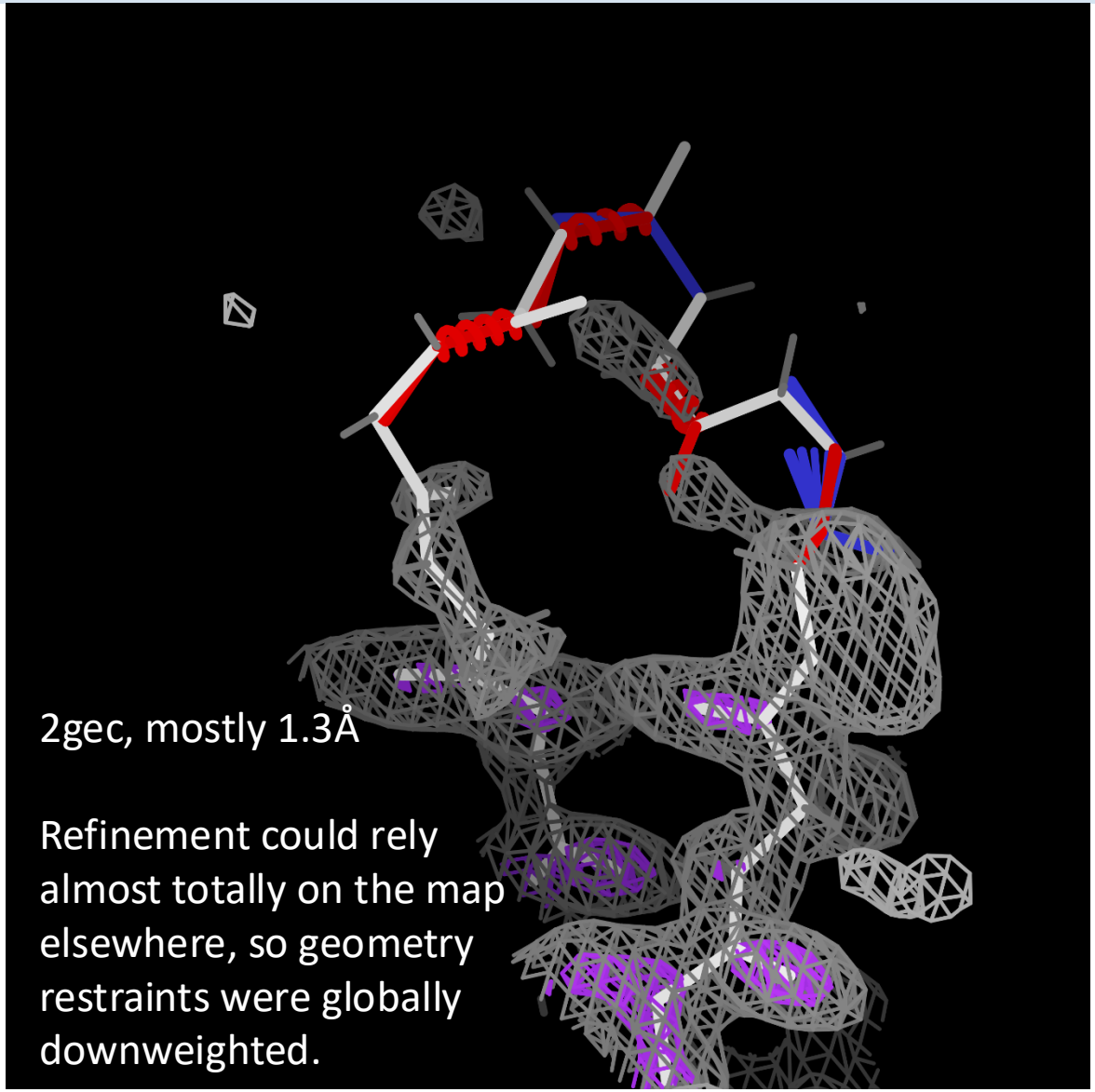


OmegaFold prediction for p81313, as of Sept 2022

Systematic

- Systematic geometry errors occur in programs with different libraries or expectations
- Be aware of what you import
- Do geometry minimization and/or re-refine.

Bond Geometry: possible causes

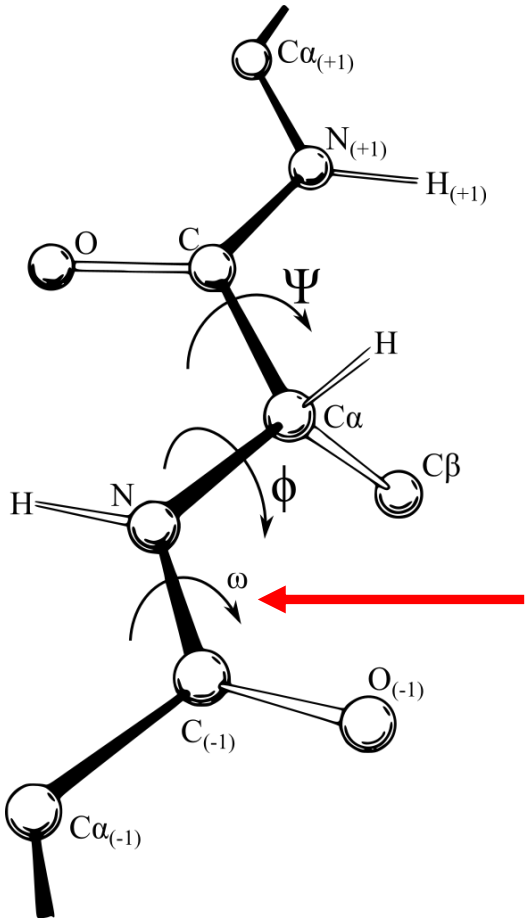


Localized

- Localized geometry outliers result from conformational strain and/or missing density
- Fix the source of strain
- Manually apply more restraints to low-data regions
- Leave it unmodeled if a good solution is impossible

Cis Peptides (molprobity.omegalyze)

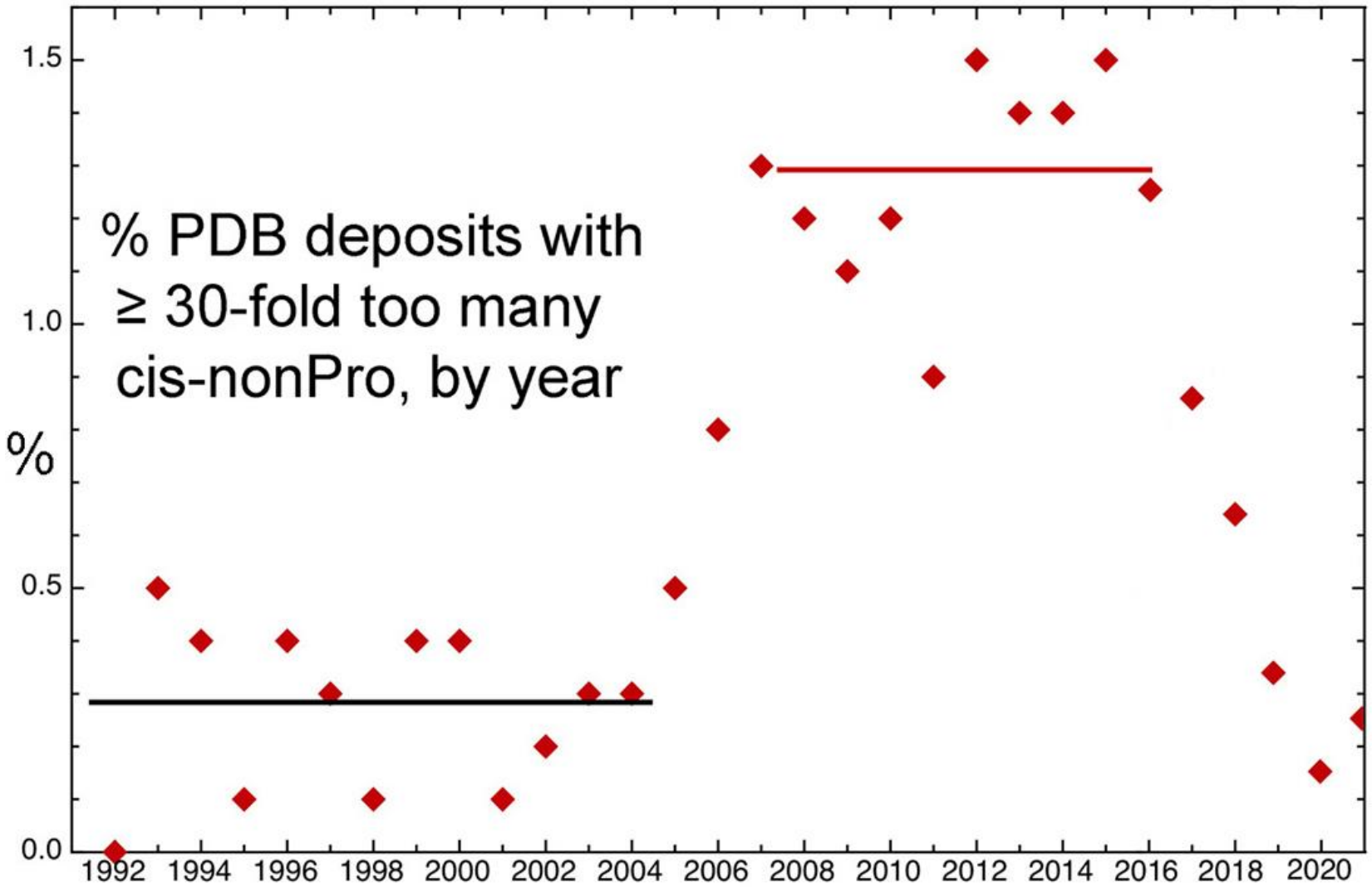
Cis Peptides: Method



- The peptide bond that joins amino acids has partial double bond character and does not rotate freely
- CA-C-N-CA torsion
  - "Omega"
- Usually *trans* (CA on opposite sides)
- Rarely *cis* (both CA on same side)

molprobity.omegalyze

The *Cis*-nonPro crisis



Cis-nonProline peptides are a very rare conformation.

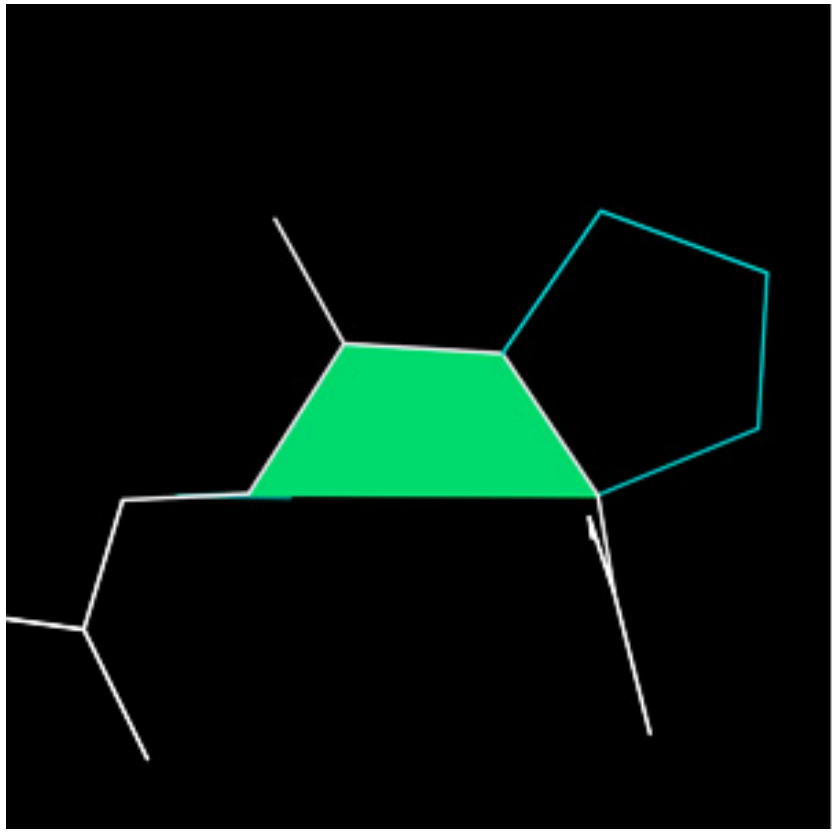
Cis peptides started being hugely over-modeled around 2006.

Once the phenomenon was noticed, an all-hands community response quickly restored normal modeling rates.

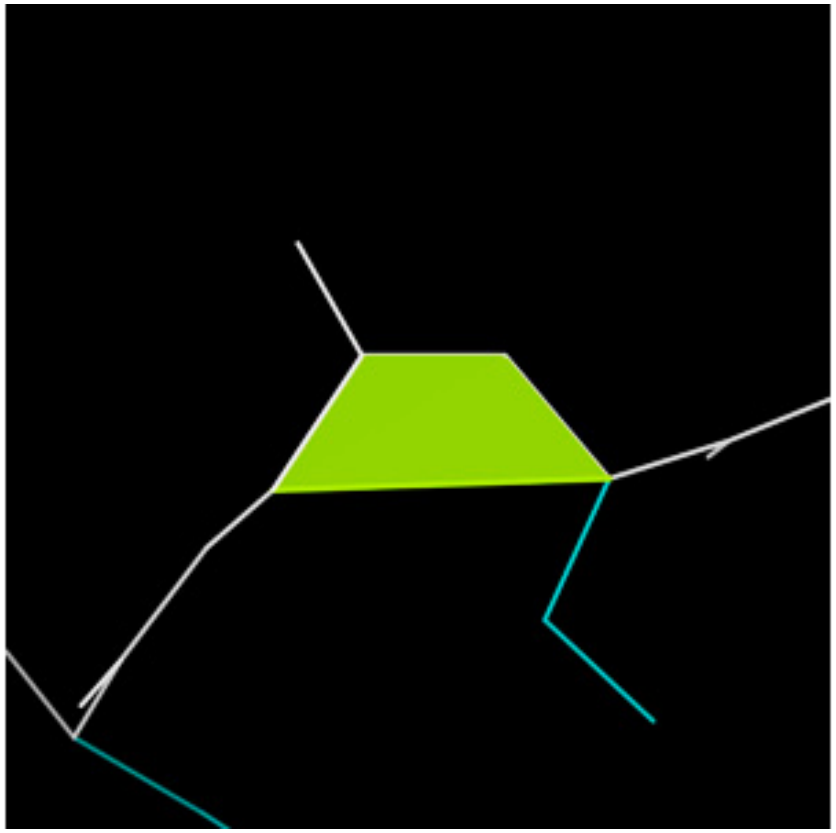
Croll, T. I. (2015). *The rate of cis–trans conformation errors is increasing in low-resolution crystal structures. Acta D, 71*(3), 706-709.

Phenix

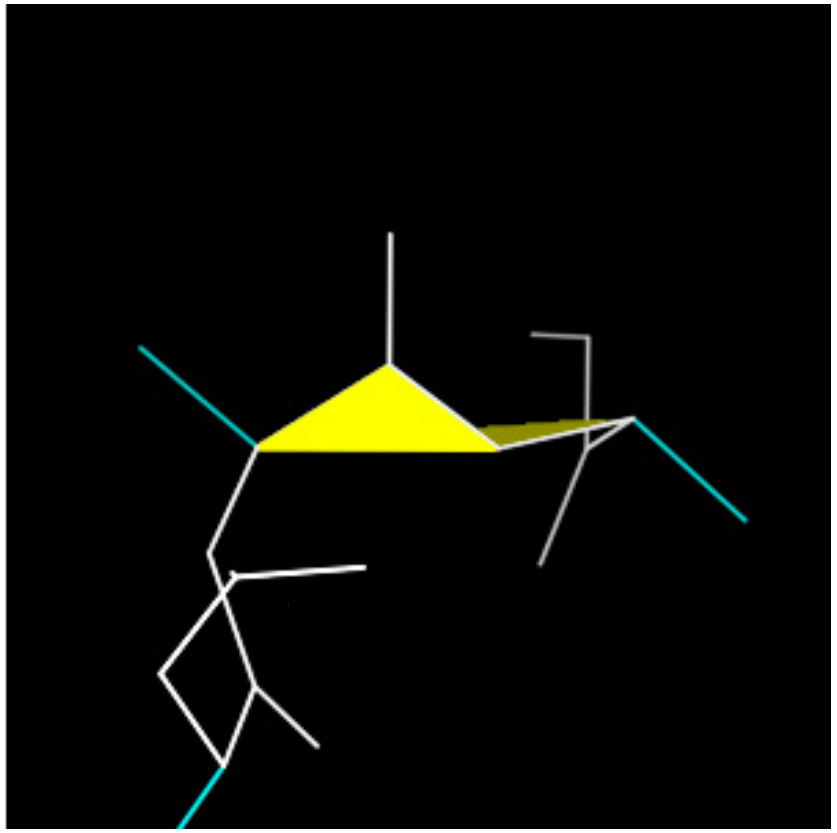
Cis Peptides: Visualization (KiNG)



- *Cis* peptide bond is much more common preceding Proline
  - ~5% of **Proline**
- Gentle green trapezoid fills the characteristic CA-CA space

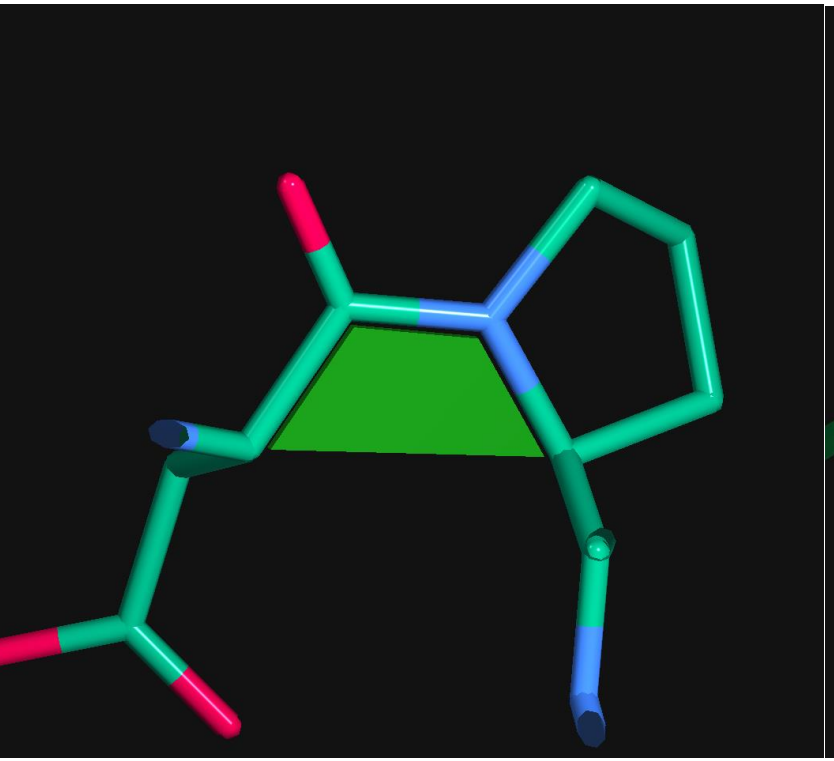


- *Cis* peptide bond is extremely rare preceding other residues
  - ~0.03% of **non-Proline**
- Unpleasantly lime trapezoid fills the characteristic CA-CA space

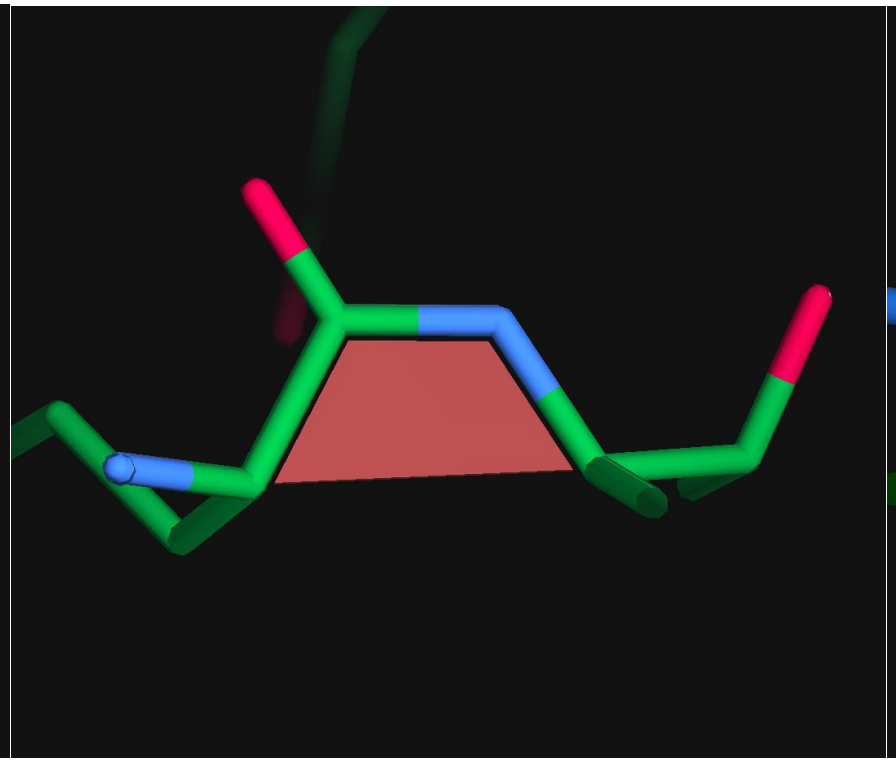


- Peptides **twisted** >30 from planar are severe geometry distortions
- Space is filled with yellow, angle between component planes approximates severity

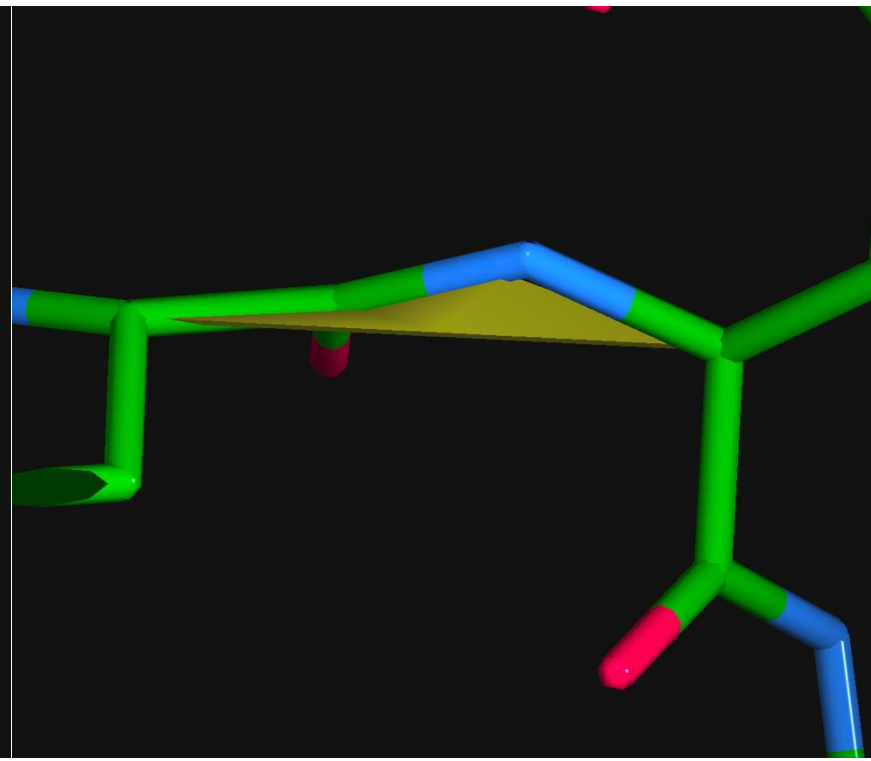
Cis Peptides: Visualization (Coot)



- *Cis* peptide bond is much more common preceding Proline
  - ~5% of **Proline**
- Gentle green trapezoid fills the characteristic CA-CA space

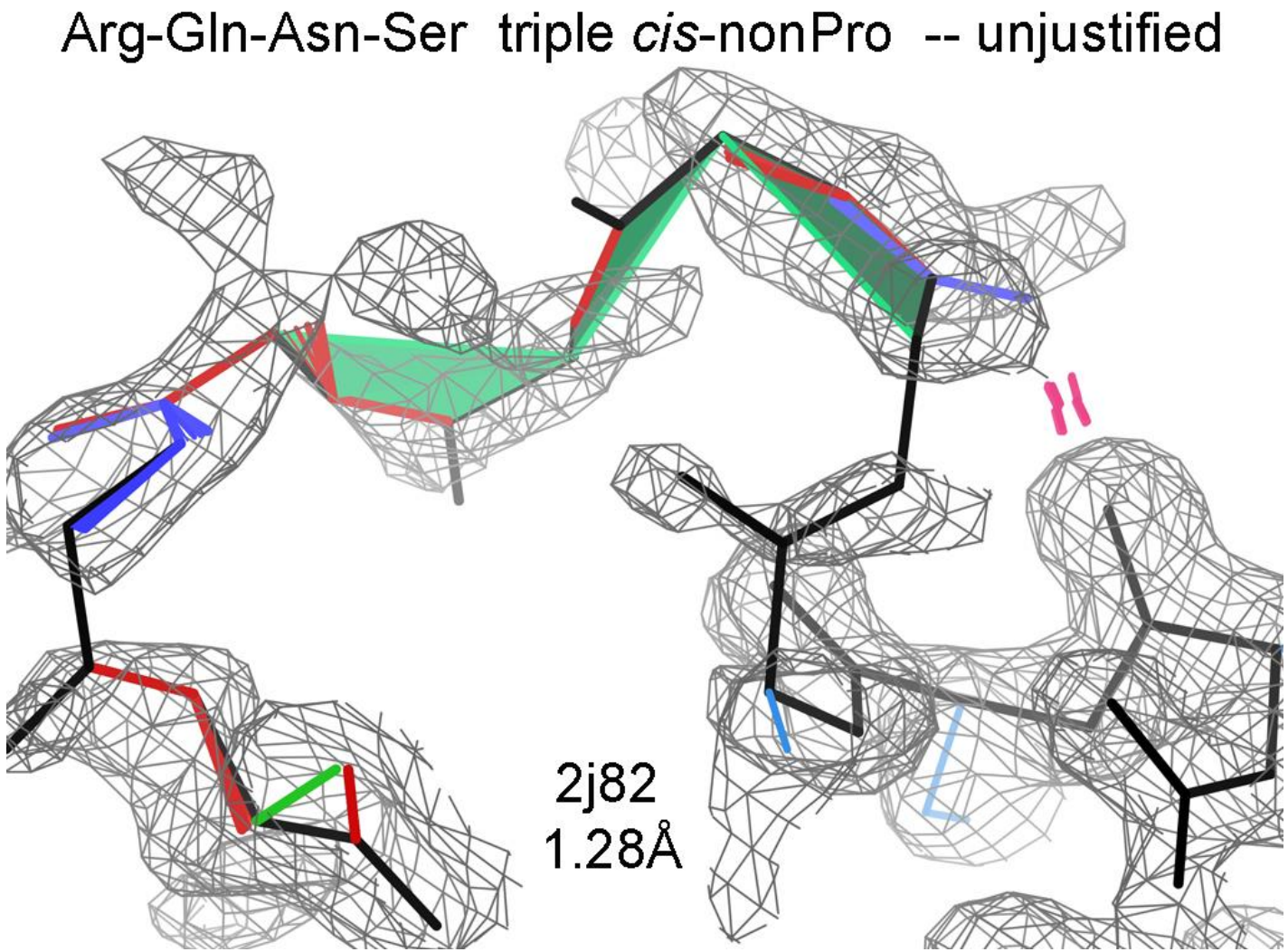


- *Cis* peptide bond is extremely rare preceding other residues
  - ~0.03% of **non-Proline**
- Warning red trapezoid fills the characteristic CA-CA space



- Peptides **twisted** >30 from planar are severe geometry distortions
- Space is filled with yellow, angle between component planes approximates severity

Cis Peptides: possible causes

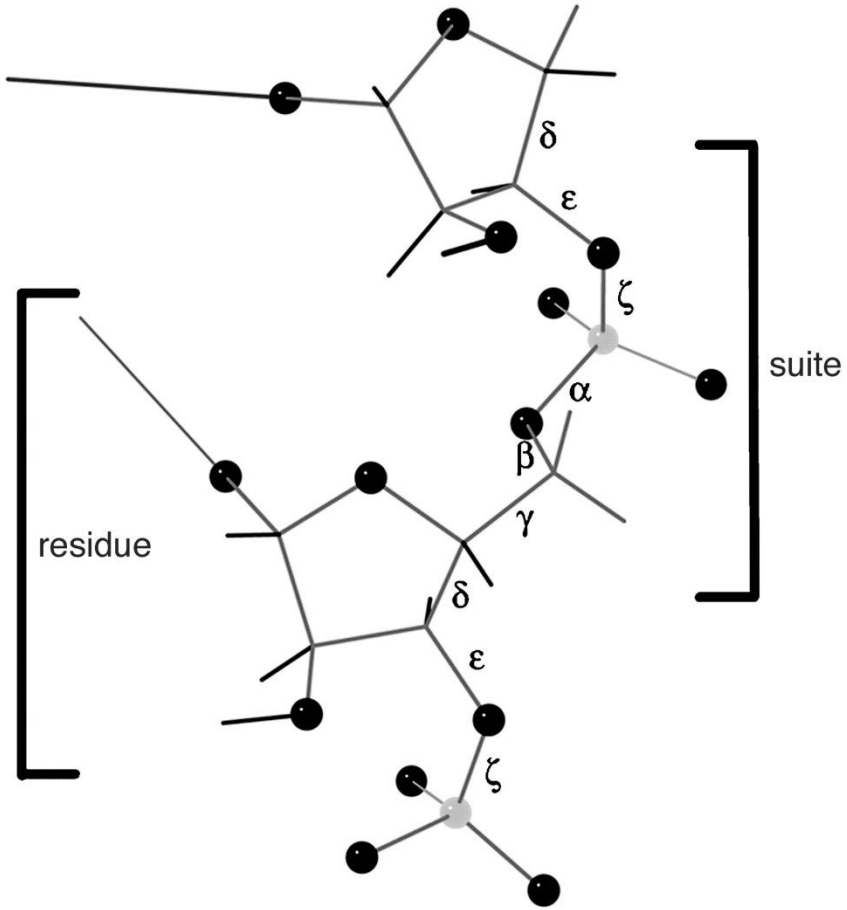


Fit to small density

- The *cis* CA-CA distance is shorter and **seems** to fit better into fragmented density
- A conformation this rare requires more justification than a marginally better fit
- Flip it to *trans* unless density, chemistry, homology, or another source gives you clear support

RNA Suites and Puckers (molprobity.rna_validate)

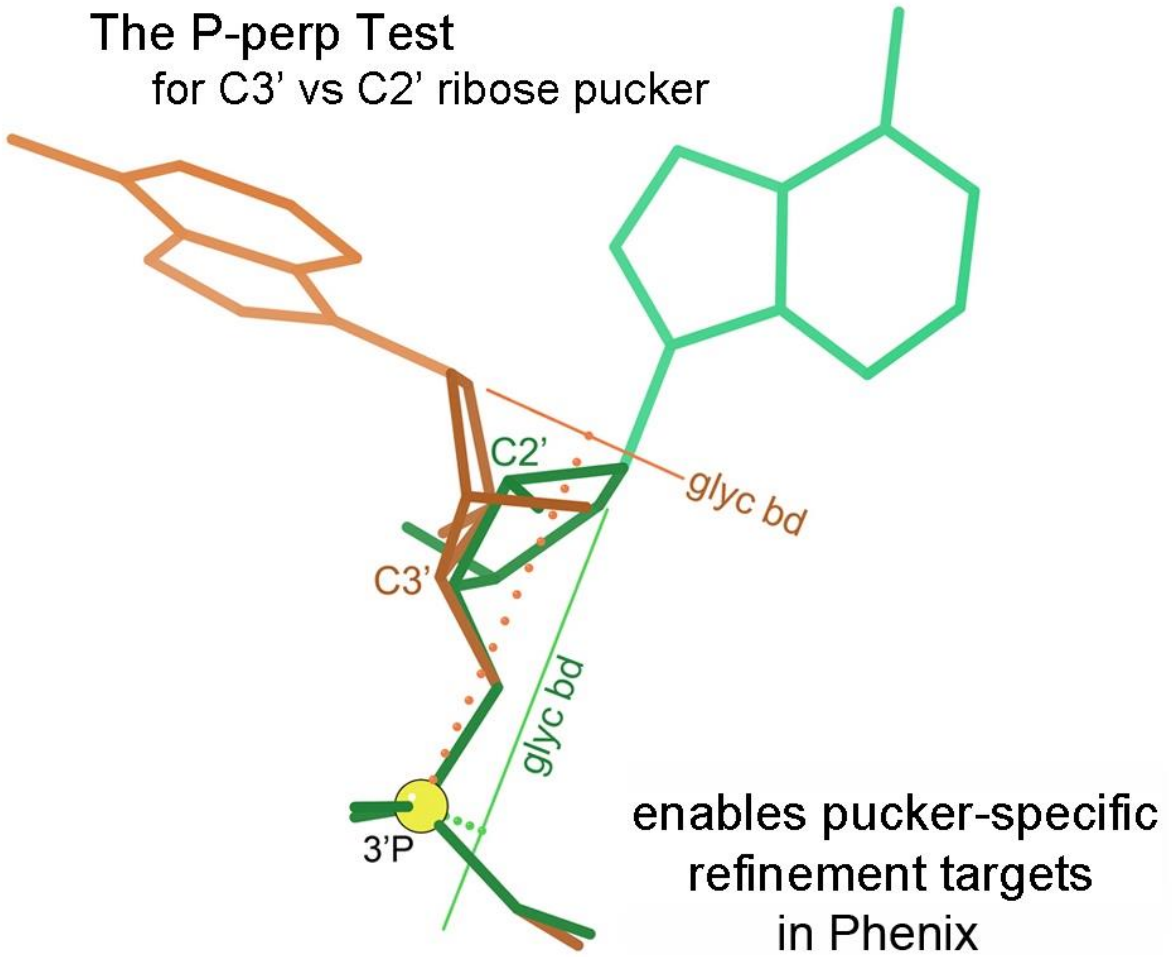
RNA Suites: Method



- Useful RNA backbone division is sugar-to-sugar suite, not P-to-P residue
- Suite conformation names are a combination of a number and a letter/character
  - e.g. 1A is the most common A-form helix conformation
- Outliers are named as !!
  - Pronounced “bang, bang”
  - Many !!’s are real, rare conformations

molprobity.suitename

RNA Ribose Puckers: Method

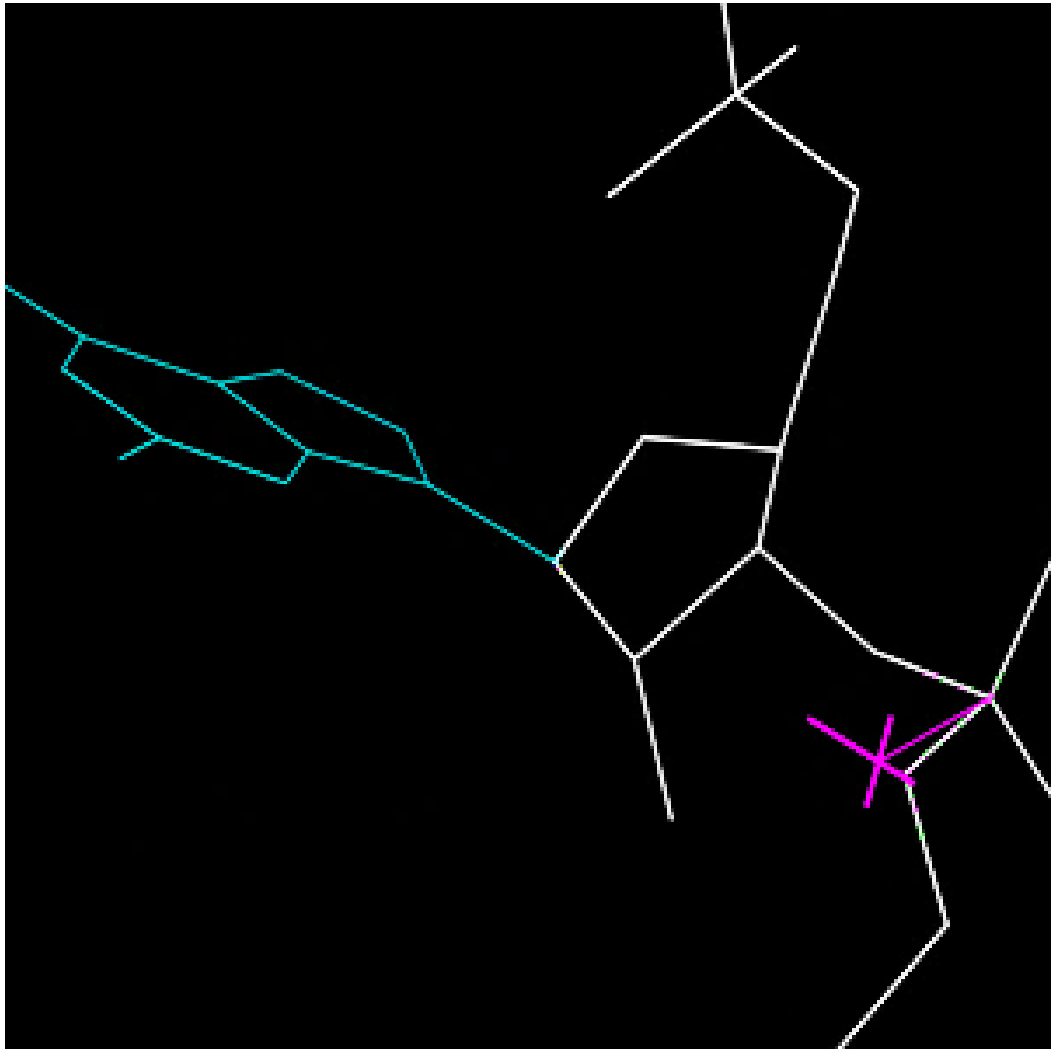


- The backbone ribose in RNA can have one of two pucker states
  - C2' endo
  - C3' endo
- Ribose pucker correlates very strongly with perpendicular distance from the 3'phosphate to the glycosidic bond vector
  - Glycosidic bond joins ribose sugar to nucleobase

- At low resolution, perpendicular distance is easy to see, ribose pucker is hard to see
- If there's a mismatch, the pucker is probably wrong

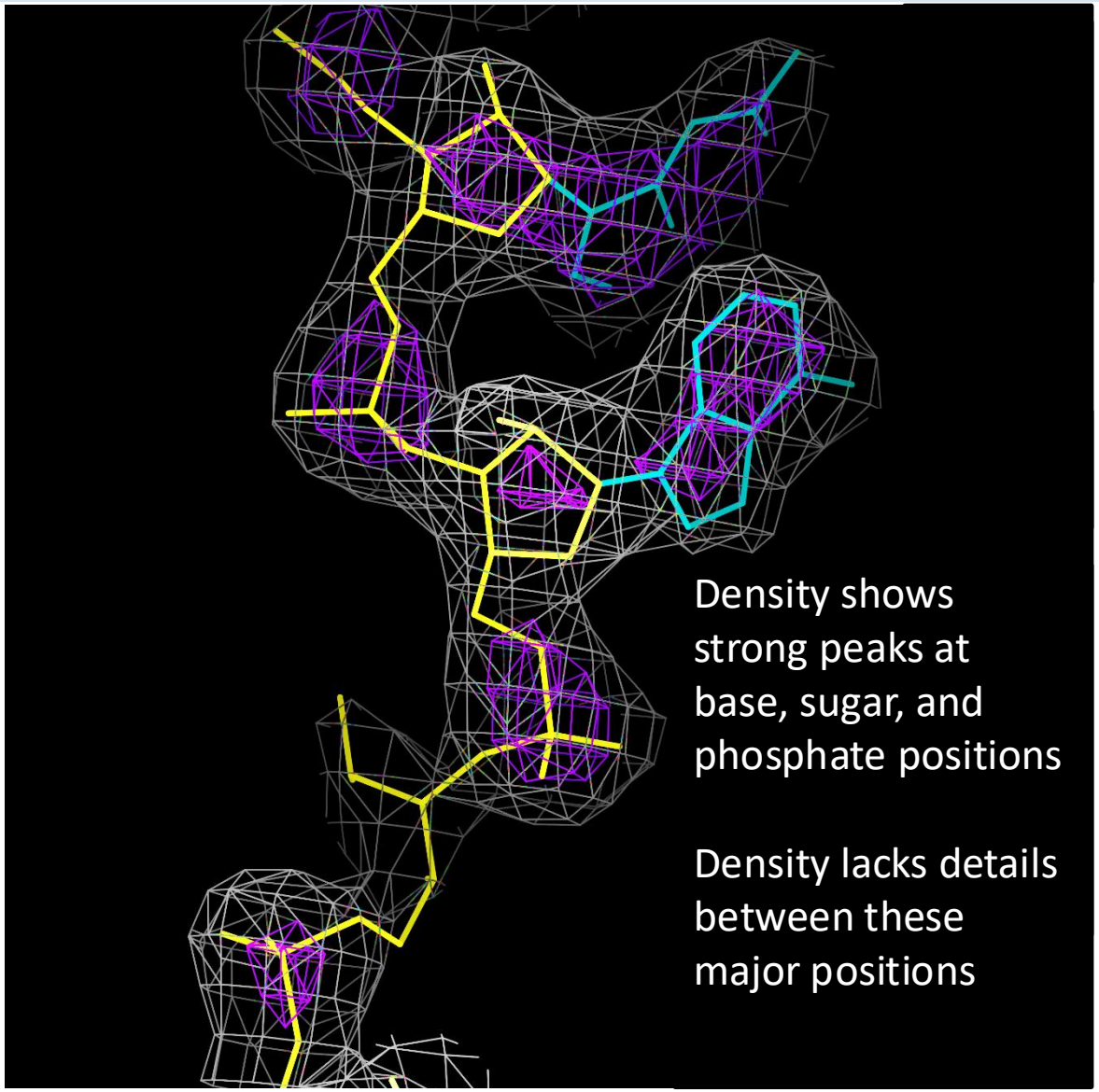
molprobity.rna_validate

RNA Ribose Puckers: Visualization



- A magenta cross is draw for each incorrect ribose pucker
  - Long end of cross points along glycosidic bond vector
  - Cross is connected to 3’phosphate by the perpendicular distance line

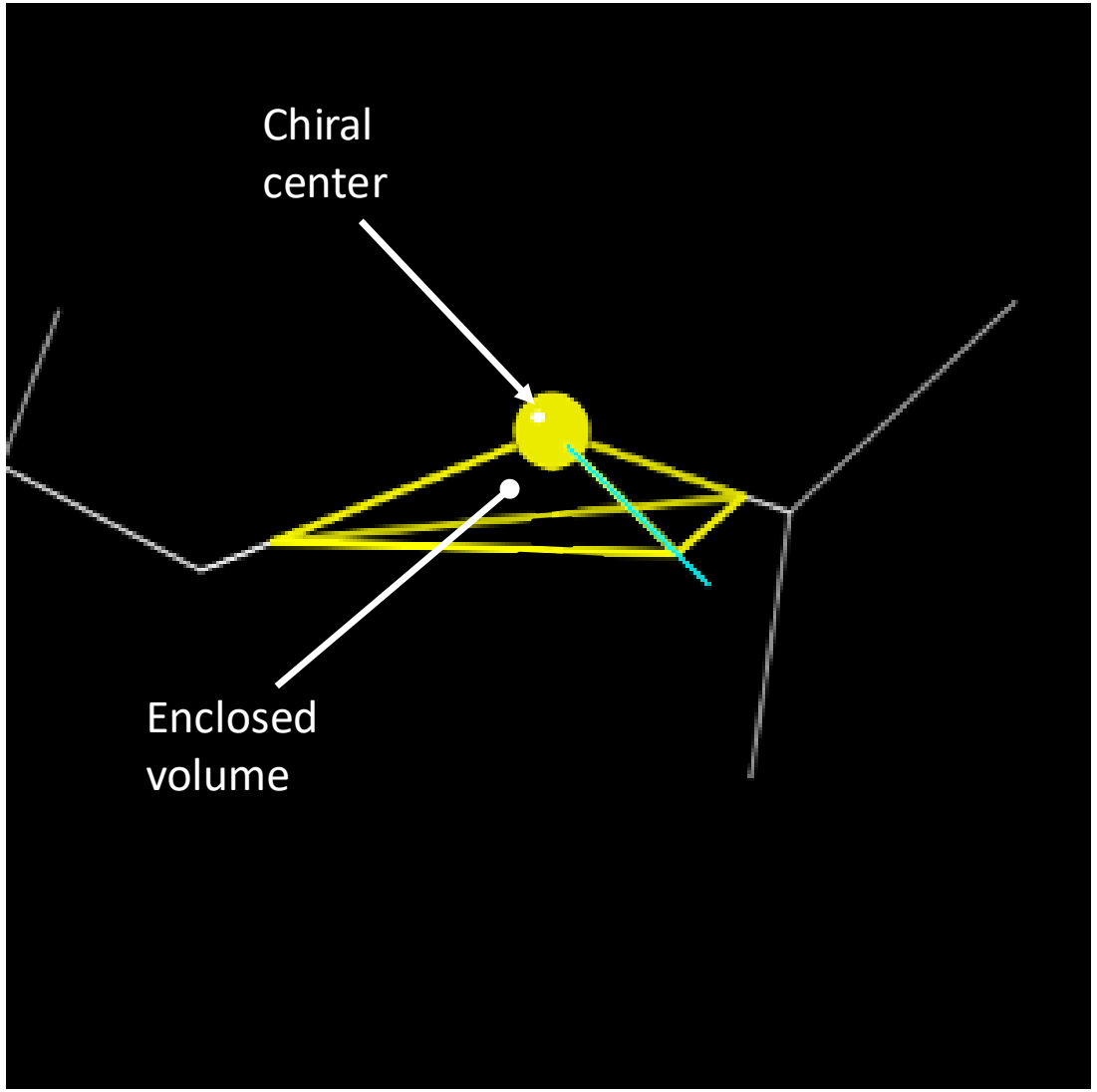
RNA Errors: possible Causes



- RNA backbone has many degrees of freedom
- Electron density often leaves RNA backbone underdetermined
  - Even when bases are better resolved
- More tools to help with this are in development

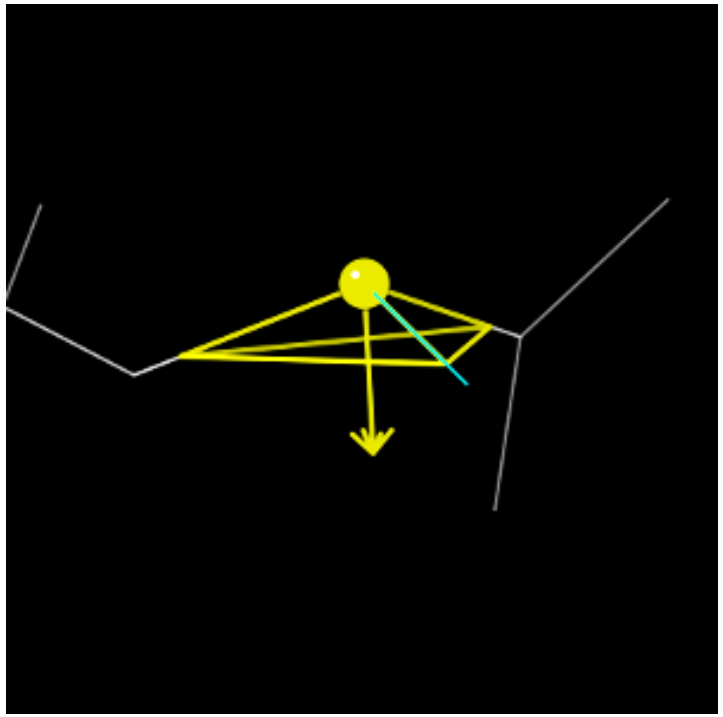
Chiral Volume Outliers
(Very rare unless something is weird)
(molprobity.chiral_validation)

Chiral Volume Outliers: Method



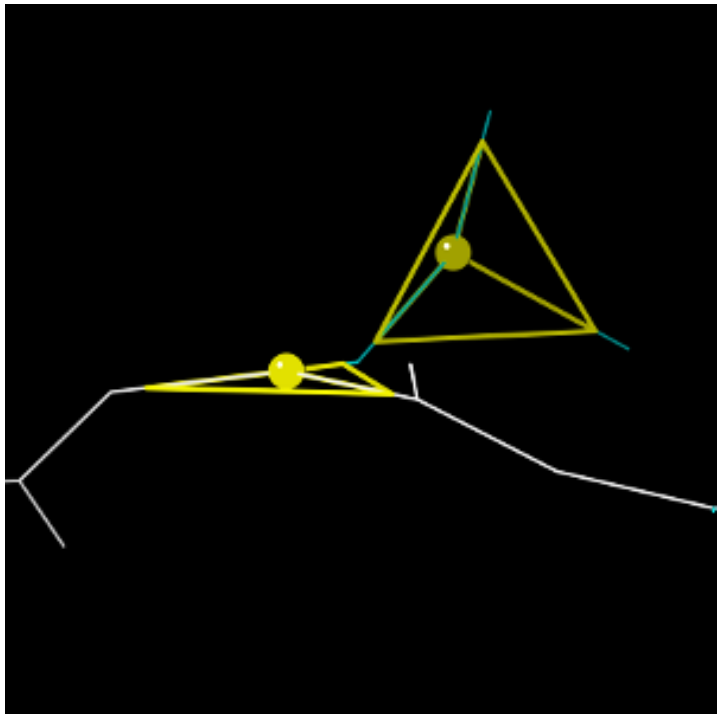
- Tetrahedral atoms with 4 distinct substituents are chiral
- Do a little light vector math to find the volume enclosed by the chiral center and its three heaviest children
  - Magnitude of volume indicates how tetrahedral the bonding is
  - Sign of volume indicates handedness (L vs D)

Chiral Volume Outliers: Visualization and Causes

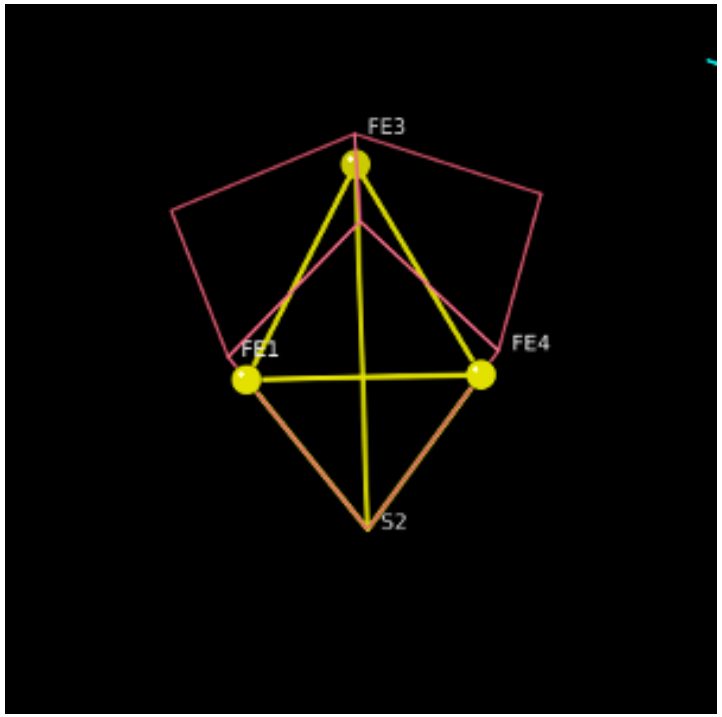


True handedness swaps

- D-amino acids with L names
- L-amino-acids with D names



Squished or flattened geometry errors

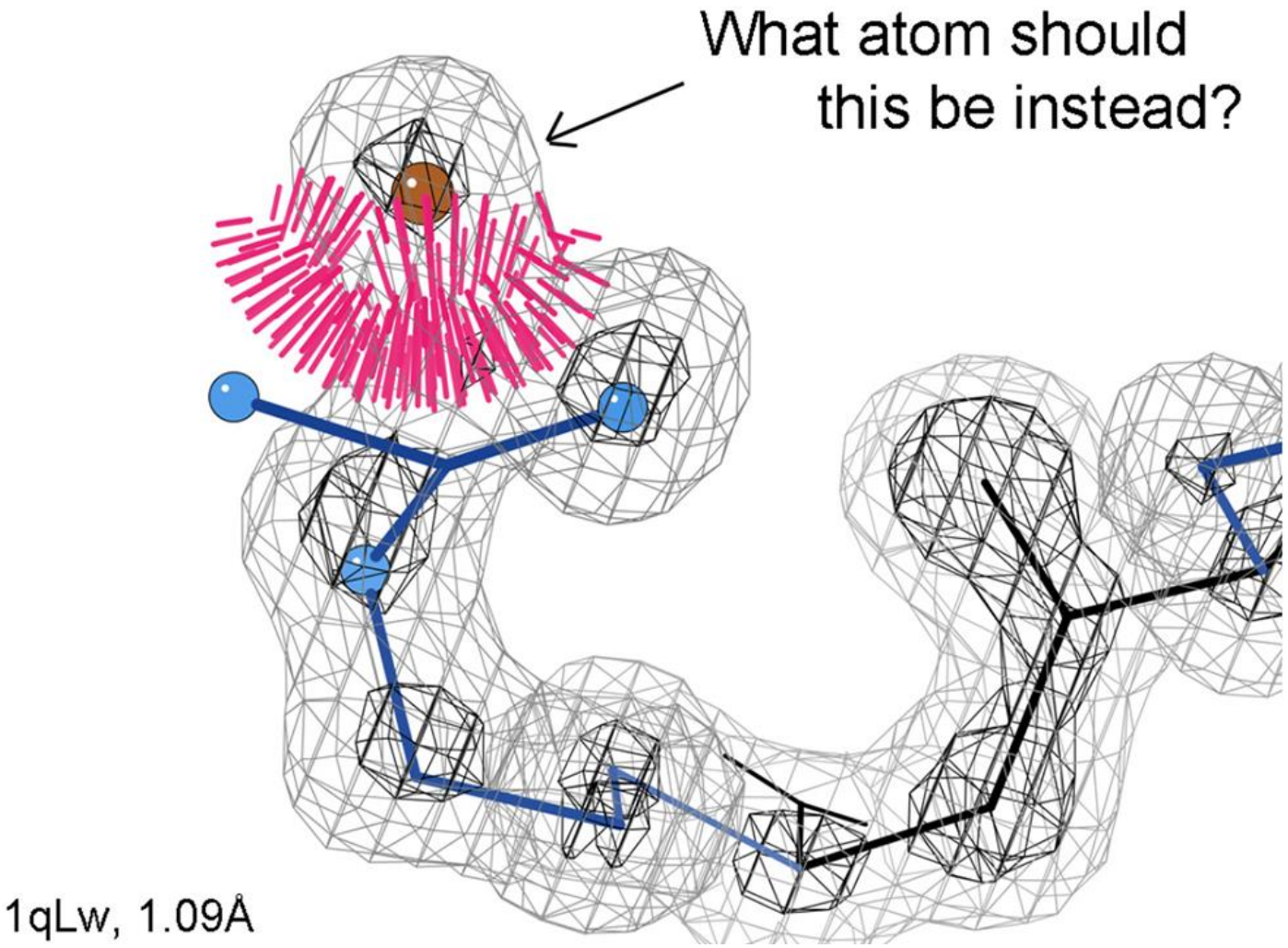


Atom naming errors

- FeS clusters
- Swapping CD1 and CD2 names in Leu

UnDowser
Water Validation
(molprobity.undowser_validation)

UnDowser: Method



- Undowser is a tool for finding incorrect waters
- Use all-atom contact analysis to find waters with steric clashes
- Identify probable substitutions for each problem water
  - Ions
  - Ligands
  - Sidechain alternates
  - Nothing!

molprobity.undowser_validation

UnDowser: Visualization

Multi-criterion visualizations

View in KiNG | View in NGL | Download (294 Kb)

View (135 Kb)

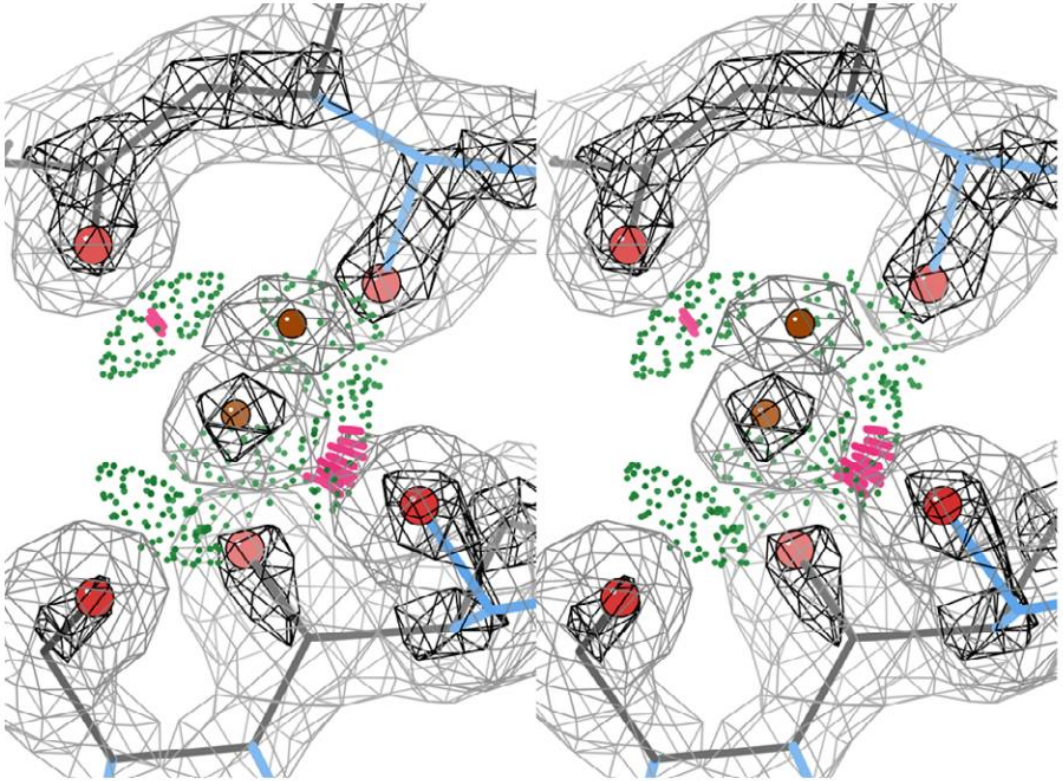
View (4.9 Kb)

SUMMARY: 6 waters out of 58 have clashes (10.34%)

Water ID	Clashes with	Water B	Contact B	Clash Severity	Clash with Polar May be ion	Clash with non-polar Unmodeled alt or noise	Clash with water Occ <1 or ligand	Clash with altloc Add or rename alts
A: 125 :HOH:	CG of A: 51 :GLU:	26.53	26.06	0.506		×		
	HG3 of A: 51 :GLU:	26.53	26.06	0.503		×		
A: 107 :HOH:	HA2 of A: 10 :GLY:	23.36	18.74	0.736		×		
A: 100 :HOH:	HE3 of A: 48 :LYS:	24.10	20.04	0.501		×		
	CE of A: 48 :LYS:	24.10	20.04	0.425		×		
A: 114 :HOH:	O of A: 122 :HOH:	27.11	26.13	0.651			×	
A: 122 :HOH:	O of A: 114 :HOH:	26.13	27.11	0.651			×	
A: 80 :HOH:	HB3 of A: 39 :ASP:	22.27	24.16	0.426		×		

- MolProbity has a dedicated chart for water analysis
- Each clashing water is listed
  - Colored by severity
  - Possible causes marked in table
- Recently added to Phenix commandline
  - *phenix.undowser_validation*
  - Coming soon to GUI

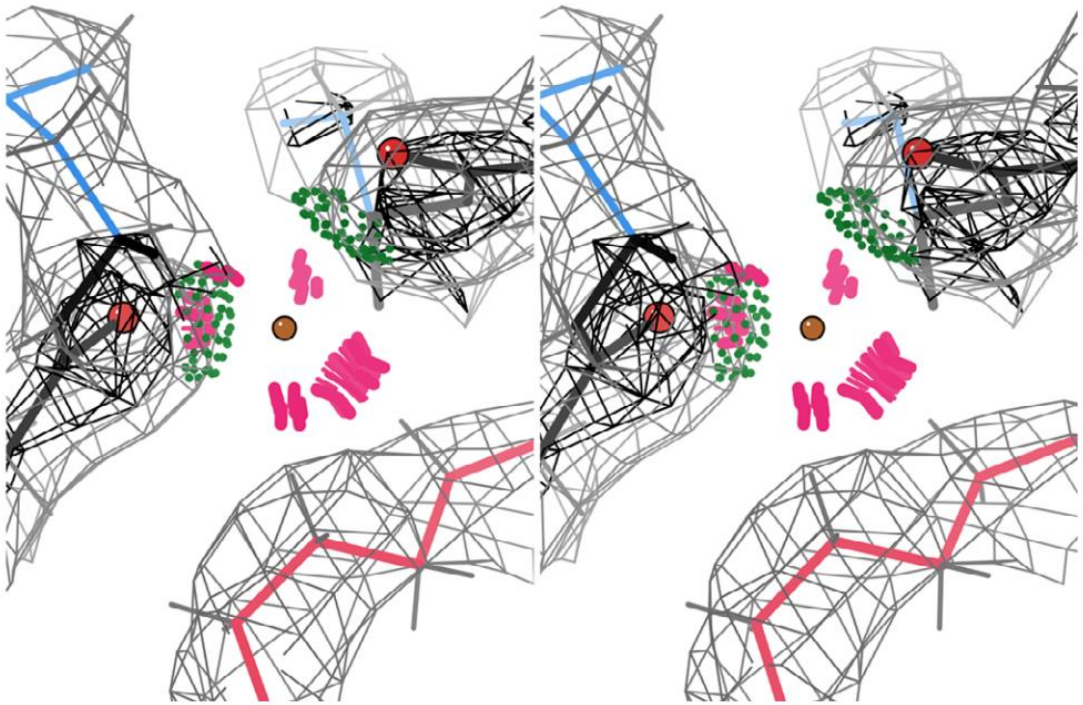
A water that should be an ion



(Stereo image)
HOH 606 from 6hhm, 1.23 Å

- Very strong density peak
- Octahedral contact geometry
  - (water is tetrahedral)
- Contacts are all polar groups (δ-)
- This is actually a + ion, probably Na+

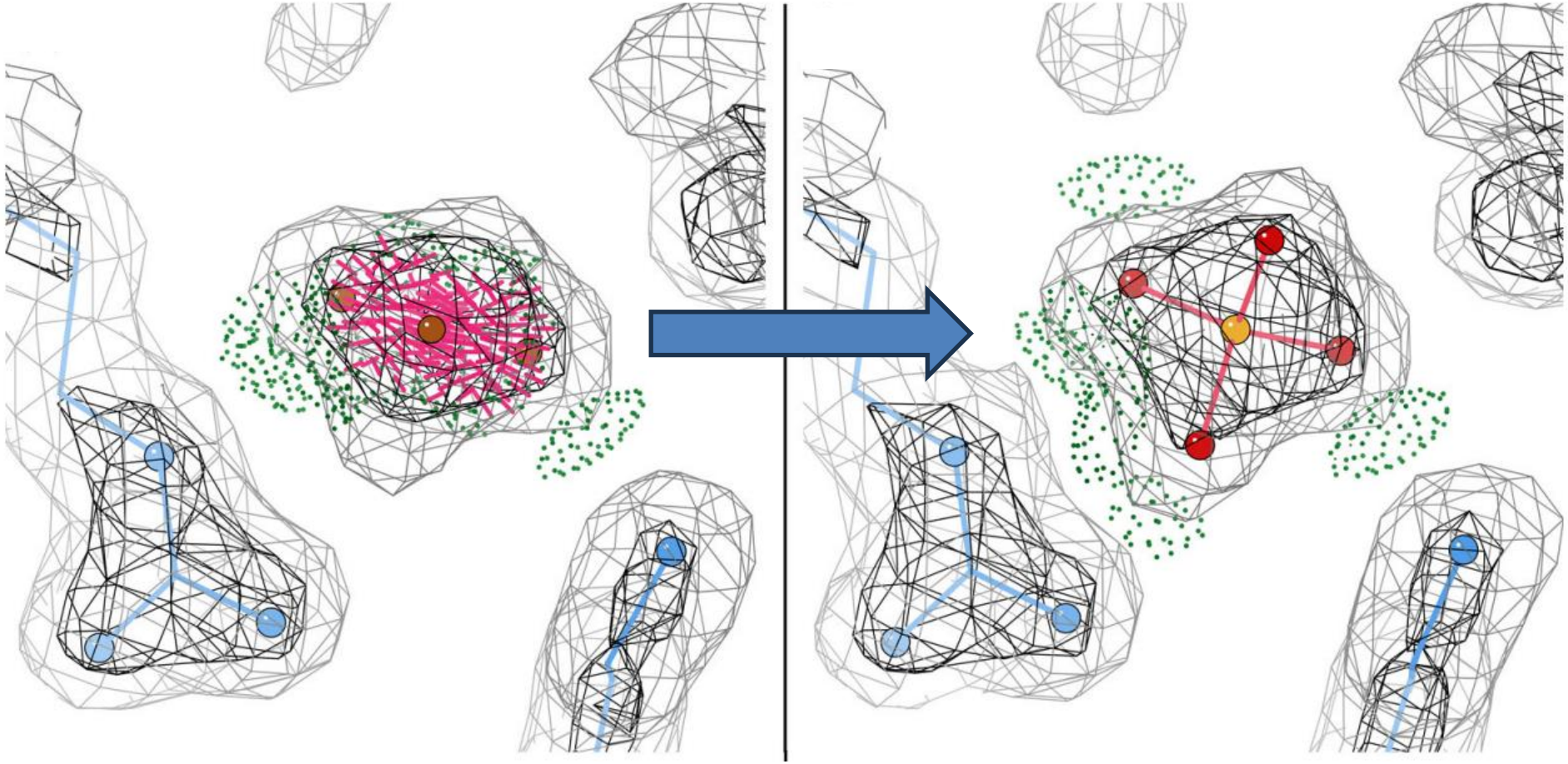
A water that should not be



(Stereo image)
HOH 504 from 5onu, 2.22 Å

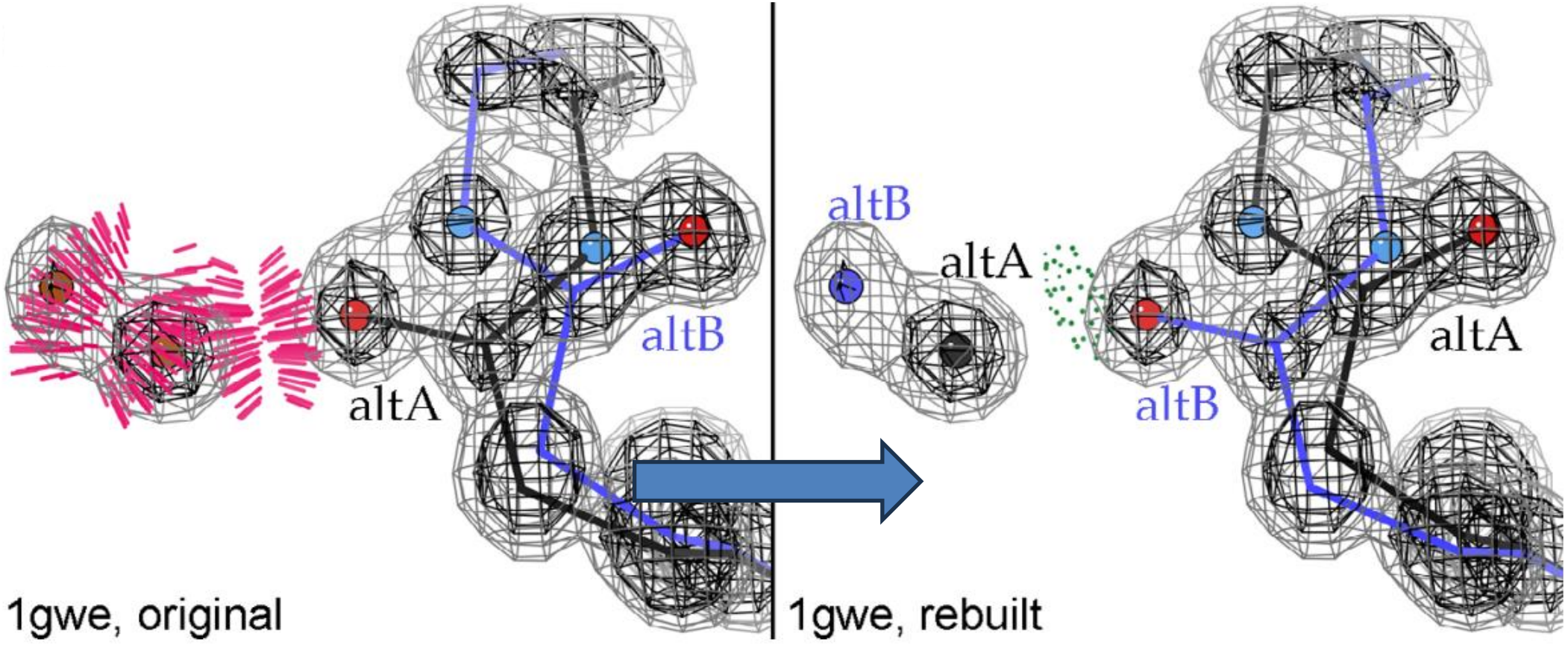
- No density peak
- Mix of polar and non-polar contacts
  - So unlikely to be a coordinated ion
- This water doesn't really exist

Waters that should be ligands



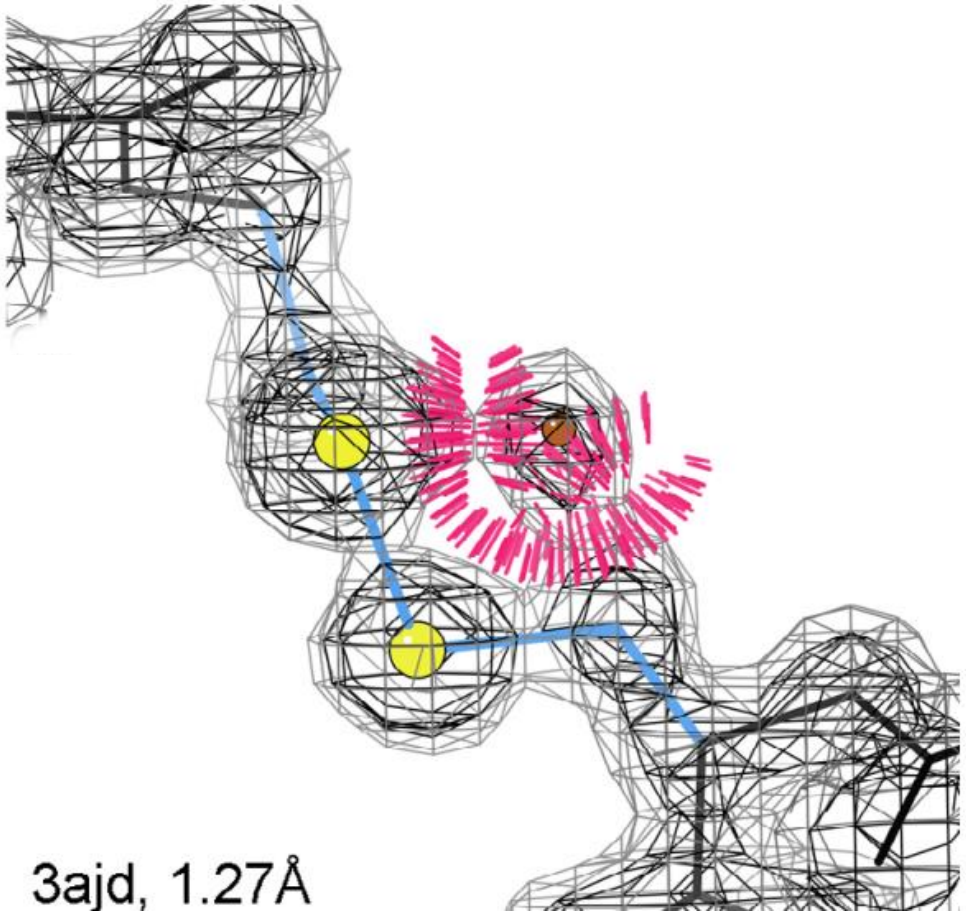
- Densely-clashing waters may actually be a ligand

Waters that should be partial occupancy



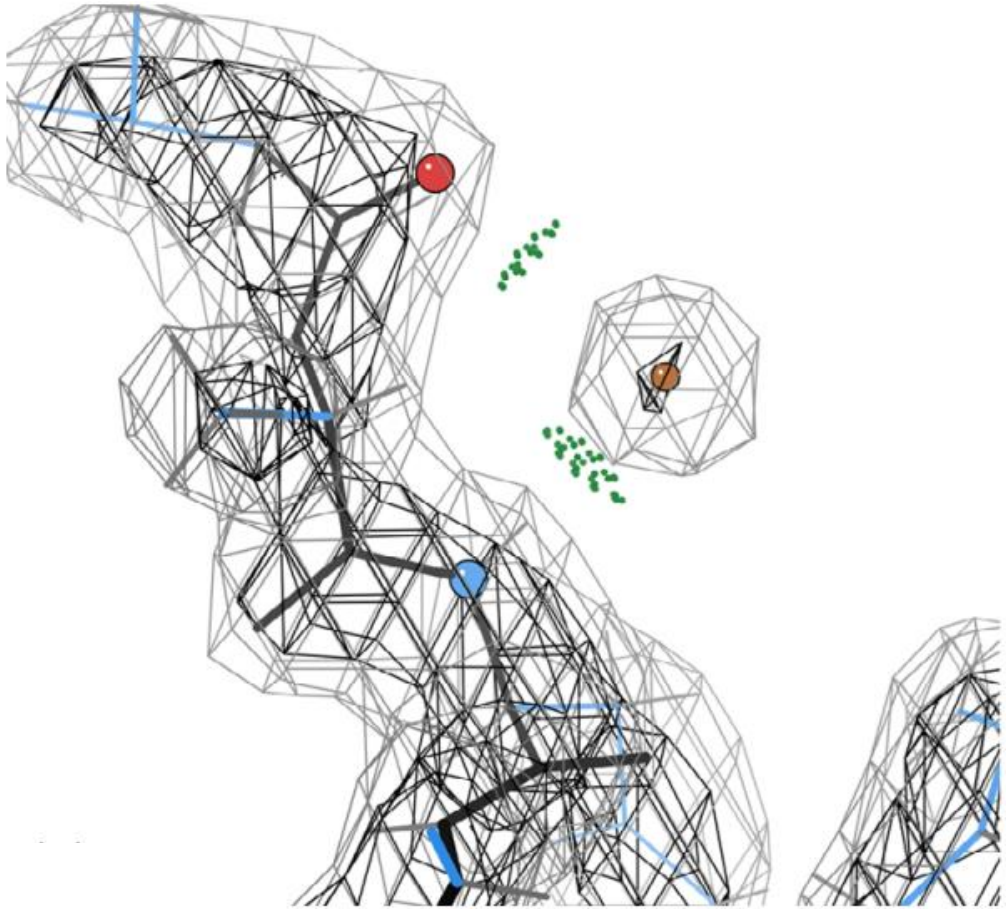
- Densely-clashing waters may actually be part of an alternate conformation network

Waters that replace alternates



- Very close contacts
  - (Covalent bond distance)
- Clash with non-terminal sidechain atoms
- Could be an unmodeled alternate conformation

Waters can be real, too!



- Clear density peak
  - Weaker than macromolecule density is fine
- Hydrogen bonds
- Contacts with both δ+ and δ- polar partners, so an ion is unlikely

MolProbity Score

- The MolProbity score combines a few validations and scales the result to look like a resolution (only for proteins)
  - Clashscore
  - Ramachandran
  - Rotamers

- MolProbity score better than model resolution is good
- MolProbity score worse than model resolution is bad

It's a global score, and thus obscures local structure quality issues!

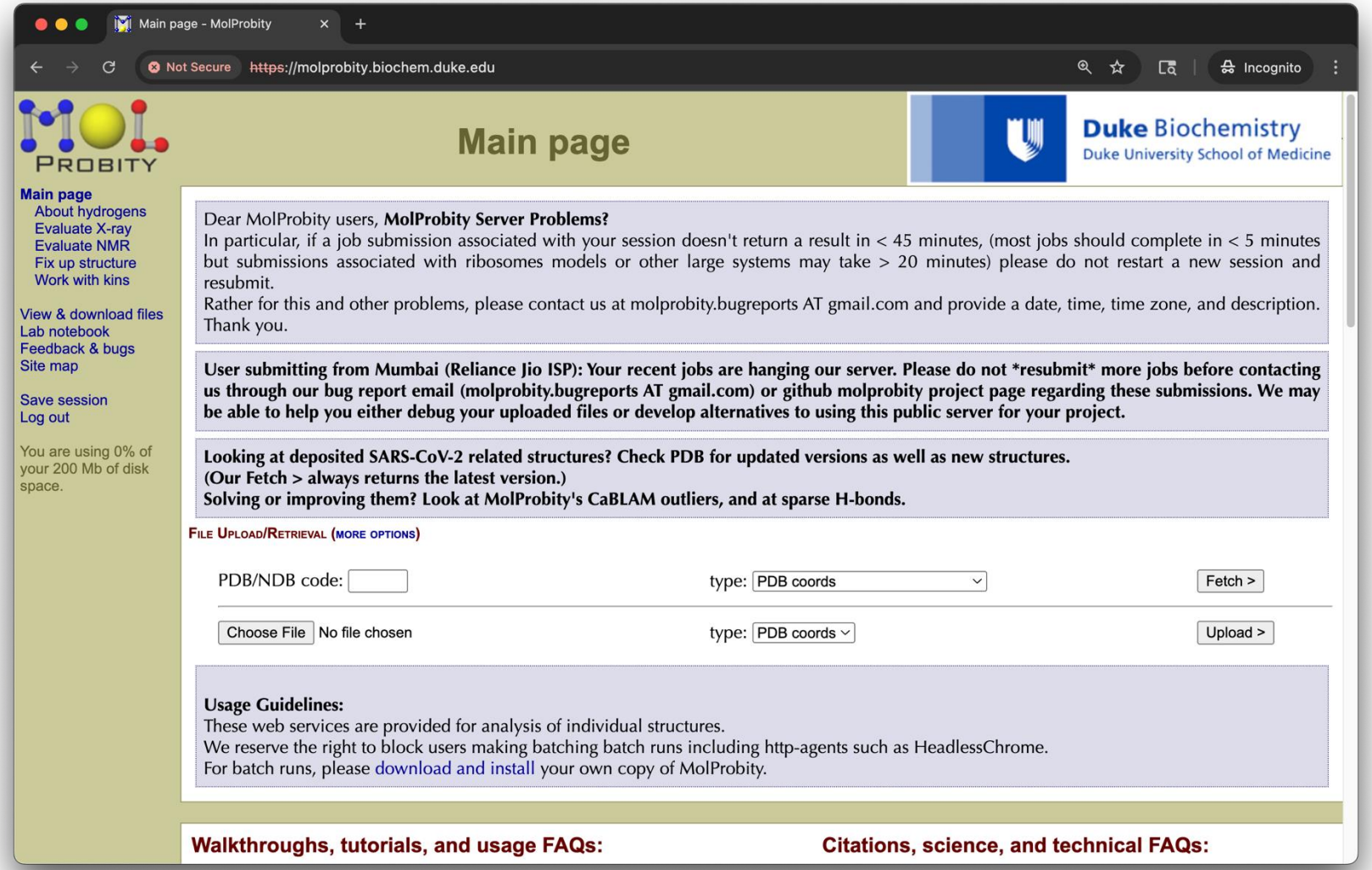
molprobity.molprobity

Phenix

MolProbity

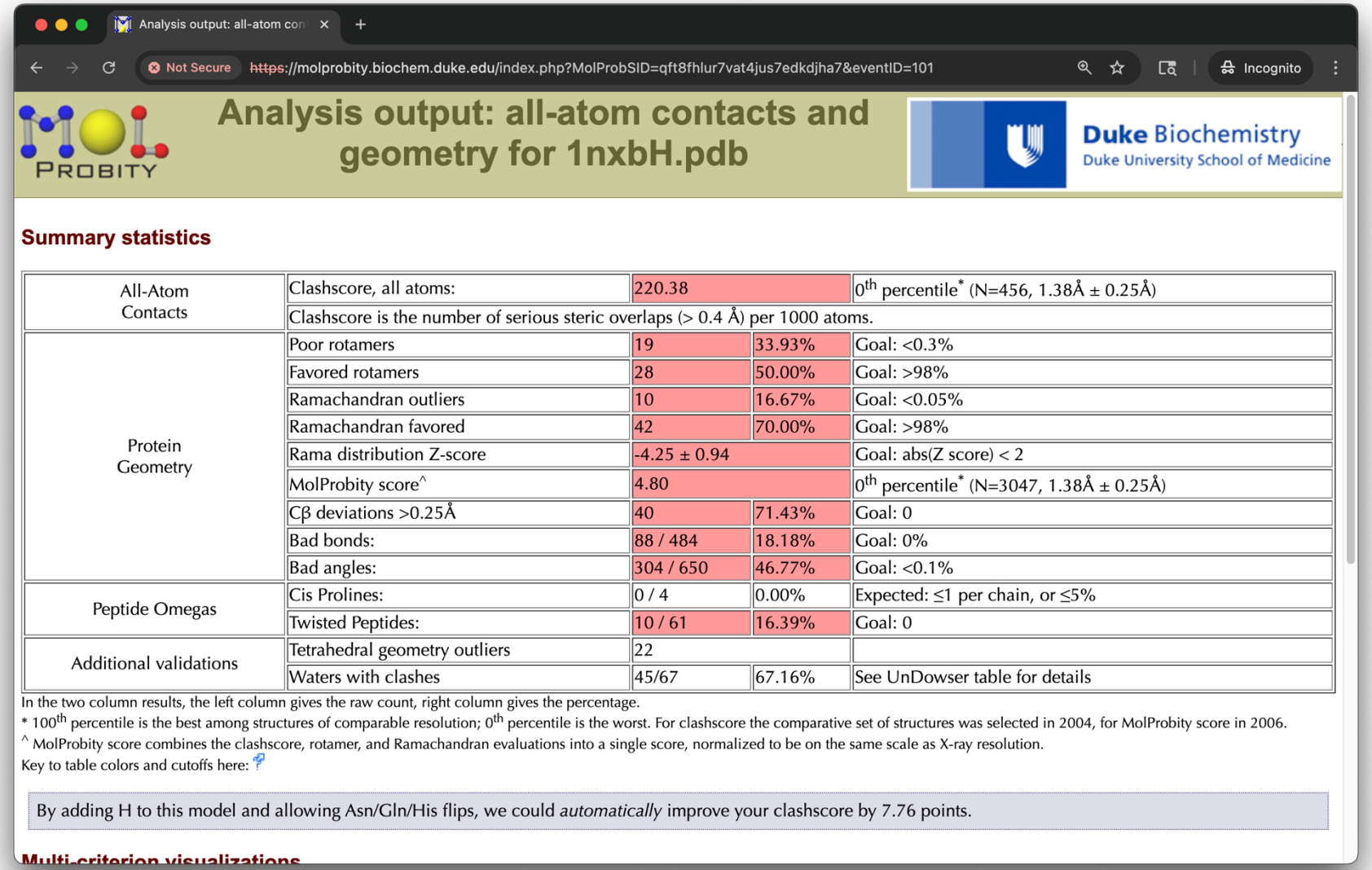
Free online server:
molprobity.biochem.duke.edu

- Also integrated into Phenix and CCP4
- Runs all the validations described today
- Adds hydrogens, evaluates all-atom contacts, rotamers, Ramachandran, etc.



The MolProbity Summary Table

- Dashboard for your structure — check this first!
- Color coding: green = good, yellow = caution, red = outlier
- Key metrics: clashscore, rotamer outliers, Rama outliers, MolProbity score
- Goal values for each metric



Analysis output: all-atom con

Not Secure https://molprobity.biochem.duke.edu/index.php?MolProbSID=qft8fhlur7vat4jus7edkdjha7&eventID=101

Incognito

MOLPROBITY

Analysis output: all-atom contacts and geometry for 1nxbH.pdb

Duke Biochemistry
Duke University School of Medicine

Summary statistics

All-Atom Contacts	Clashscore, all atoms:	220.38		0th percentile* (N=456, 1.38Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	19	33.93%	Goal: <0.3%
	Favored rotamers	28	50.00%	Goal: >98%
	Ramachandran outliers	10	16.67%	Goal: <0.05%
	Ramachandran favored	42	70.00%	Goal: >98%
	Rama distribution Z-score	-4.25 ± 0.94		Goal: abs(Z score) < 2
	MolProbity score^	4.80		0th percentile* (N=3047, 1.38Å ± 0.25Å)
	Cβ deviations >0.25Å	40	71.43%	Goal: 0
	Bad bonds:	88 / 484	18.18%	Goal: 0%
	Bad angles:	304 / 650	46.77%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
	Twisted Peptides:	10 / 61	16.39%	Goal: 0
Additional validations	Tetrahedral geometry outliers	22		
	Waters with clashes	45/67	67.16%	See UnDowser table for details

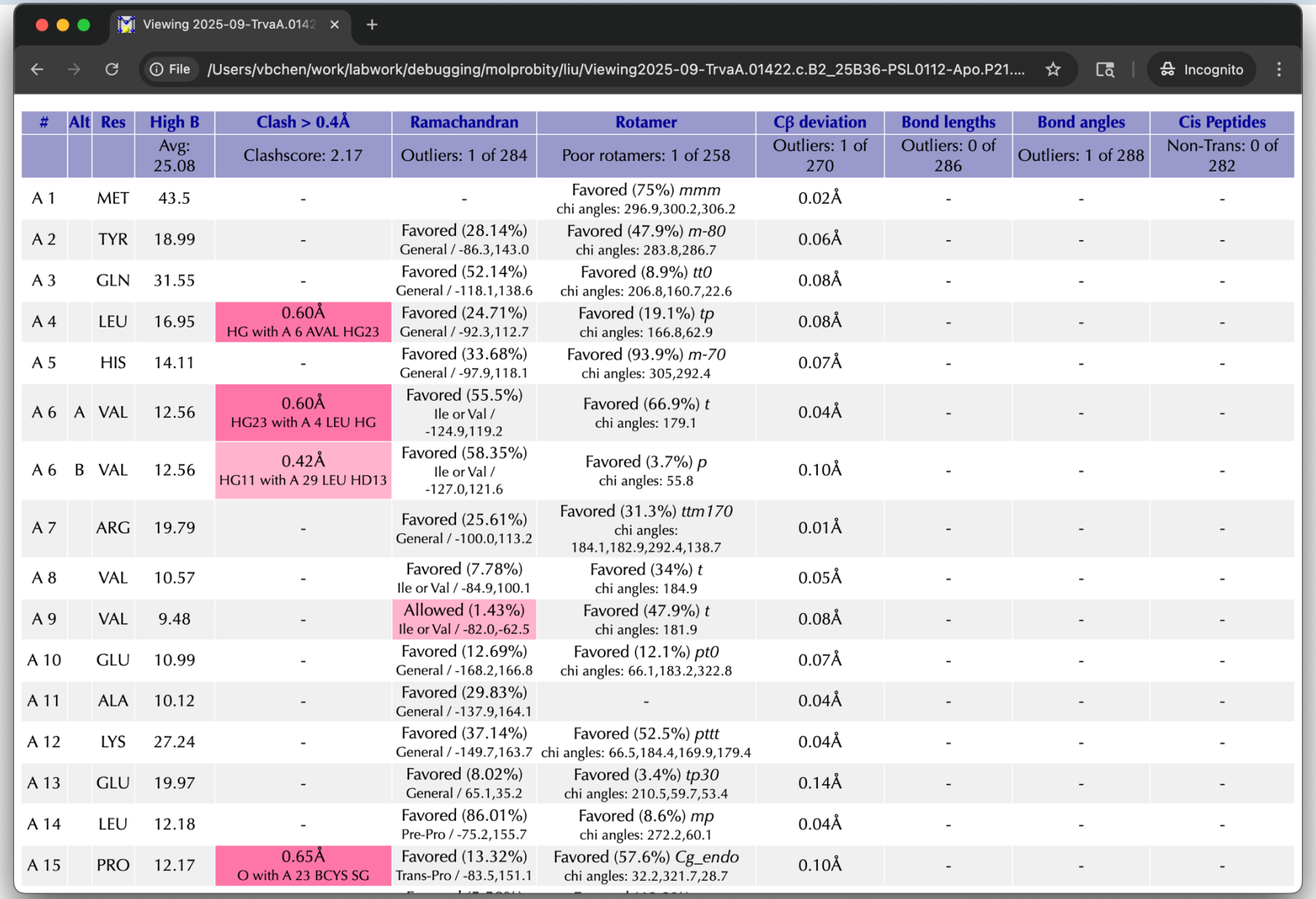
In the two column results, the left column gives the raw count, right column gives the percentage.
* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.
^ MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.
Key to table colors and cutoffs here:

By adding H to this model and allowing Asn/Gln/His flips, we could *automatically* improve your clashscore by 7.76 points.

Multi-criterion visualizations

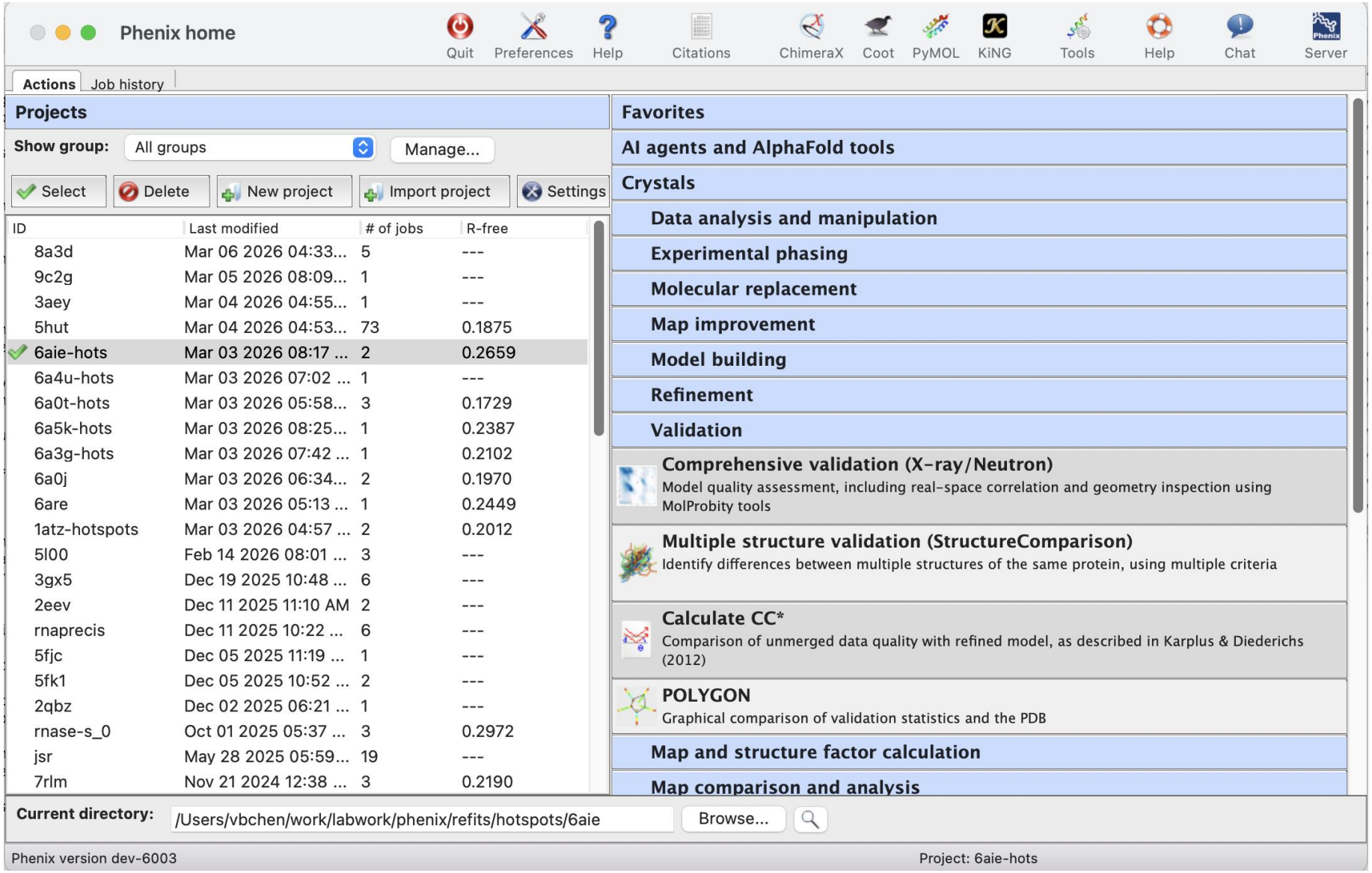
Phenix

The MolProbity Residue Table

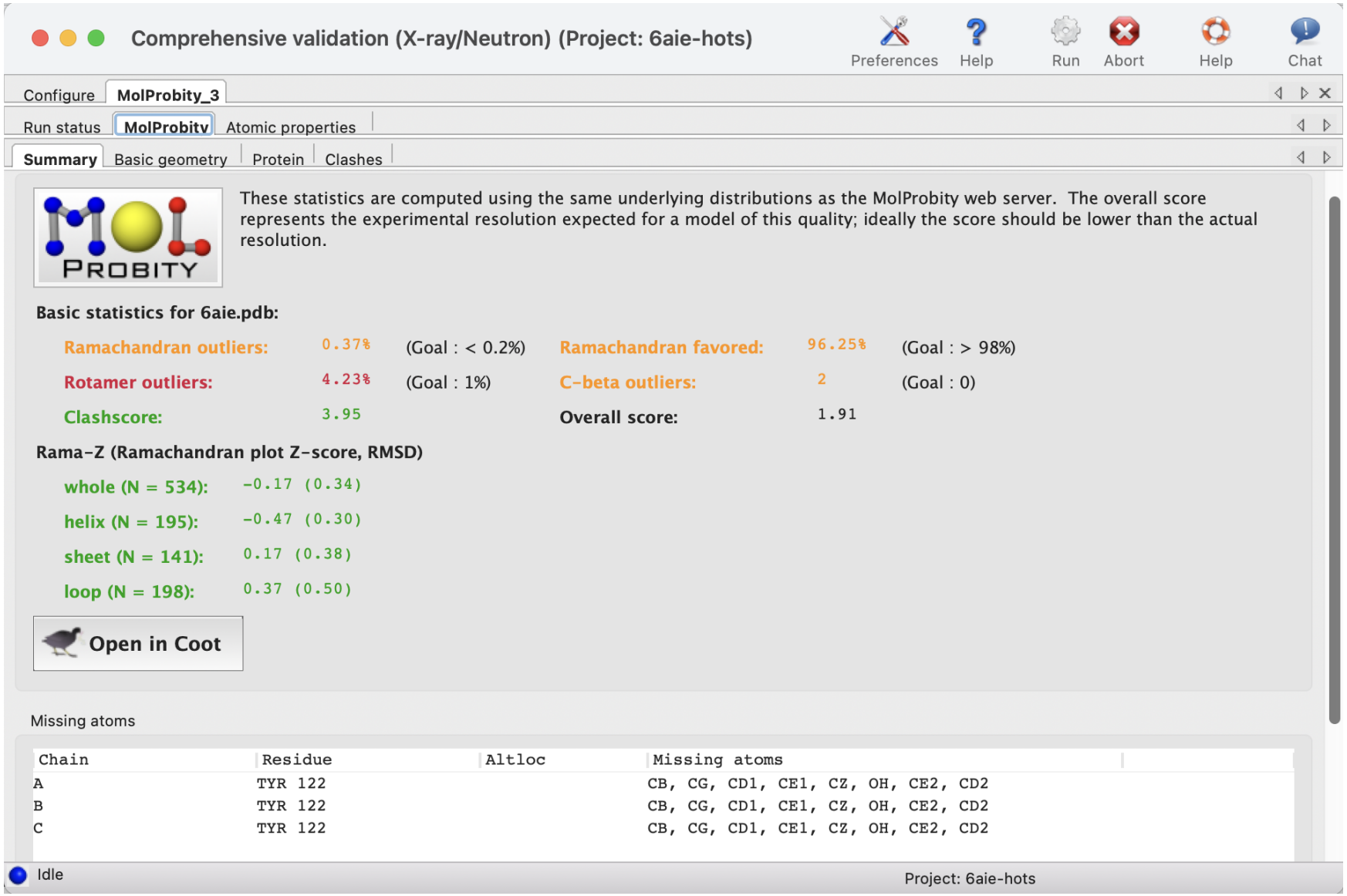


#	Alt	Res	High B	Clash > 0.4Å	Ramachandran	Rotamer	Cβ deviation	Bond lengths	Bond angles	Cis Peptides
			Avg: 25.08	Clashscore: 2.17	Outliers: 1 of 284	Poor rotamers: 1 of 258	Outliers: 1 of 270	Outliers: 0 of 286	Outliers: 1 of 288	Non-Trans: 0 of 282
A 1		MET	43.5	-	-	Favored (75%) *mmm* chi angles: 296.9,300.2,306.2	0.02Å	-	-	-
A 2		TYR	18.99	-	Favored (28.14%) General / -86.3,143.0	Favored (47.9%) *m-80* chi angles: 283.8,286.7	0.06Å	-	-	-
A 3		GLN	31.55	-	Favored (52.14%) General / -118.1,138.6	Favored (8.9%) *tt0* chi angles: 206.8,160.7,22.6	0.08Å	-	-	-
A 4		LEU	16.95	0.60Å HG with A 6 AVAL HG23	Favored (24.71%) General / -92.3,112.7	Favored (19.1%) *tp* chi angles: 166.8,62.9	0.08Å	-	-	-
A 5		HIS	14.11	-	Favored (33.68%) General / -97.9,118.1	Favored (93.9%) *m-70* chi angles: 305,292.4	0.07Å	-	-	-
A 6	A	VAL	12.56	0.60Å HG23 with A 4 LEU HG	Favored (55.5%) Ile or Val / -124.9,119.2	Favored (66.9%) *t* chi angles: 179.1	0.04Å	-	-	-
A 6	B	VAL	12.56	0.42Å HG11 with A 29 LEU HD13	Favored (58.35%) Ile or Val / -127.0,121.6	Favored (3.7%) *p* chi angles: 55.8	0.10Å	-	-	-
A 7		ARG	19.79	-	Favored (25.61%) General / -100.0,113.2	Favored (31.3%) *ttm170* chi angles: 184.1,182.9,292.4,138.7	0.01Å	-	-	-
A 8		VAL	10.57	-	Favored (7.78%) Ile or Val / -84.9,100.1	Favored (34%) *t* chi angles: 184.9	0.05Å	-	-	-
A 9		VAL	9.48	-	Allowed (1.43%) Ile or Val / -82.0,-62.5	Favored (47.9%) *t* chi angles: 181.9	0.08Å	-	-	-
A 10		GLU	10.99	-	Favored (12.69%) General / -168.2,166.8	Favored (12.1%) *pt0* chi angles: 66.1,183.2,322.8	0.07Å	-	-	-
A 11		ALA	10.12	-	Favored (29.83%) General / -137.9,164.1	-	0.04Å	-	-	-
A 12		LYS	27.24	-	Favored (37.14%) General / -149.7,163.7	Favored (52.5%) *pttt* chi angles: 66.5,184.4,169.9,179.4	0.04Å	-	-	-
A 13		GLU	19.97	-	Favored (8.02%) General / 65.1,35.2	Favored (3.4%) *tp30* chi angles: 210.5,59.7,53.4	0.14Å	-	-	-
A 14		LEU	12.18	-	Favored (86.01%) Pre-Pro / -75.2,155.7	Favored (8.6%) *mp* chi angles: 272.2,60.1	0.04Å	-	-	-
A 15		PRO	12.17	0.65Å O with A 23 BCYS SG	Favored (13.32%) Trans-Pro / -83.5,151.1	Favored (57.6%) *Cg_endo* chi angles: 32.2,321.7,28.7	0.10Å	-	-	-

Validation in the Phenix GUI



Validation in the Phenix GUI



PUTTING IT ALL TOGETHER

Part 3

Useful links

- For the quick-and-dirty webpage version of this material:
  - http://molprobity.biochem.duke.edu/help/validation_options/validation_options.html
  - This also includes links to many of the relevant publications

- If you want to see more specific target values for all the metrics:
  - http://molprobity.biochem.duke.edu/help/validation_options/summary_table_guide.html

When do you stop?

- Realistically? Do as much as you can.
  - Ideally stop when you – and refinement – can't make the structure better
- Zero outliers is not the goal!
  - Some outliers are justified
  - Some outliers are not justified, but can't be fixed
- If you can't obtain a physically-reasonable solution, consider deleting the region.

Intervention Philosophy

Refinement: good at optimizing details within a local minimum

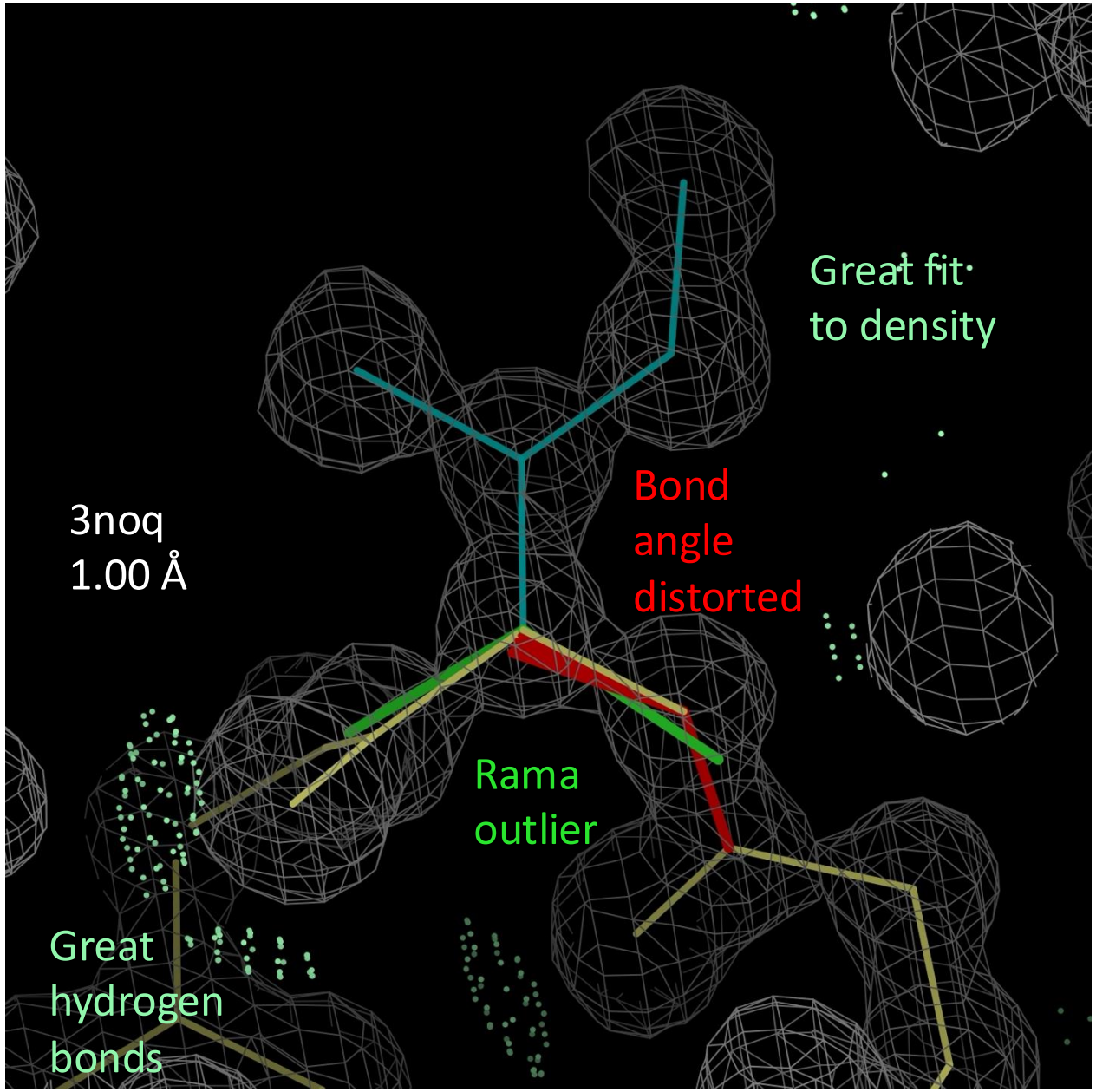
Refinement: bad at escaping wrong local minima

Human: find the right minimum, let refinement optimize

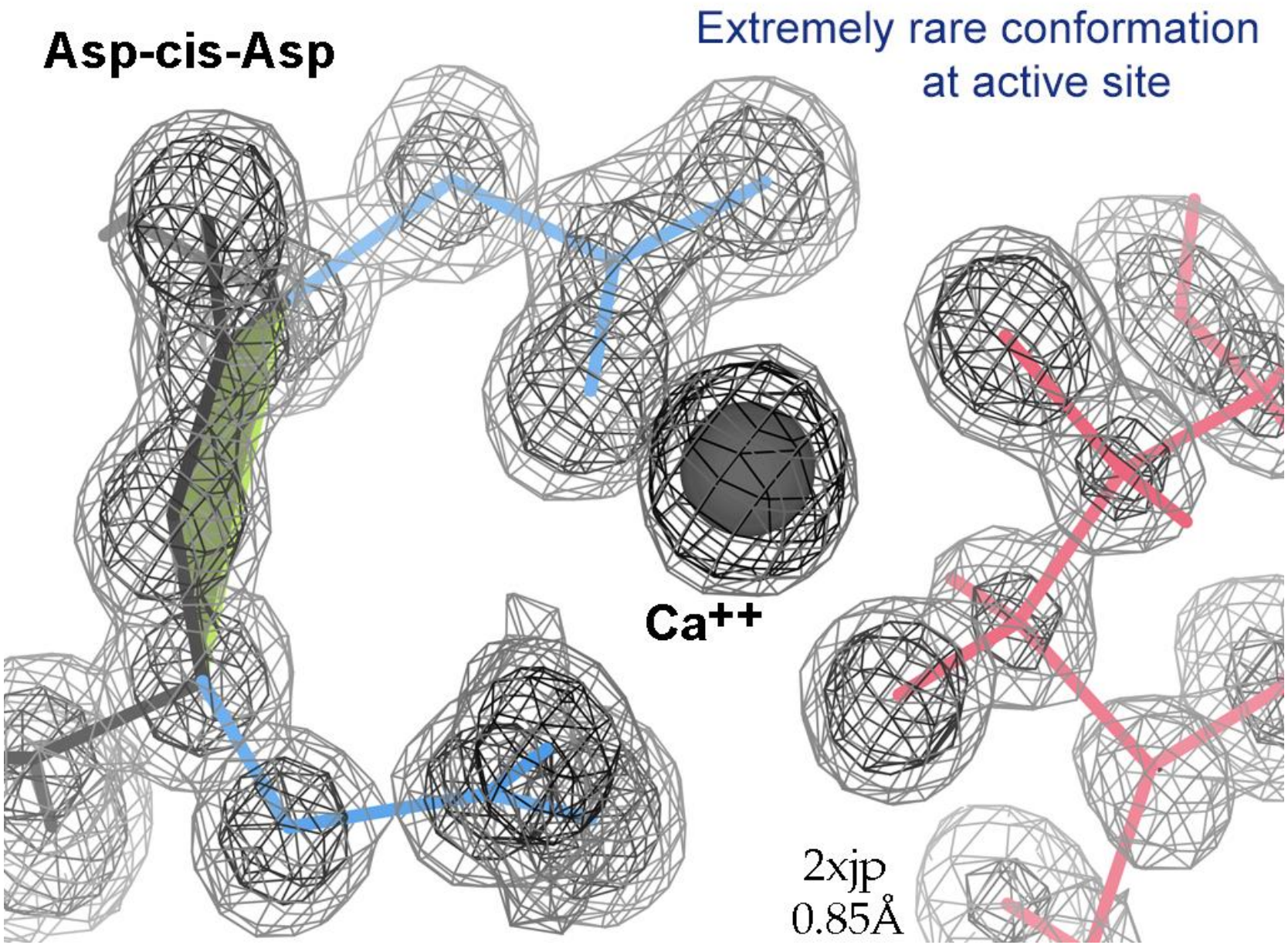
Workflow: validation → manual fix → refinement → repeat

Outliers are not always wrong

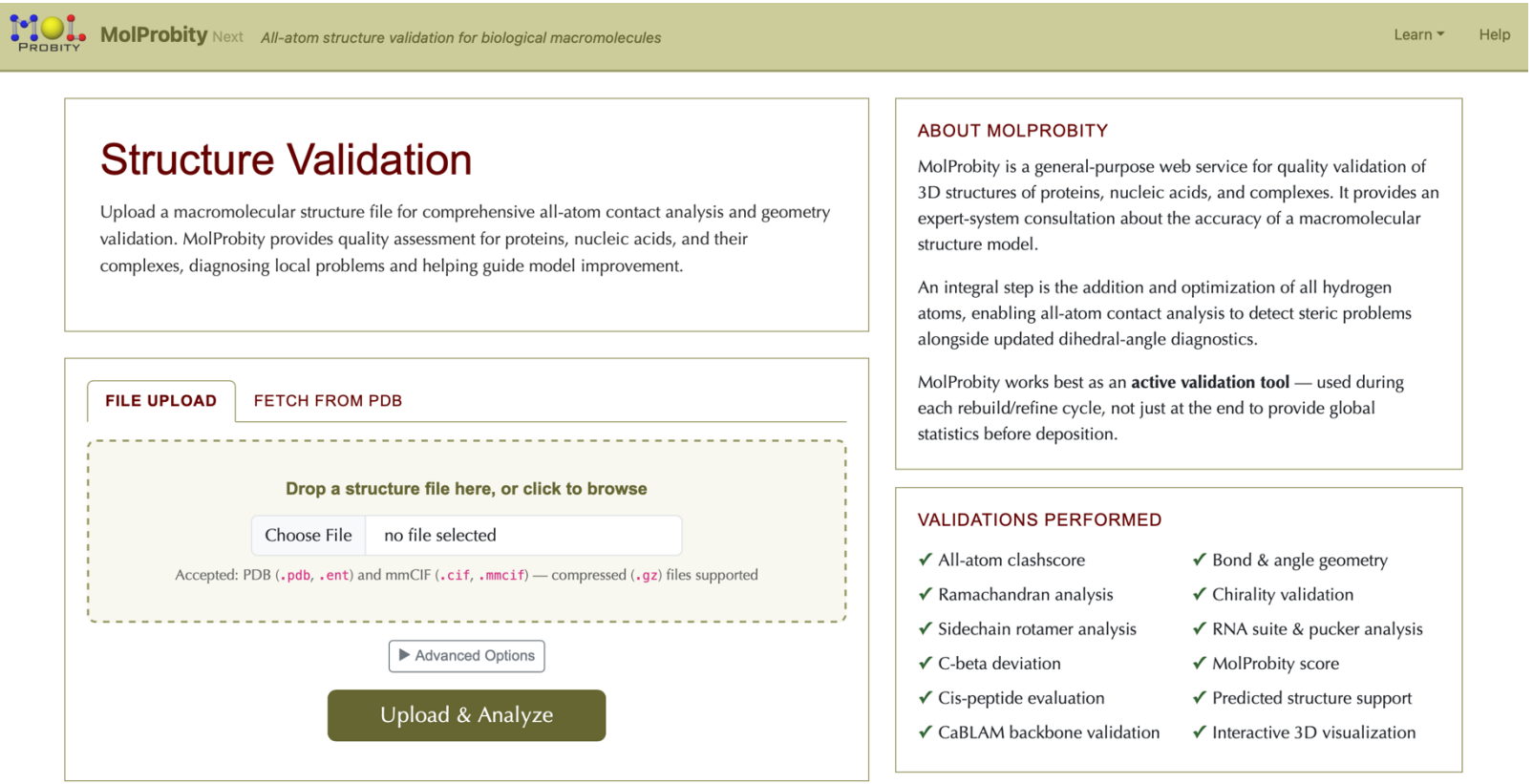
- Some outliers are real and biologically meaningful
- Ramachandran outlier in active site
- Key: outliers need justification (density support, functional role)
- Don't "fix" what's actually correct!



Outliers are not always wrong



Changes are coming….look for MolProbity Next soon



Changes are coming….look for MolProbity Next soon

MolProbity Results for 6m4j.cif.gz

Downloads | Calculated at: 2026-05-05 13:49:27

Overview | Residue Details

mmCIF Header Info

Title: SspA in complex with cysteine
Experiment: X-RAY DIFFRACTION
Resolution: 1.80 Å
Program: REFMAC
Deposition Date: 2020-03-07
Space Group: P 63 2 2 (No. 182)
R-Free: 0.1902
R-Work: 0.1657

Structure Statistics

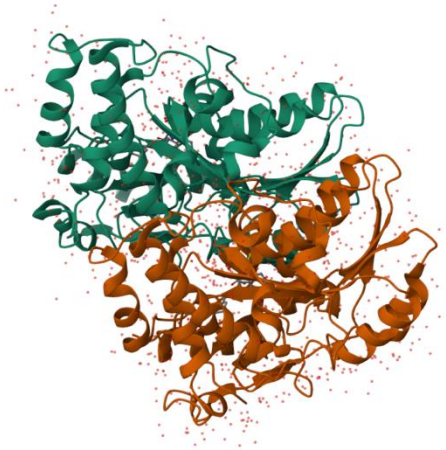
Total Chains: 6
Total Atoms: 6293
Hydrogen Atoms: N/A
Total Residues: 694
Water Molecules: 863

3D Structure Viewer — 6m4j

Mol* | Moorhen | Kin

Outliers: Clashes | Rama | Rotamer | Cβ | Cis non-Pro | Twisted | CaBLAM | CA Geom | Bonds | Angles | Hotspots | Show All

Repr: Show Ball & Stick



Summary Statistics

Metric	Count	Result	Goal
All-Atom Contacts			
Clashscore (?)	17.33	35th %ile	

Changes are coming….look for MolProbity Next soon

Overview | Residue Details

Compact View

Click a row to zoom the 3D viewer

Open Full View →

Show Outliers Only Show All Metrics Showing all 1731 residues

25 entries per page

Search:

	Residue	MP-Triage (?)	Issues
Key: Severe Outlier Minor			
+	A 259 SER	19.7	Clash, Cβ, Twist, CA Geom, Bond, Angle, Hot#1
+	B 264 GLU	19.2	Clash, Rota, Twist, Angle, Hot#3
+	A 260 ALA	18.0	Clash, Rama, Twist, CA Geom, Angle, Hot#1
+	B 190 MET.B	15.4	Clash, Rota, Hot#2
+	B 159 GLU	14.6	Clash, Hot#2
+	B 257 GLU	13.5	Clash, Rama, Cis, CaBLAM, Angle, Missing, Hot#4
+	A 265 PHE	13.1	Rama, Cβ, Cis, CA Geom