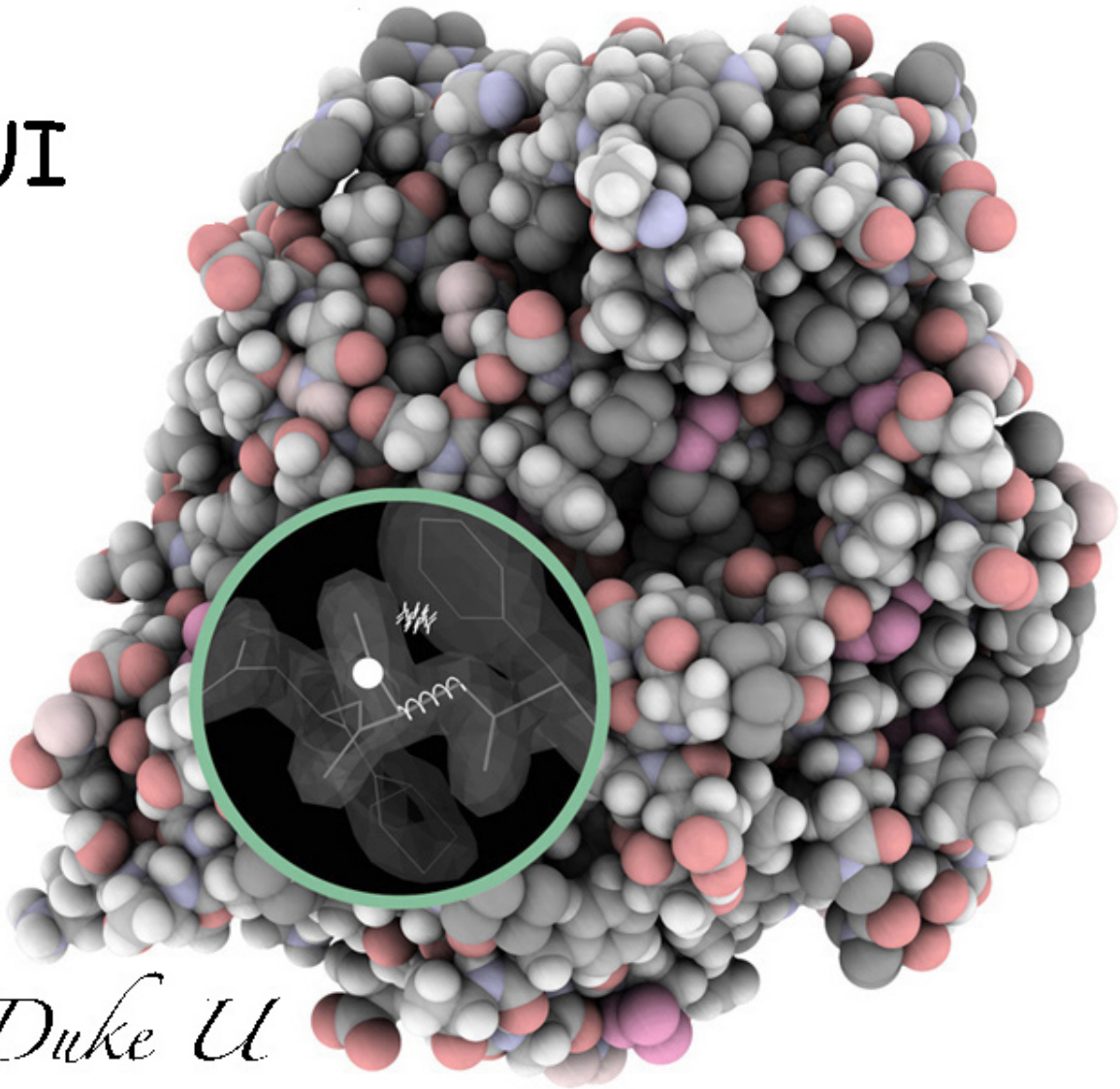
Molecular Therapy for 2.5-4Å Models:
CryoEM (or Xray)

1. Validation GUI in Phenix
2. What is its basis ?
3. How do I use it ?



Jane Richardson, Duke U

PHENIX home

Quit | Preferences | Help | Citations | Coot | PyMOL | KiNG | Other tools | Ask for help

Actions | Job history

Projects

Show group: All groups | Manage...

Select | Delete | New project | Settings

ID	Last modified	# of jobs	R-free
3ja8	Jul 19 2019 04:37 ...	1	---
jsr	Jul 19 2019 04:40 ...	3	---
cryoem2019	Jun 10 2019 02:59...	1	---
5hut	Jun 10 2019 02:36...	56	0.1875
test	May 29 2019 12:2...	6	---
sec17-sad	Feb 21 2018 12:10...	0	---
5a1atest	Dec 06 2017 04:3...	4	---
rnase-s	Nov 01 2017 03:2...	23	0.2852
1ubq	Sep 15 2017 02:4...	5	0.2198
1pe9_jsr	May 12 2017 01:4...	8	0.1854
1pe9	May 11 2017 12:5...	1	0.1868
4loz	May 09 2017 04:4...	4	---
phenix-test	May 02 2017 09:3...	6	---
4uoj	Mar 14 2017 05:30...	1	---
4mh1	Mar 14 2017 04:10...	6	0.2861
5hj0	Mar 11 2017 03:47...	1	---
4q8j	Feb 16 2017 02:48...	3	---
214l	Jan 27 2017 03:16...	3	---
4xr7	Jan 26 2017 03:11...	4	---
3js8	Sep 14 2016 12:2...	3	---

Favorites

Data analysis

Experimental phasing

Molecular replacement

Model building

Refinement

Cryo-EM

Validation

Comprehensive validation (X-ray/Neutron)
Model quality assessment, including real-space correlation and geometry inspection using MolProbity tools

Comprehensive validation (cryo-EM)
Model quality assessment, including real-space correlation, for cryo-EM structures

Structure comparison
Identify differences between multiple structures of the same protein, using multiple criteria

Calculate CC*
Comparison of unmerged data quality with refined model, as described in Karplus & Diederichs (2012)

EMRinger
Model validation for de novo electron microscopy structures

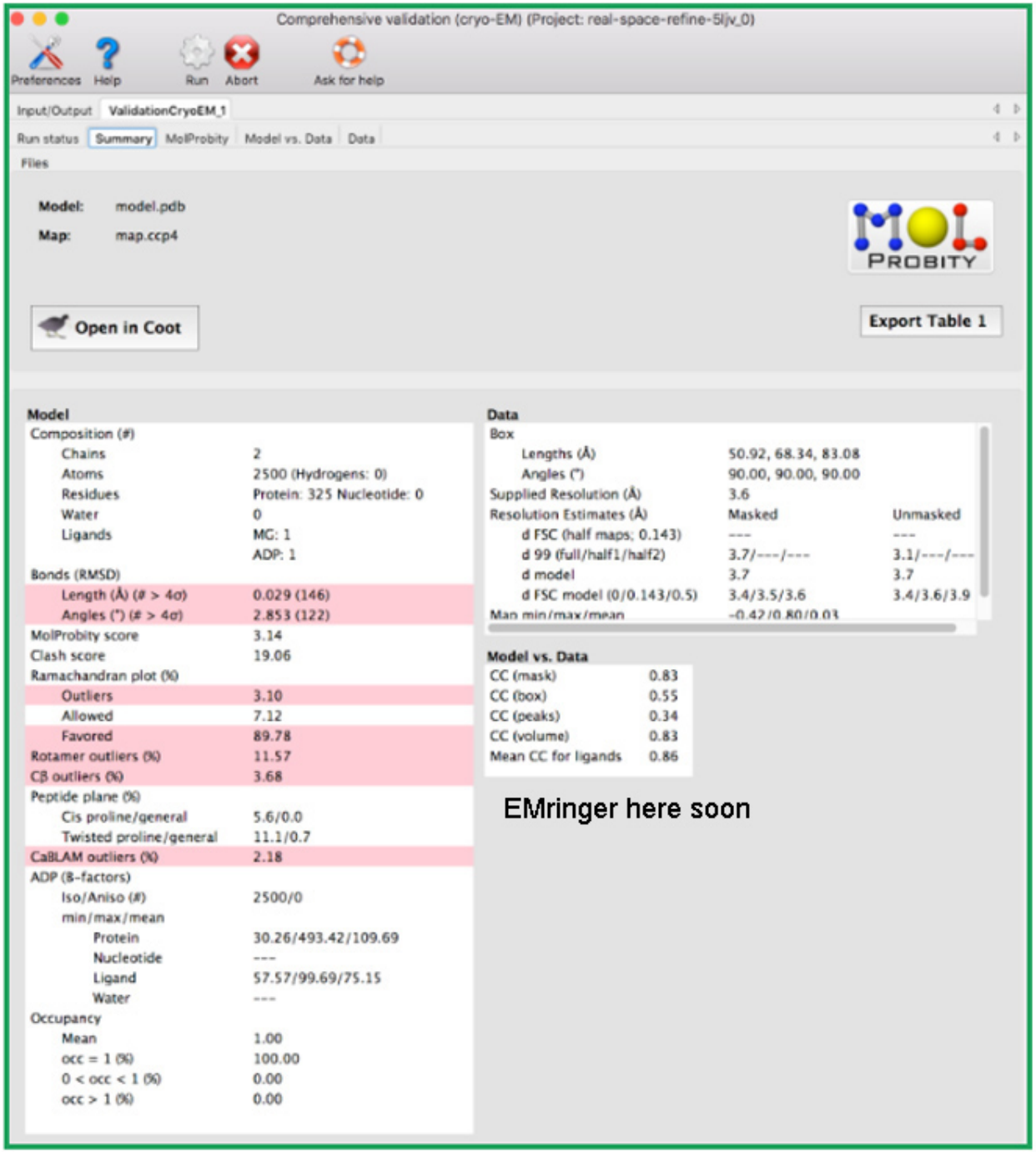
Ligands

Summary page

for cryoEM

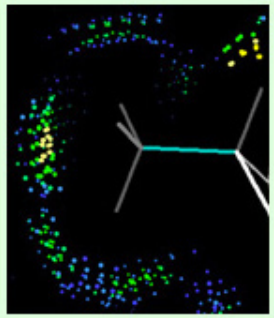
validation:

model, data, & model-vs-data

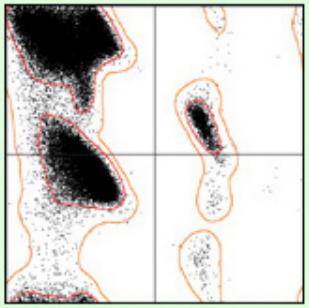


MolProbity Structure Improvement Criteria

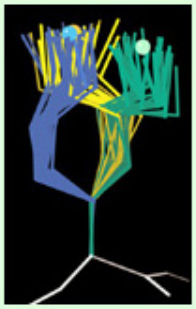
All-atom contacts, clashscore



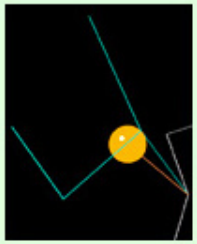
Ramachandran criteria



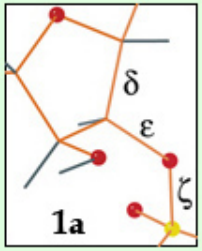
Sidechain rotamers

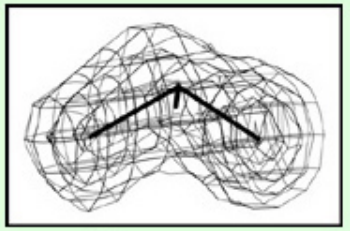


Geometry



RNA bb



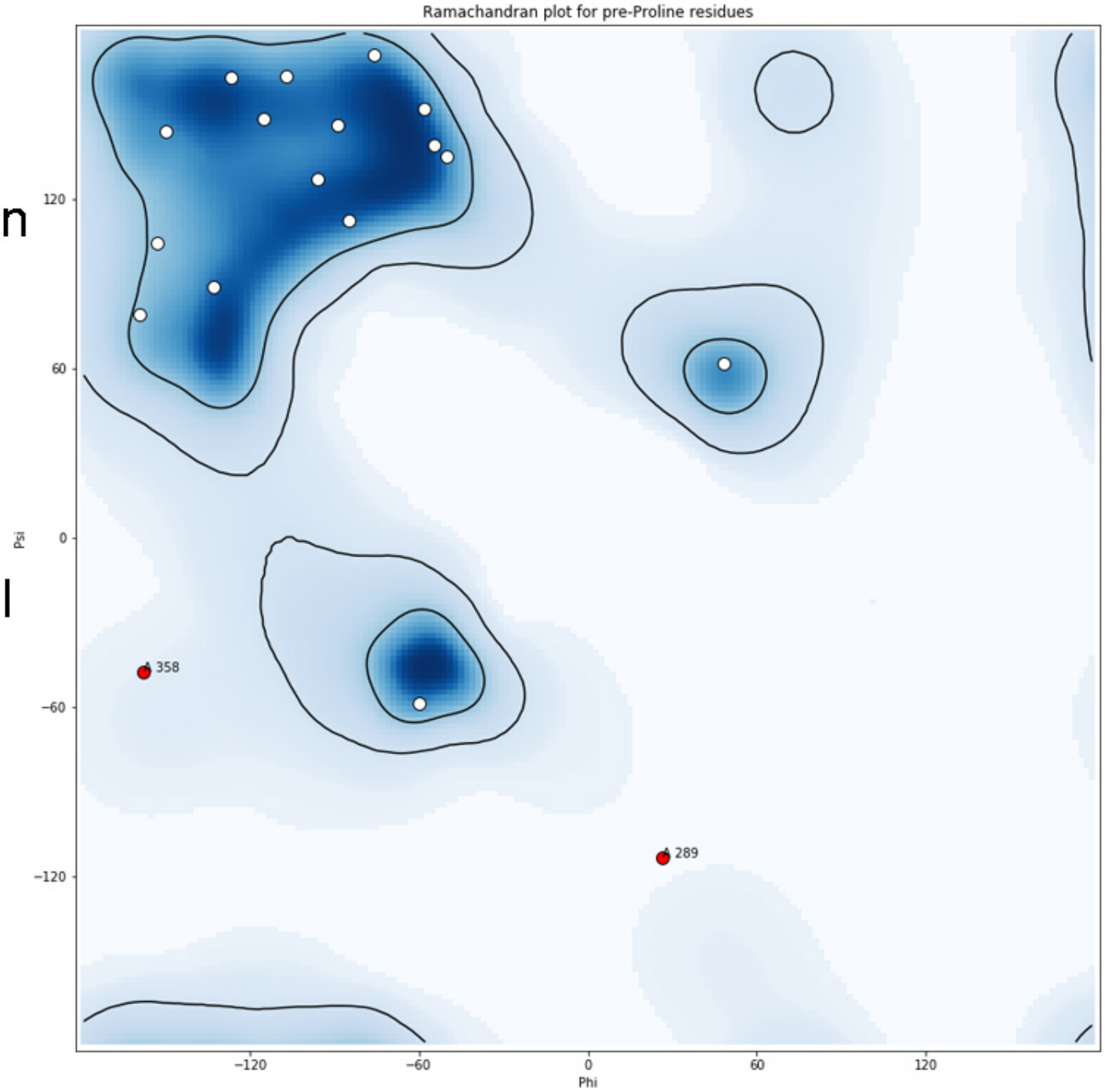


Crystallographic: R_{free}, electron density fit

for low resolution: CaBLAM, *cis*-nonPro

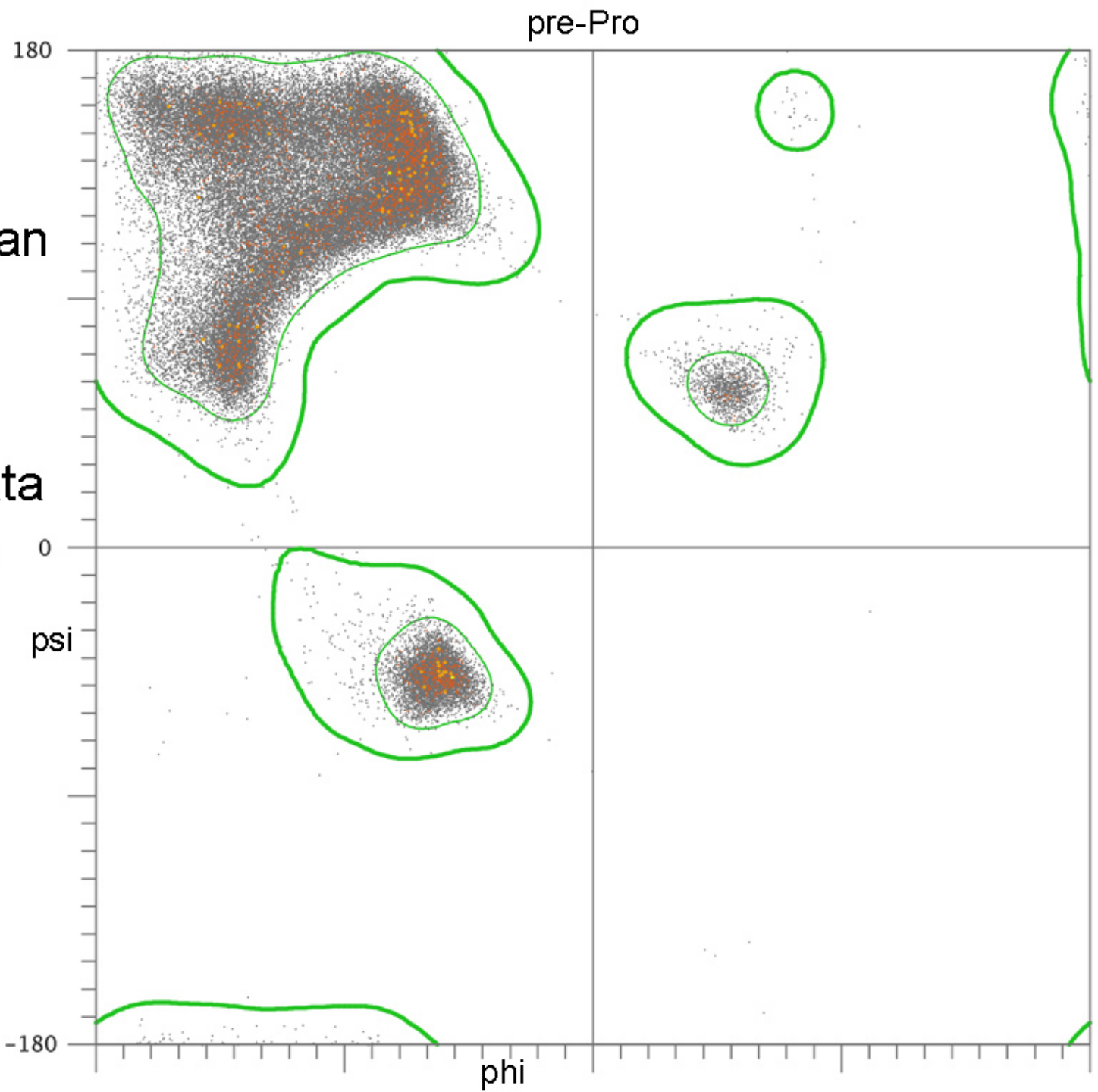
Ramachandran plot --

pre-Pro for query model in Phenix validation GUI



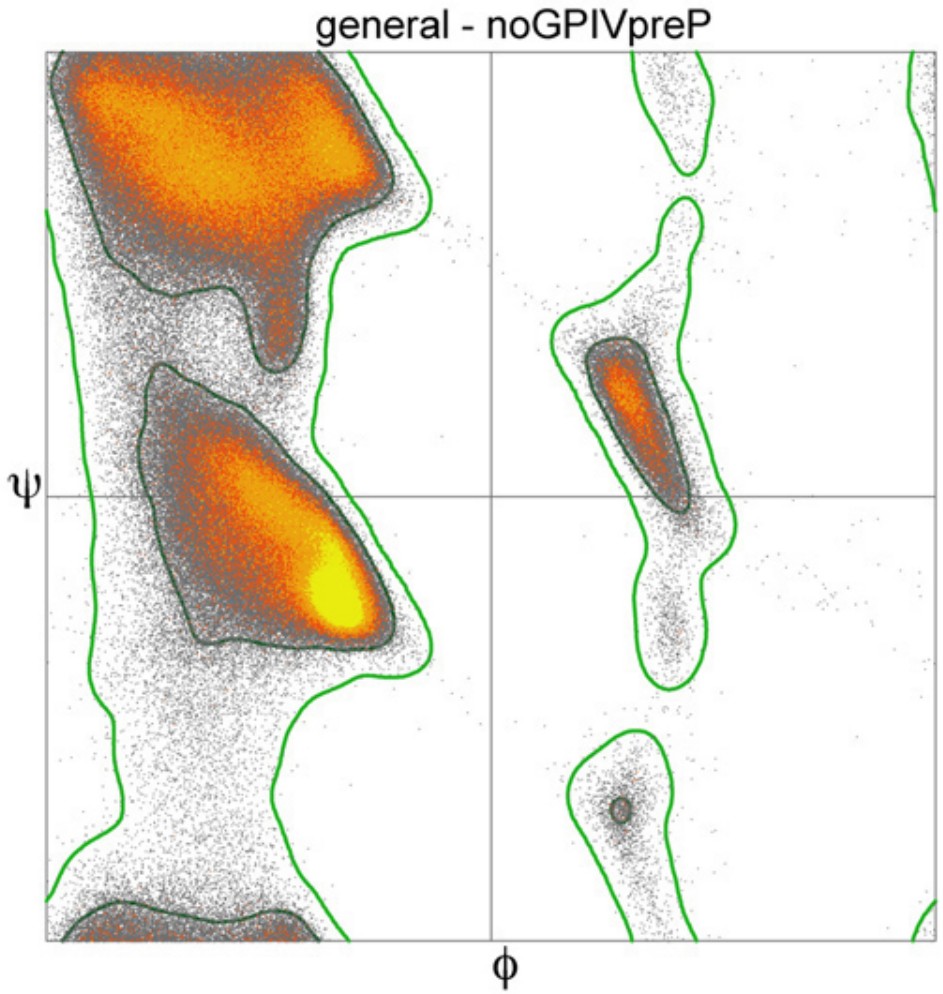
Ramachandran plot --

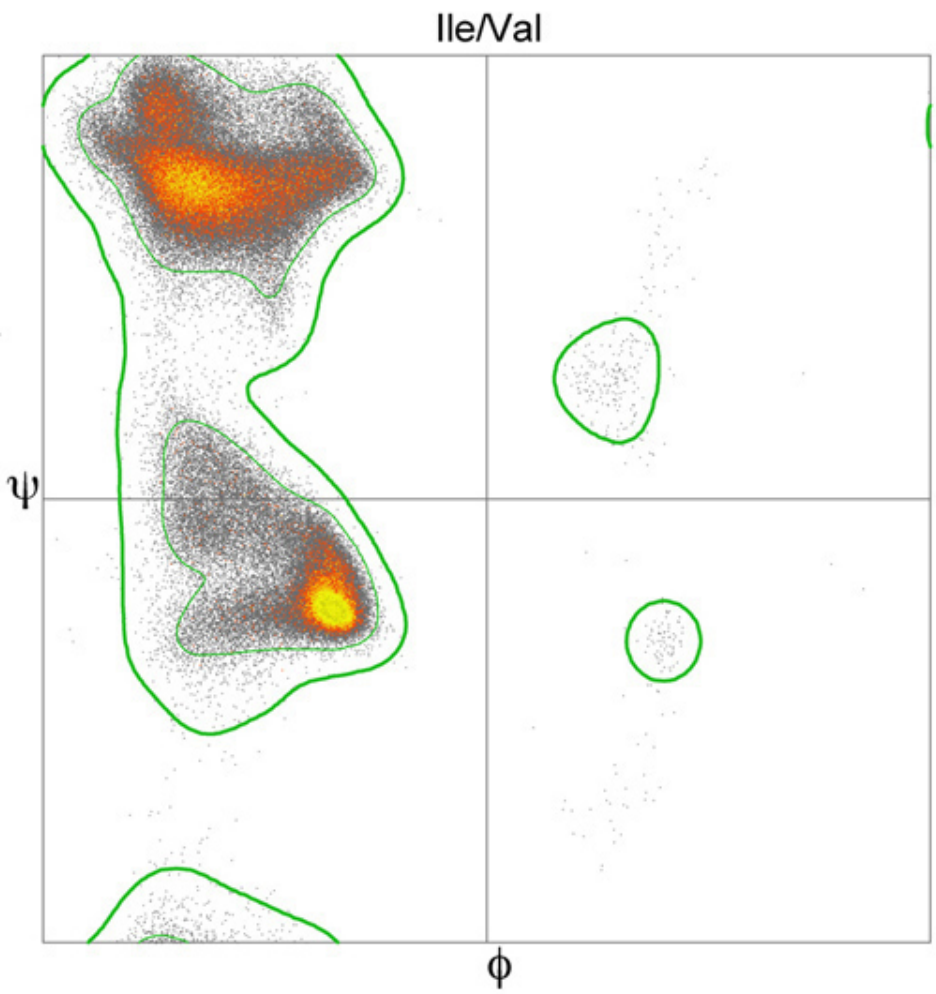
Top8000 reference data for pre-Pro



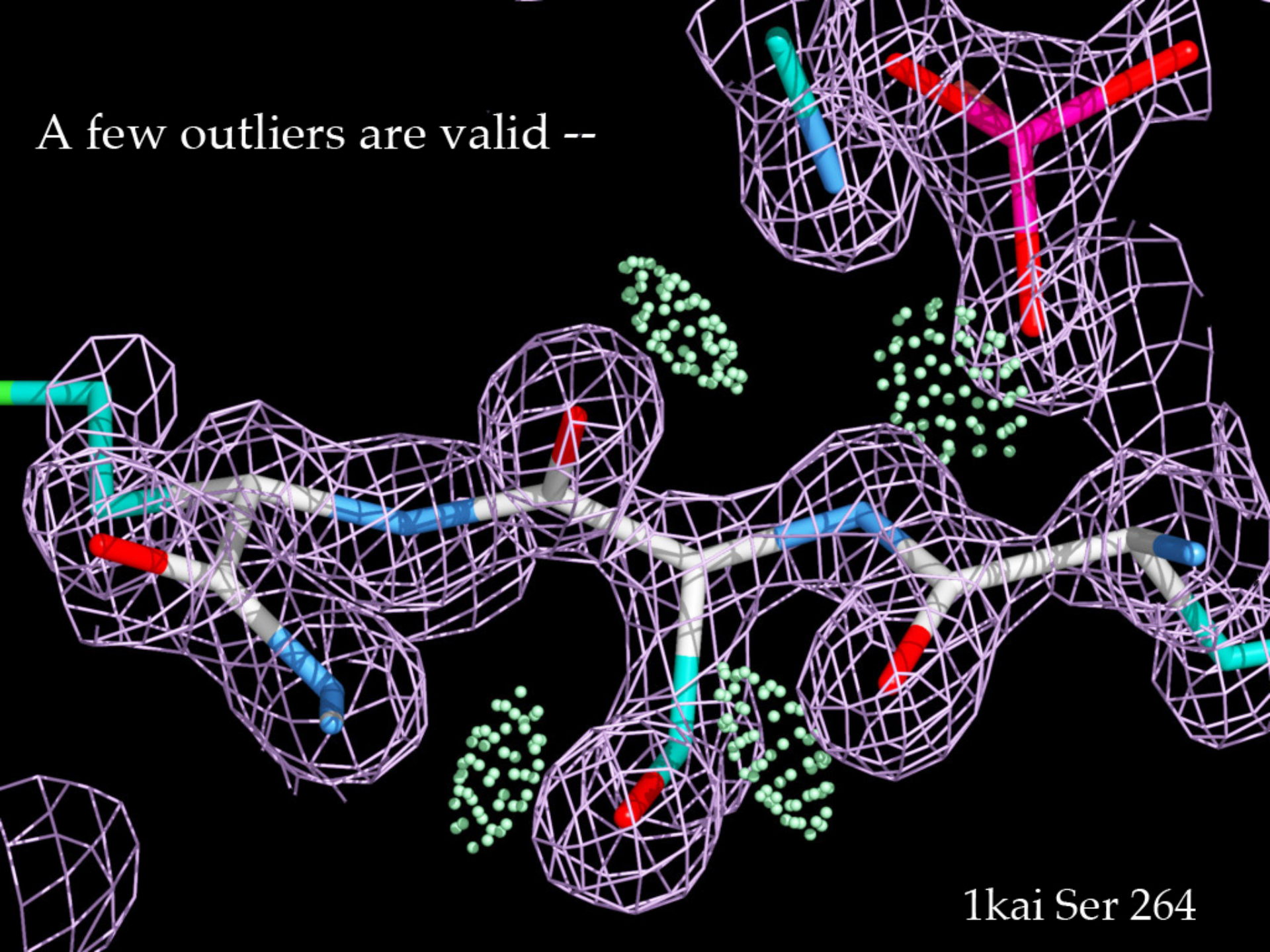
MolProbity, Phenix and wwPDB validation now use 6-category Ramachandran plots

>million residues: <70% homol, <2.0Å, B<30

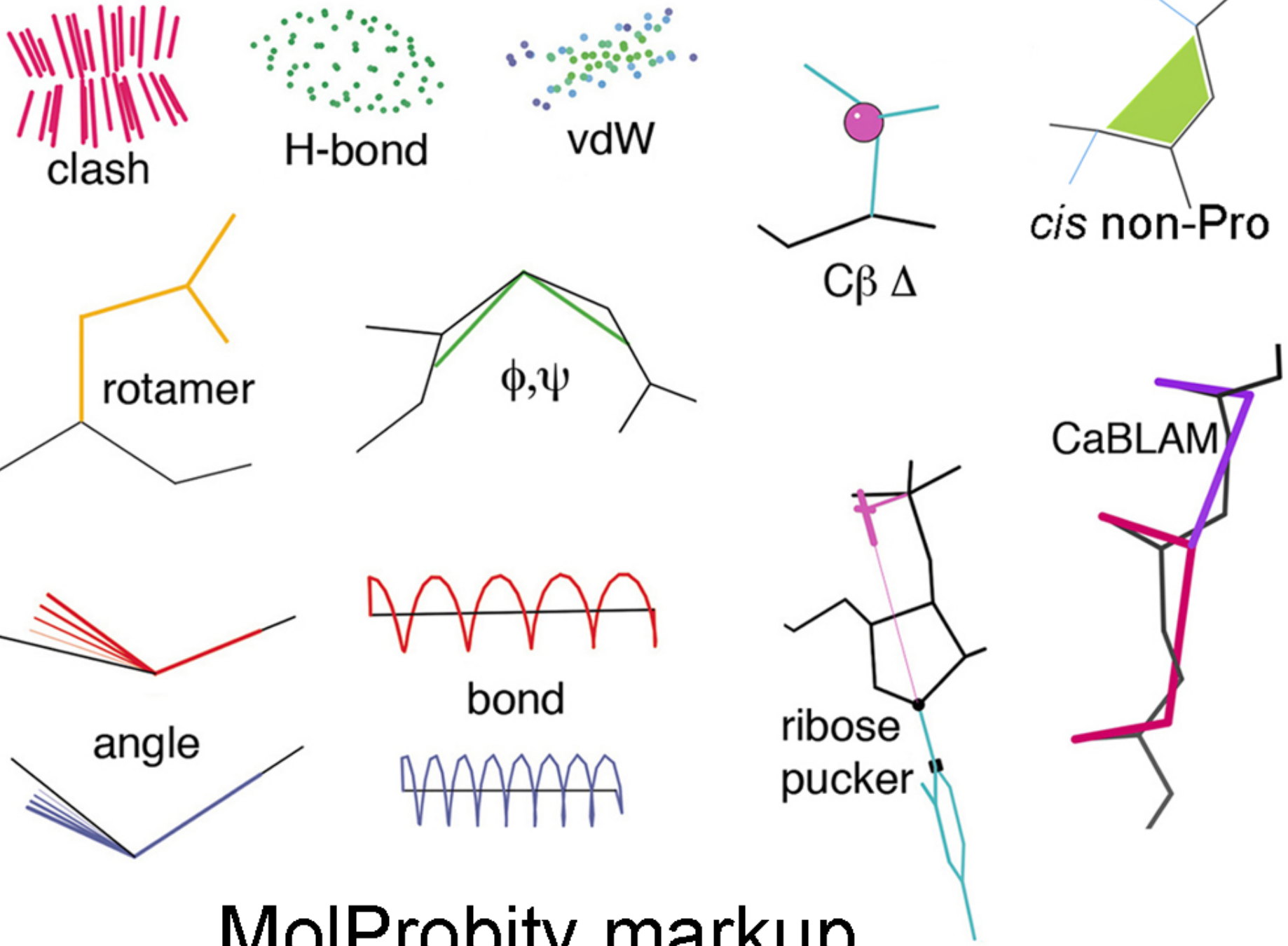


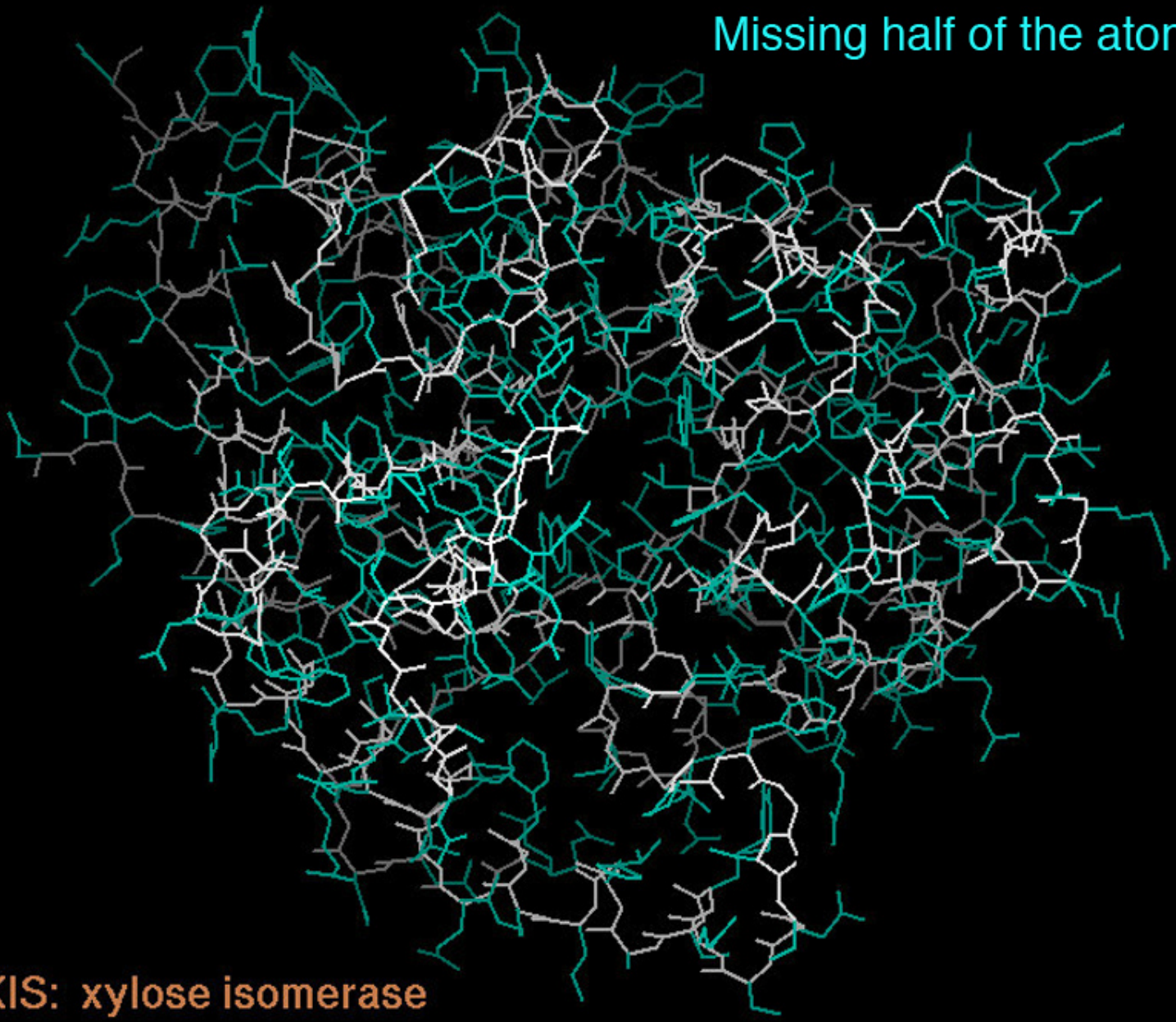


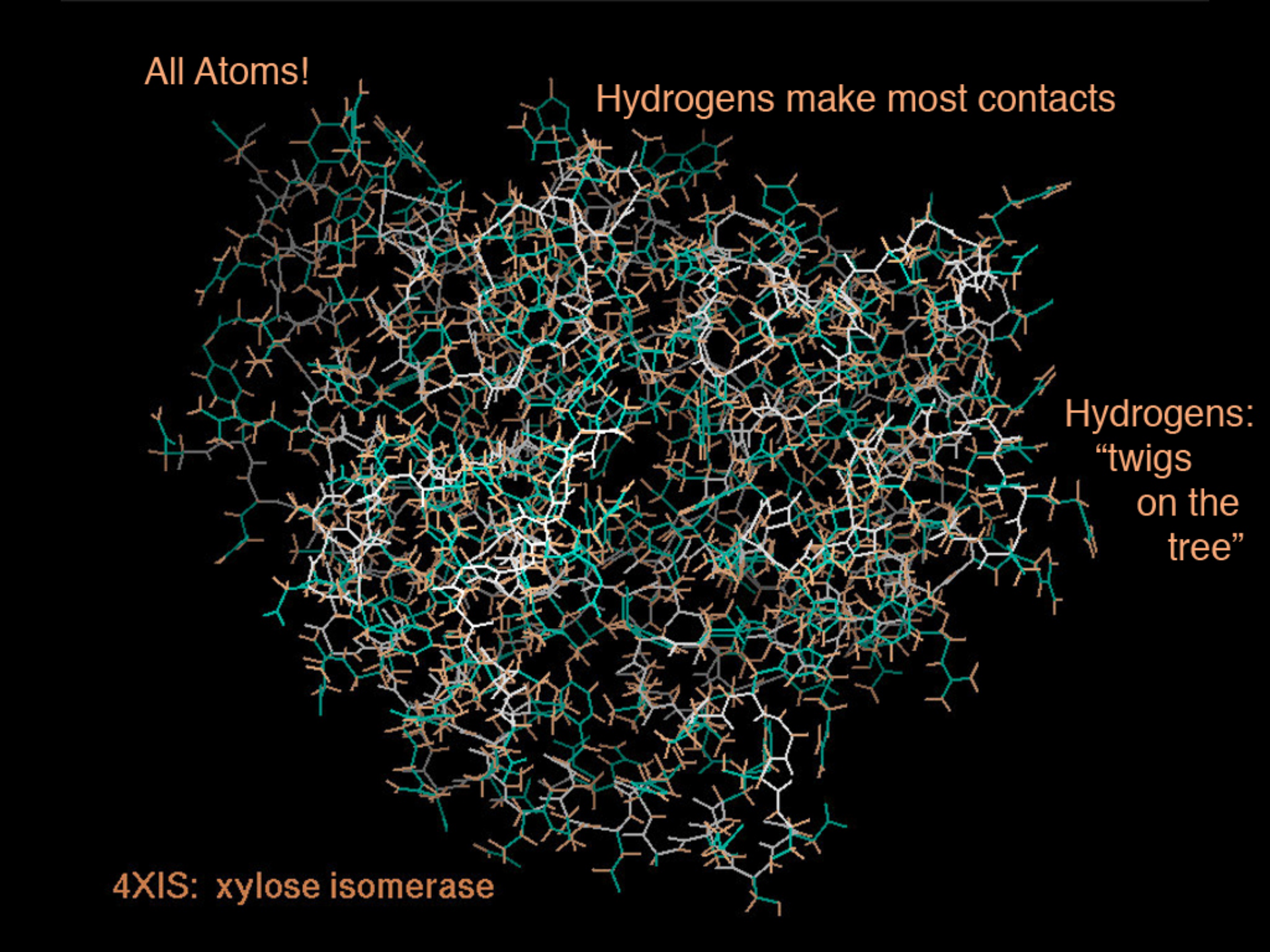
A few outliers are valid --



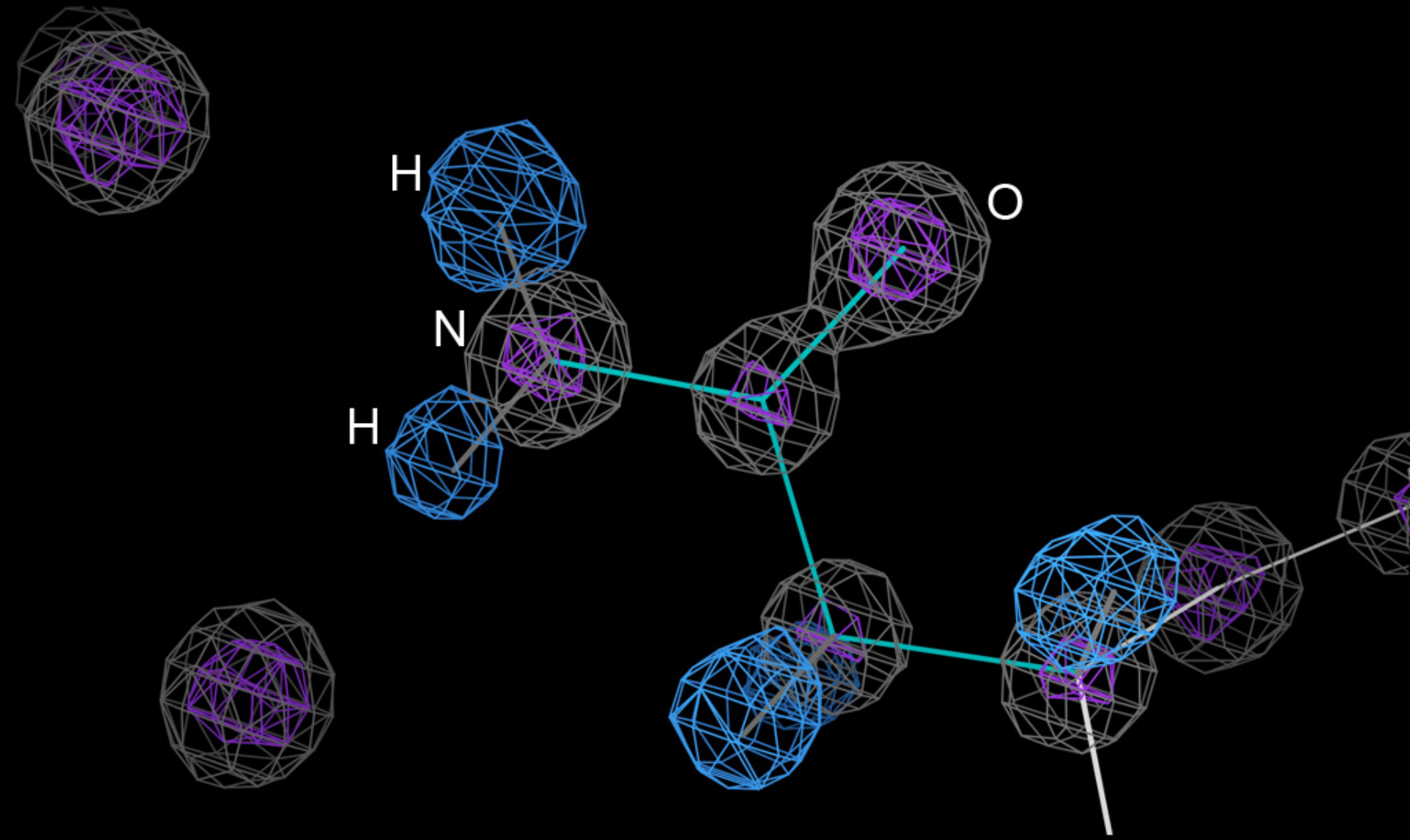
1kai Ser 264

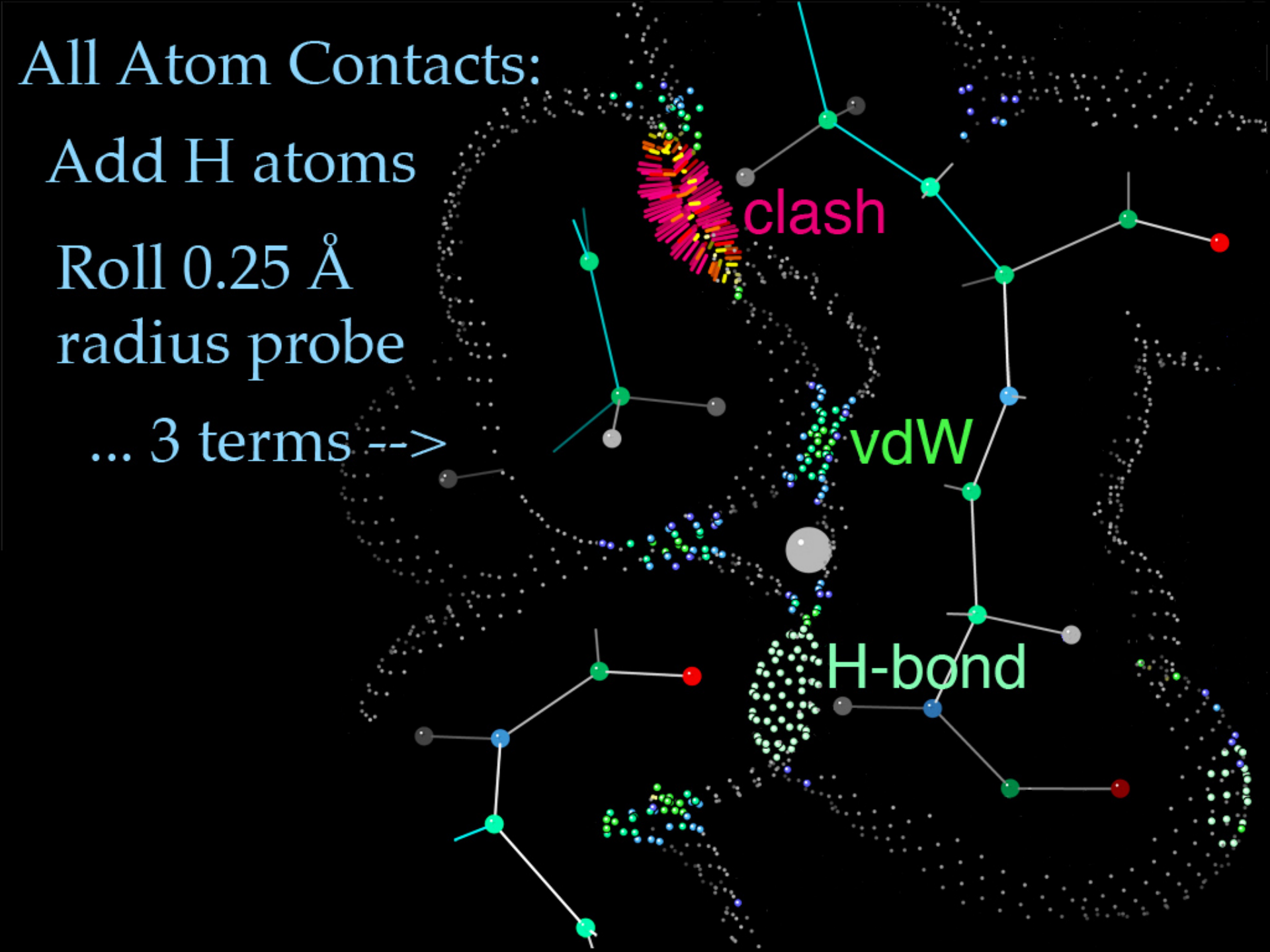

clash
H-bond
vdW
Cβ Δ
cis non-Pro
rotamer
ϕ,ψ
CaBLAM
angle
bond
ribose
pucker
MolProbity markup

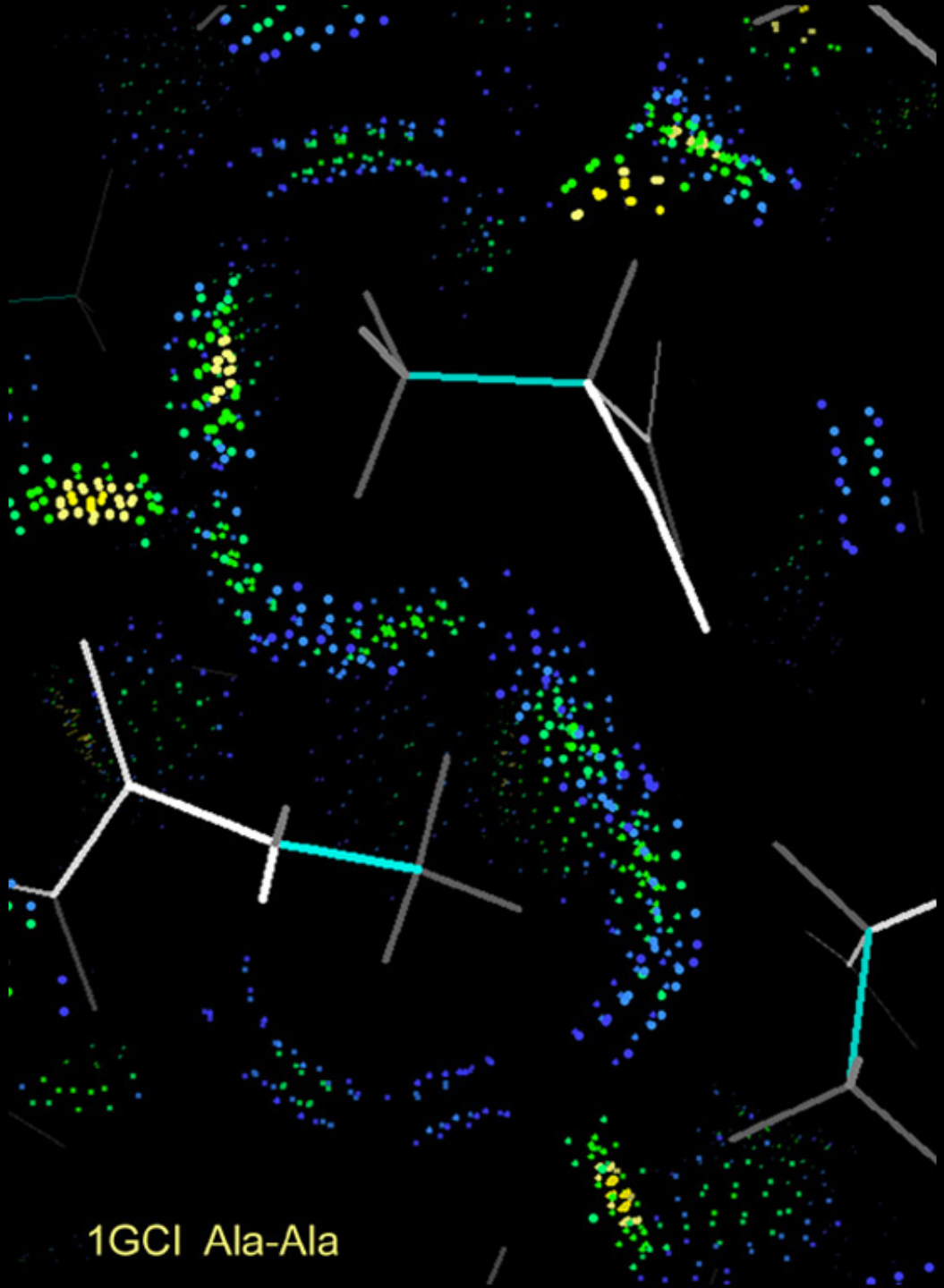

Missing half of the atoms!
4XIS: xylose isomerase

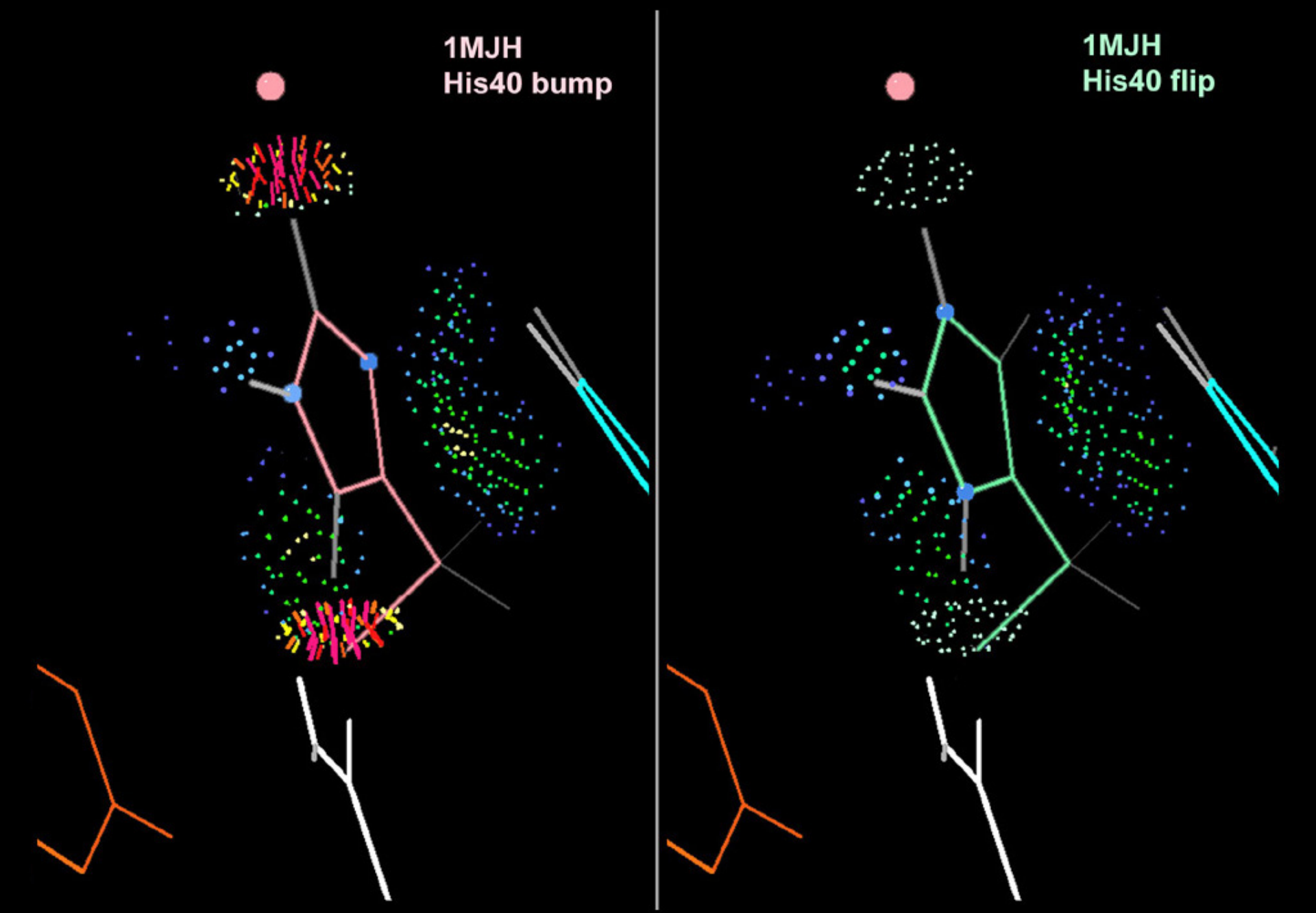

All Atoms!
Hydrogens make most contacts
Hydrogens:
"twigs
on the
tree"
4XIS: xylose isomerase

-- but the H atoms are really there !




All Atom Contacts:
Add H atoms
Roll 0.25 Å
radius probe
... 3 terms -->
clash
vdW
H-bond


1GCI Ala-Ala



His flips change protonation, H-bonds, & even charge

Comprehensive validation (cryo-EM) (Project: jsr)

Preferences Help Run Abort Ask for help

Input/Output | **ValidationCryoEM_3**

Run status | Summary | **MolProbity**

Clashes | CaBLAM | Cβ | Cis/Twisted | Rotamers | Ramachandran | Geometry Restraints

All-atom contact analysis

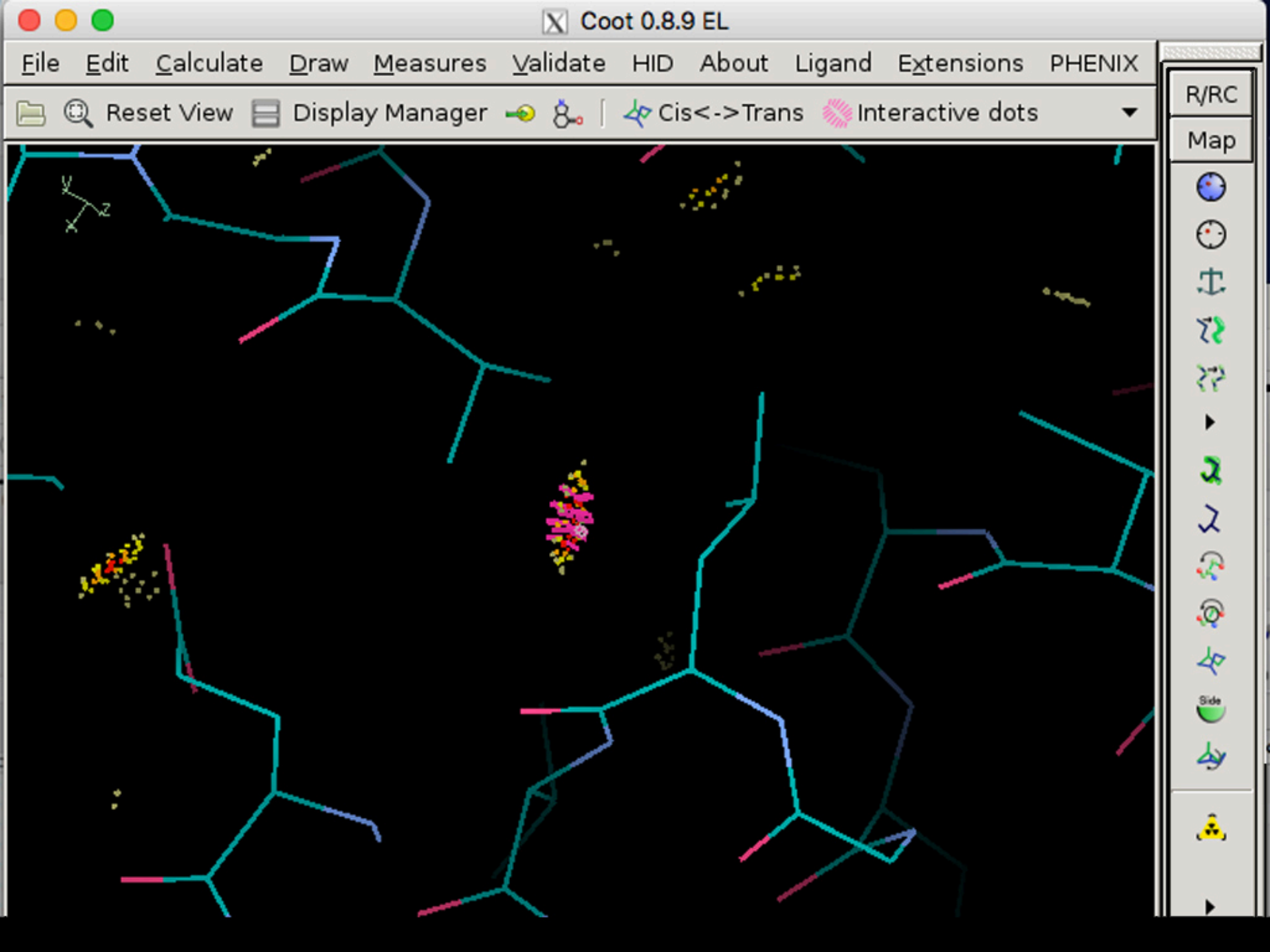
Coot display

☑ Show Probe dots in Coot ☑ Only show bad overlaps Reload data Re-run PROBE

Bad contacts from PROBE: 116 overlapping atom pairs

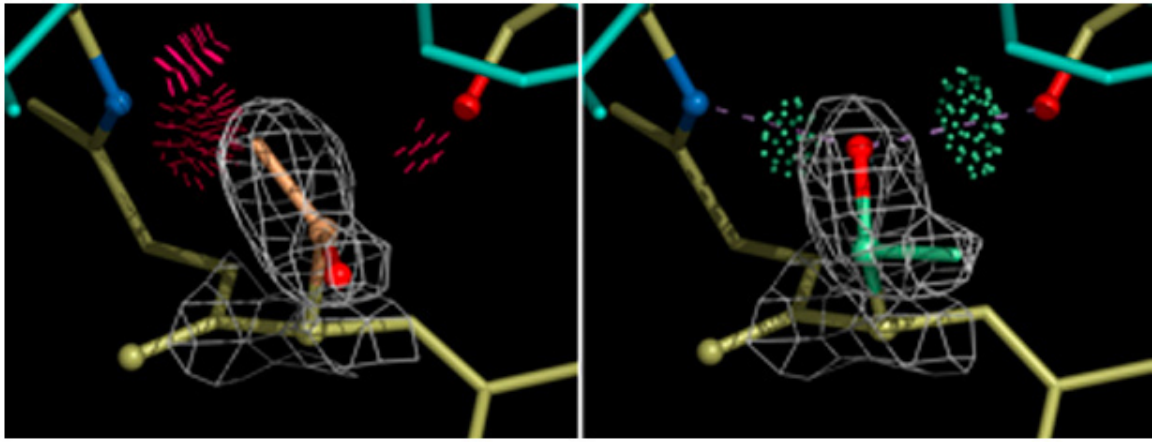
This list summarizes all severe clashes (more than 0.4 Angstrom non-H-bond overlap) found by PROBE; you can view these graphically in Coot. If no hydrogens were present, REDUCE was used to add them prior to running PROBE.

Atom 1	Atom 2	Overlap
C 61 VAL HG11	C 496 LEU HB3	0.598
B 13 GLY HA3	B 14 PRO HA	0.587
D 599 LYS HE3	D 731 ALA HB2	0.586
C 719 PRO HB3	C 817 SER HB2	0.584
D 201 LEU HD11	D 220 ARG HH11	0.579
C 733 GLU HB3	C 823 LEU H	0.57
D 268 LEU HD13	D 306 LEU HA	0.548
D 254 PRO HD3	I 11 C H41	0.545
D 43 THR HG22	D 56 LEU HB2	0.542
A 57 THR HB	A 157 ARG HH12	0.541
A 27 VAL HG22	A 201 VAL HG12	0.539
B 178 PRO O	B 207 ASN ND2	0.536
D 491 LEU HA	D 498 PRO HA	0.536
D 648 PRO HG3	D 701 THR HG21	0.534
C 706 ARG NH2	C1112 GLY O	0.533
C 179 ALA HB3	C 191 PHE HB2	0.528
C 870 ARG NH1	I 9 A OP2	0.528
C 536 SER HB2	G 22 DC H4'	0.525


Coot 0.8.9 EL
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Ligand
Extensions
PHENIX
Reset View
Display Manager
Cis<->Trans
Interactive dots
R/RC
Map
Side

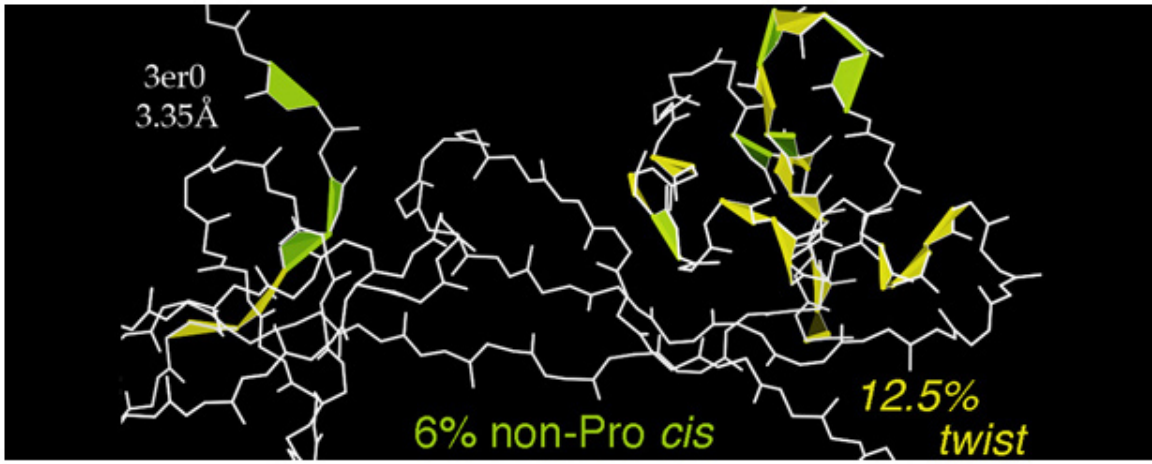
Common, systematic errors

Asn/Gln/His flips

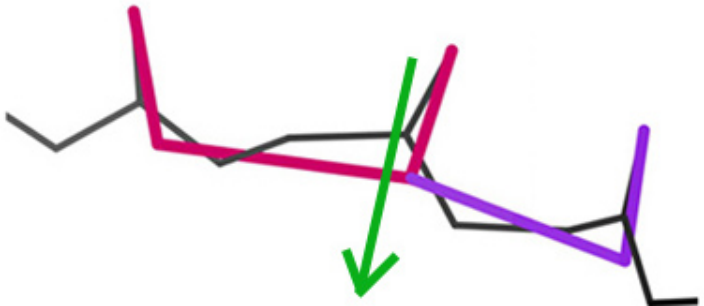


Backward Cβ-branch sidechains

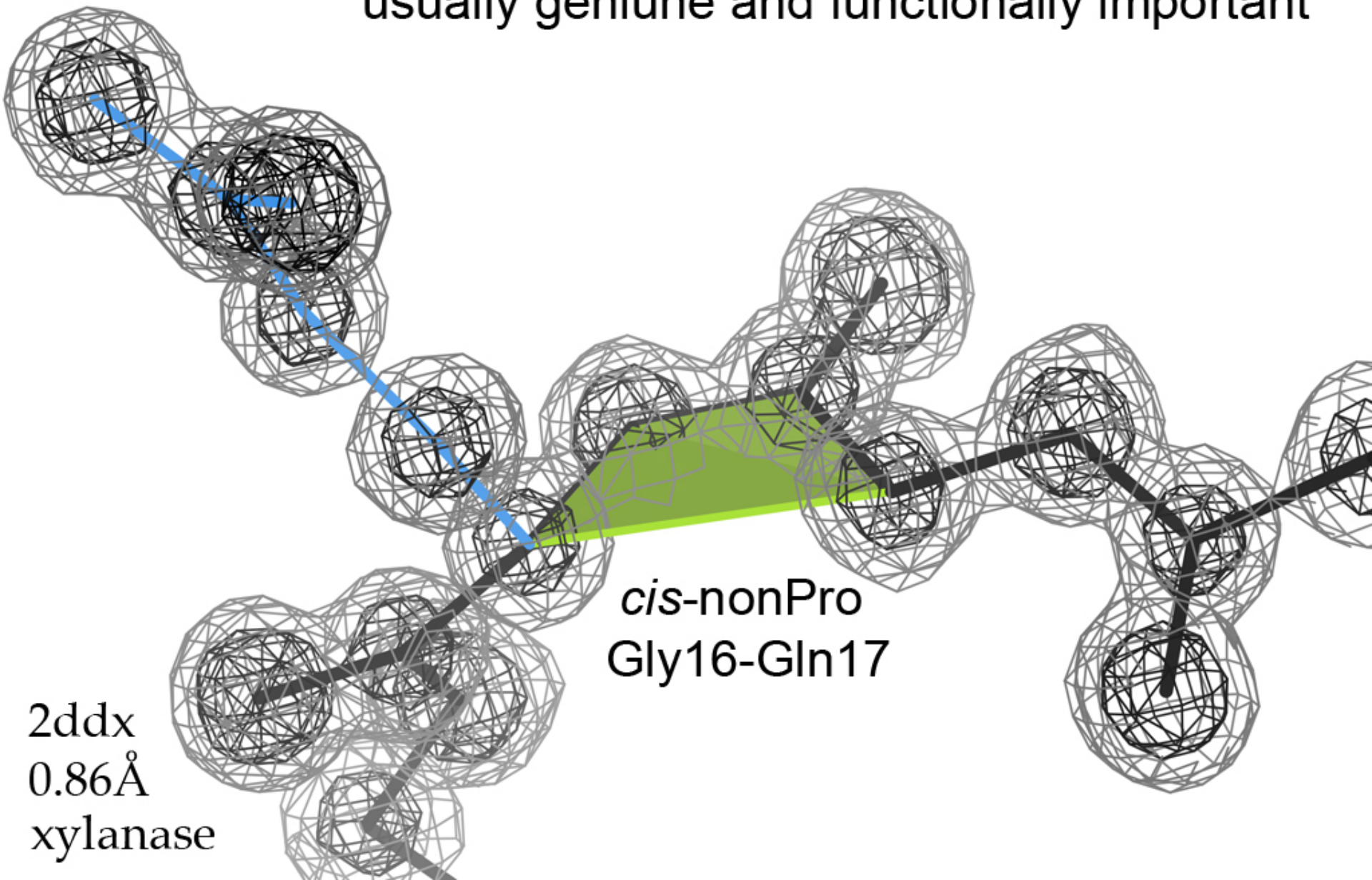
Overused *cis*-nonPro

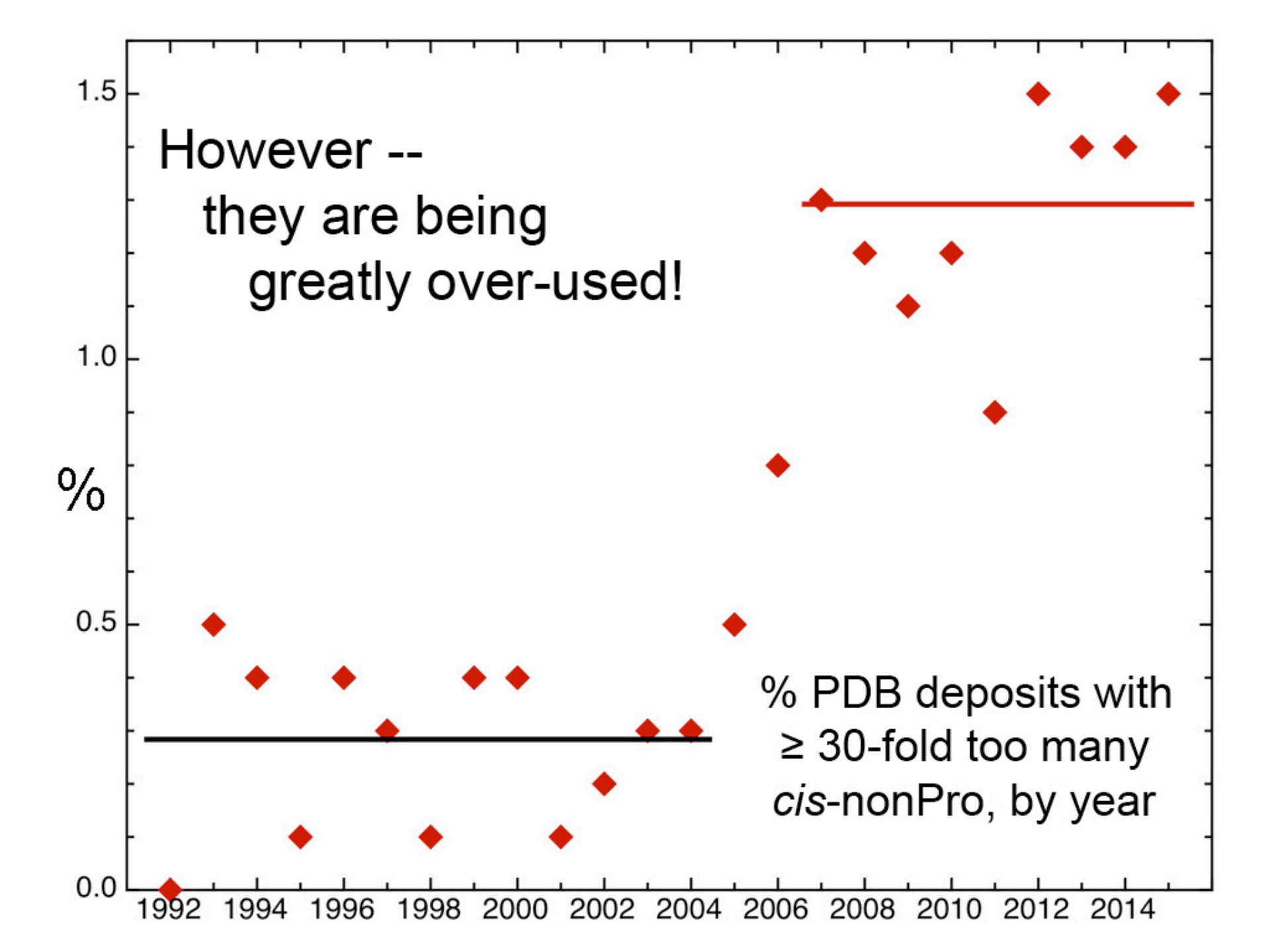


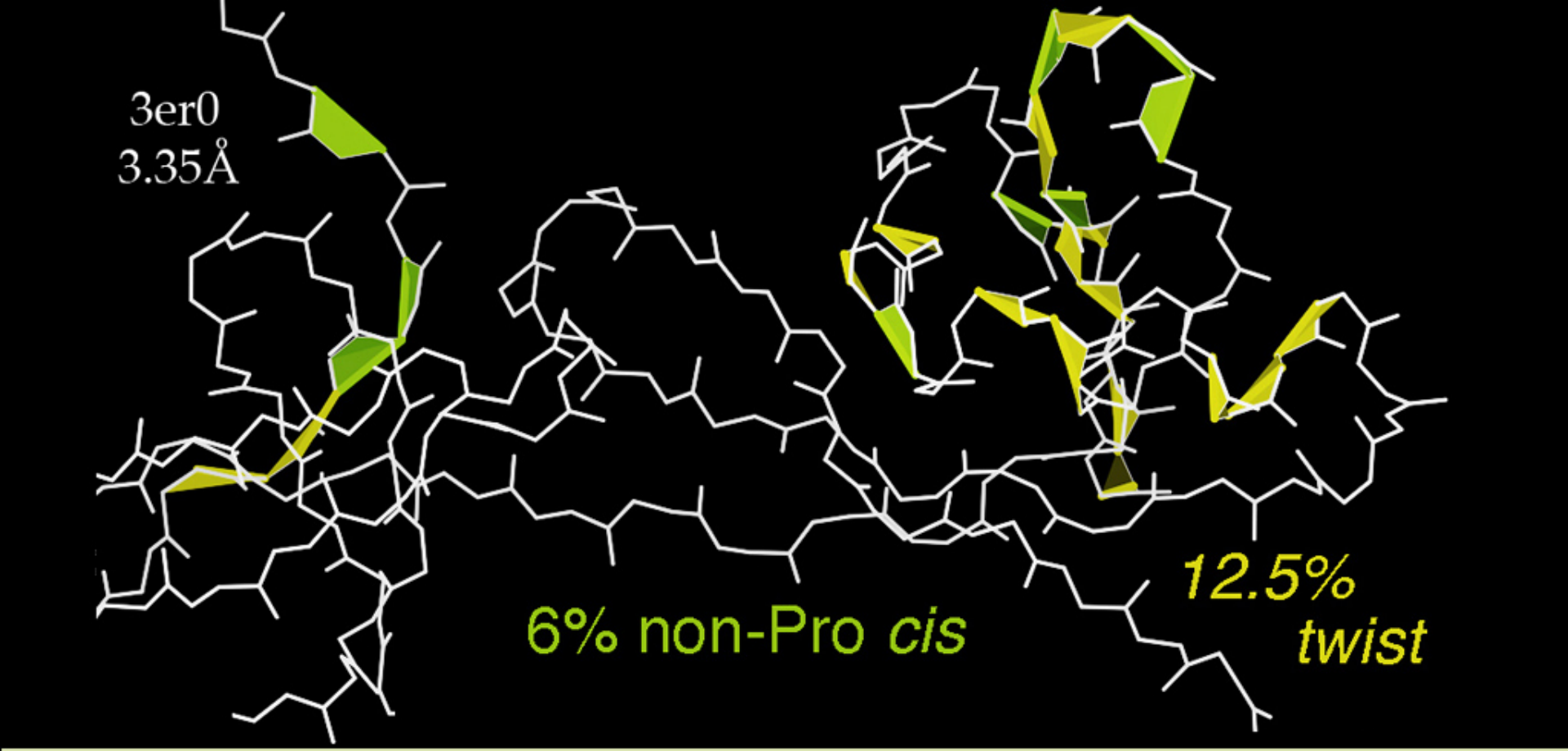
Bad CO orientation



cis-nonPro peptides are very rare (~0.03%),
usually geniune and functionally important



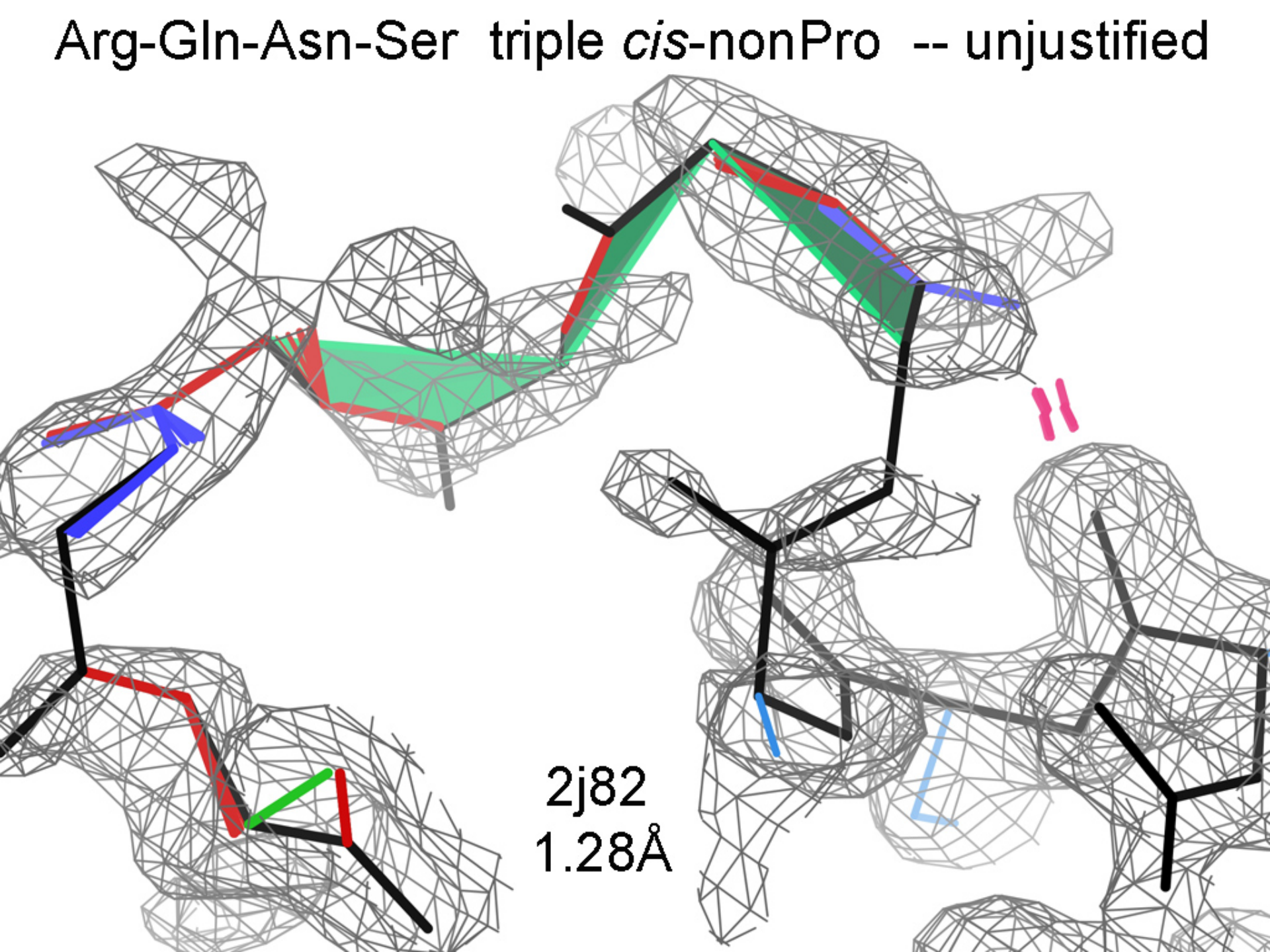

However --
they are being
greatly over-used!
%
1.5
1.0
0.5
0.0
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
% PDB deposits with
≥ 30-fold too many
cis-nonPro, by year



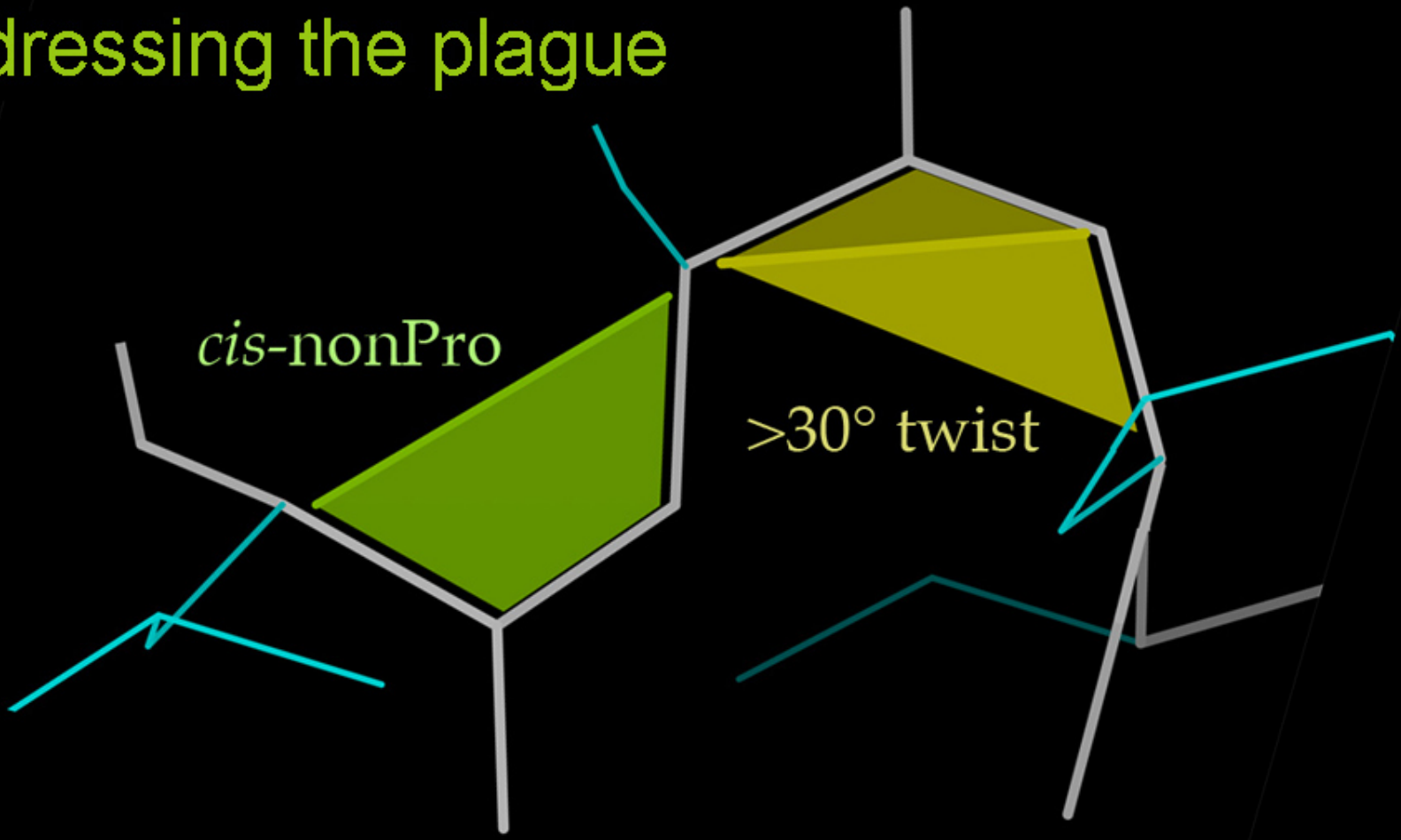
Conformations usually chosen one-at-a-time,
with no weighting by probability

Dangerous at low resolution or high B-factor

Even systematic bias: *cis* often stays better inside
shrunken, low-resolution density

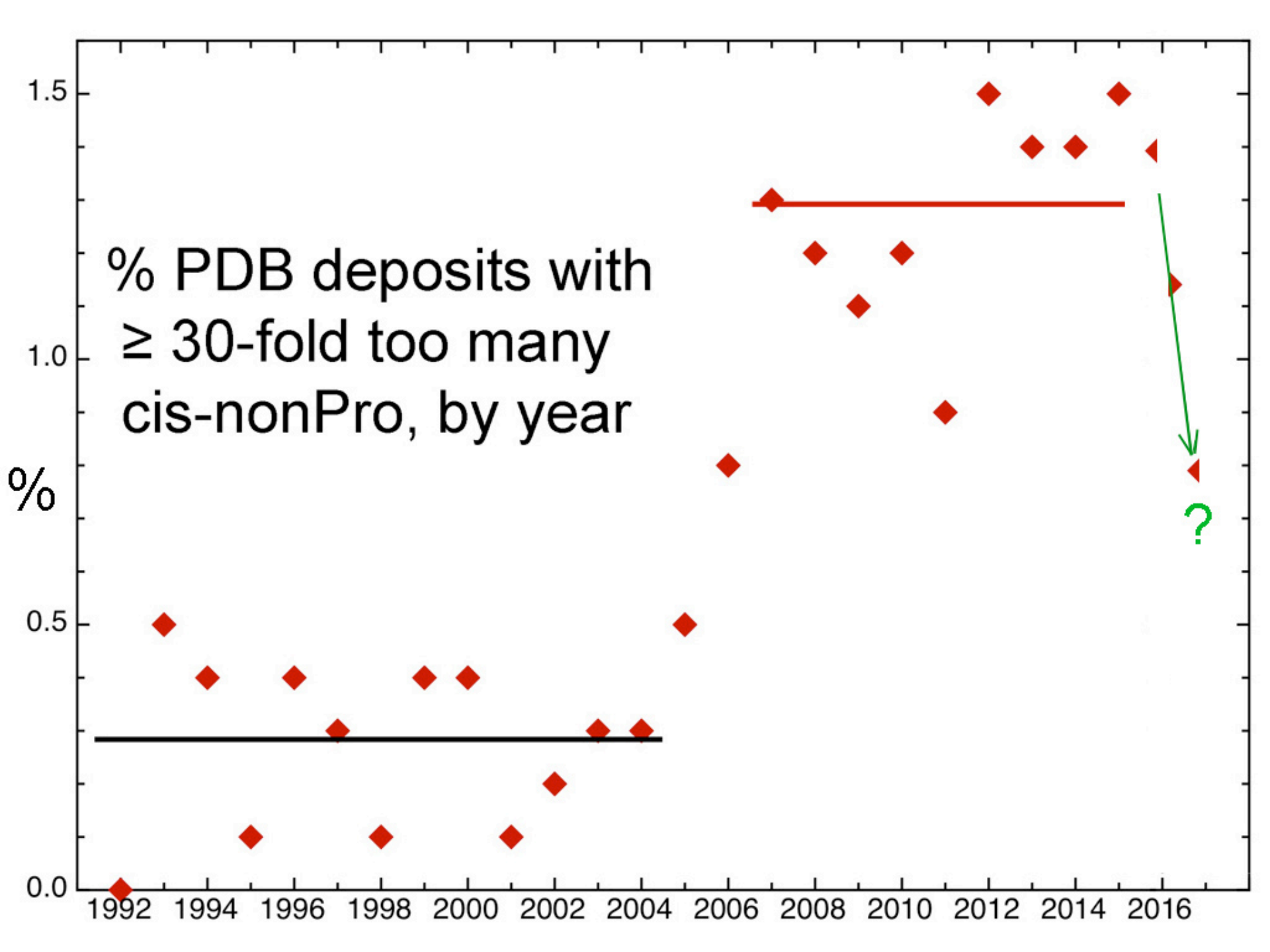

Arg-Gln-Asn-Ser triple *cis*-nonPro -- unjustified
2j82
1.28Å

Addressing the plague

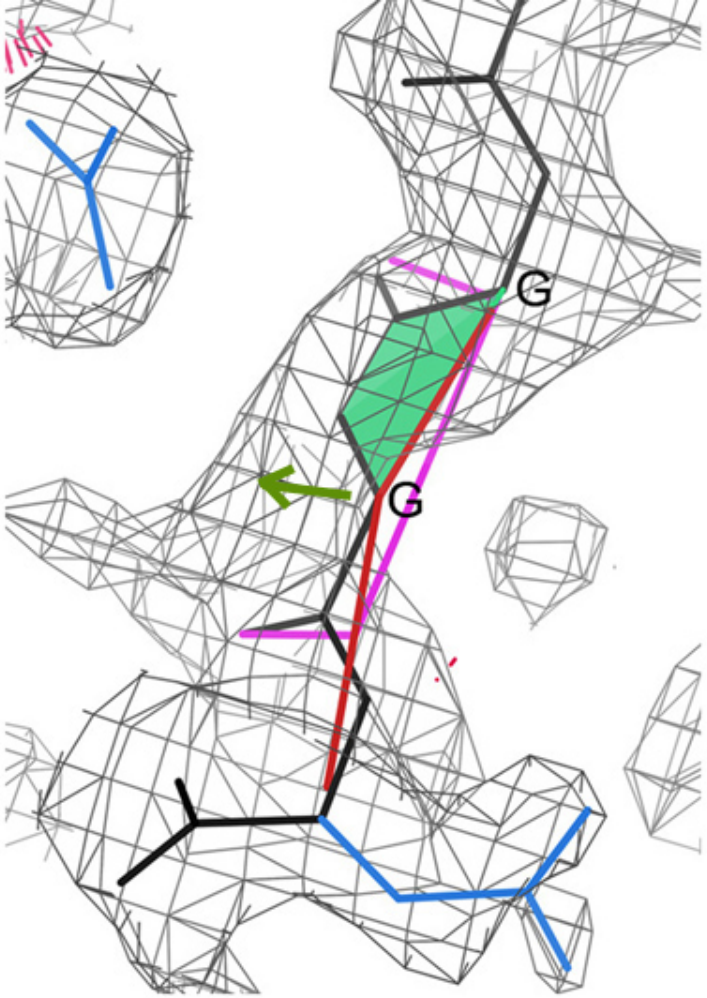


Initial step: making *cis*-nonPro & twisted peptides obvious in MolProbity, Phenix, & Coot (done)

But better to avoid them in the first place!

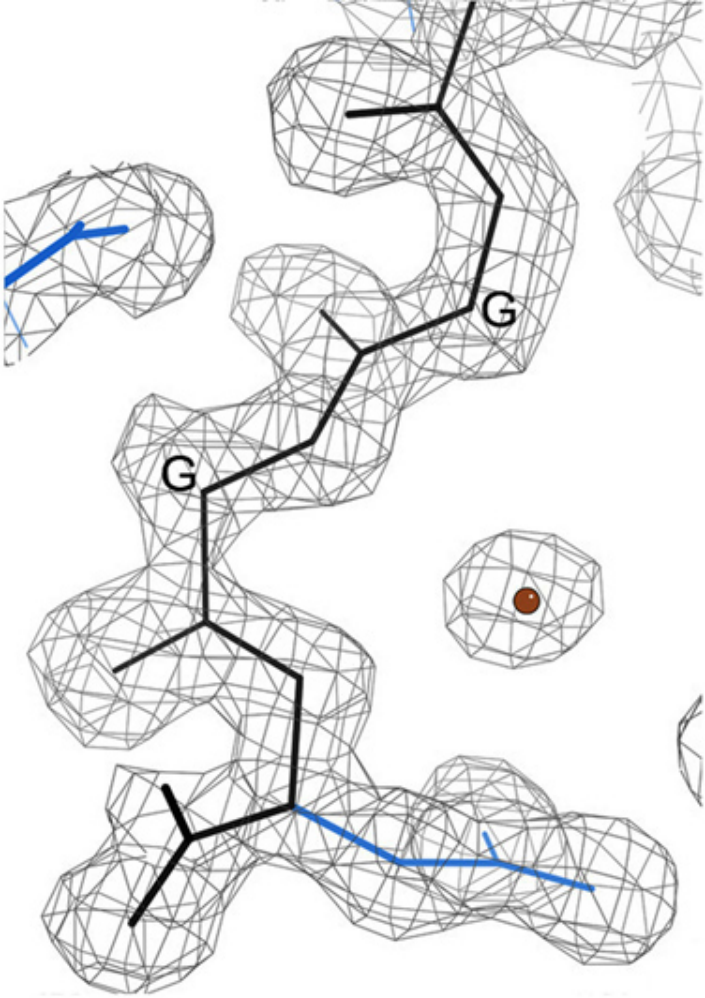

% PDB deposits with
≥ 30-fold too many
cis-nonPro, by year
%
1.5
1.0
0.5
0.0
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
?

CryoEM
β-gal
2.2Å
5a1a



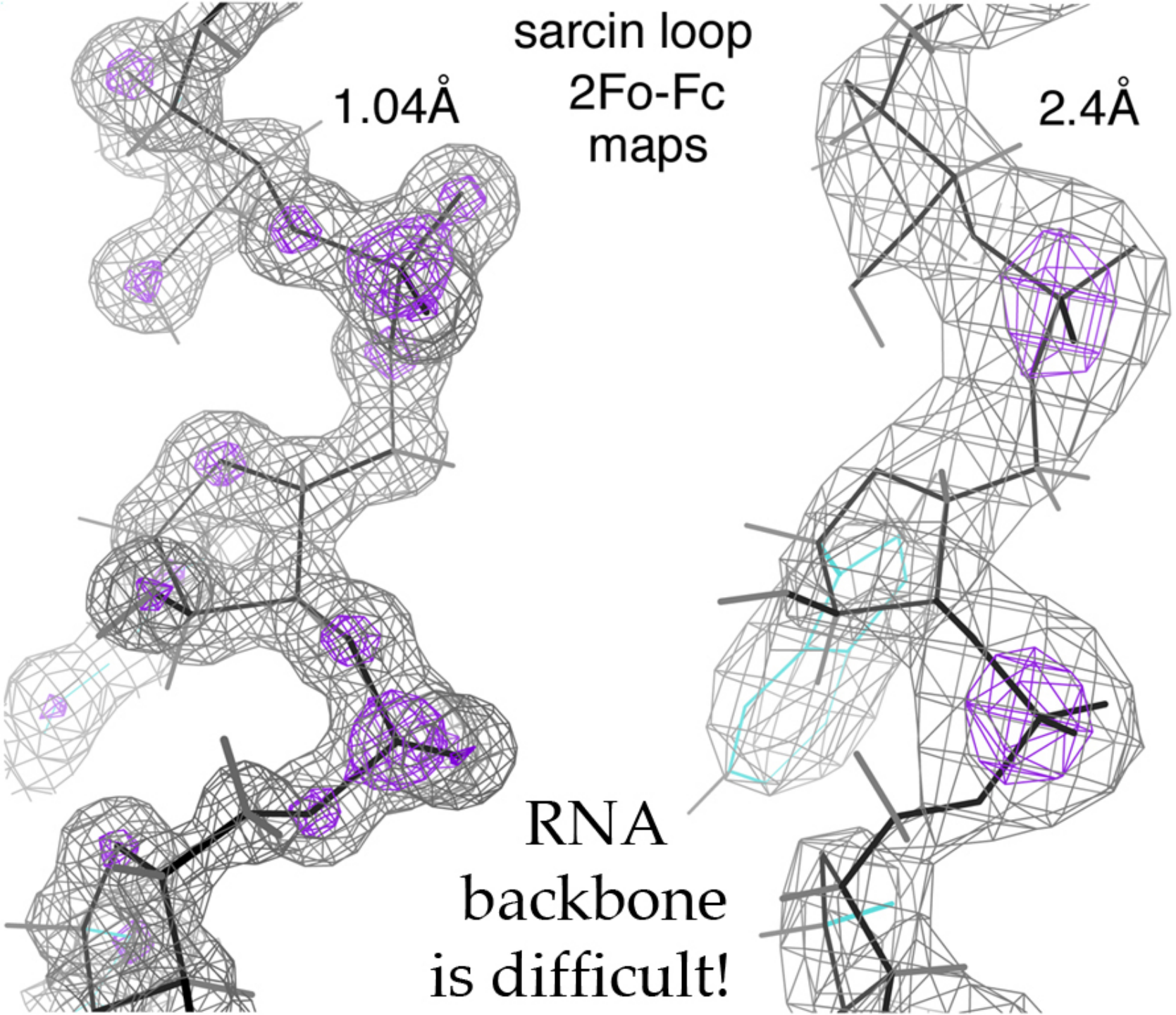
GG995 *cis* nonPro, bad fit

X-ray
β-gal
1.6Å
4ttg

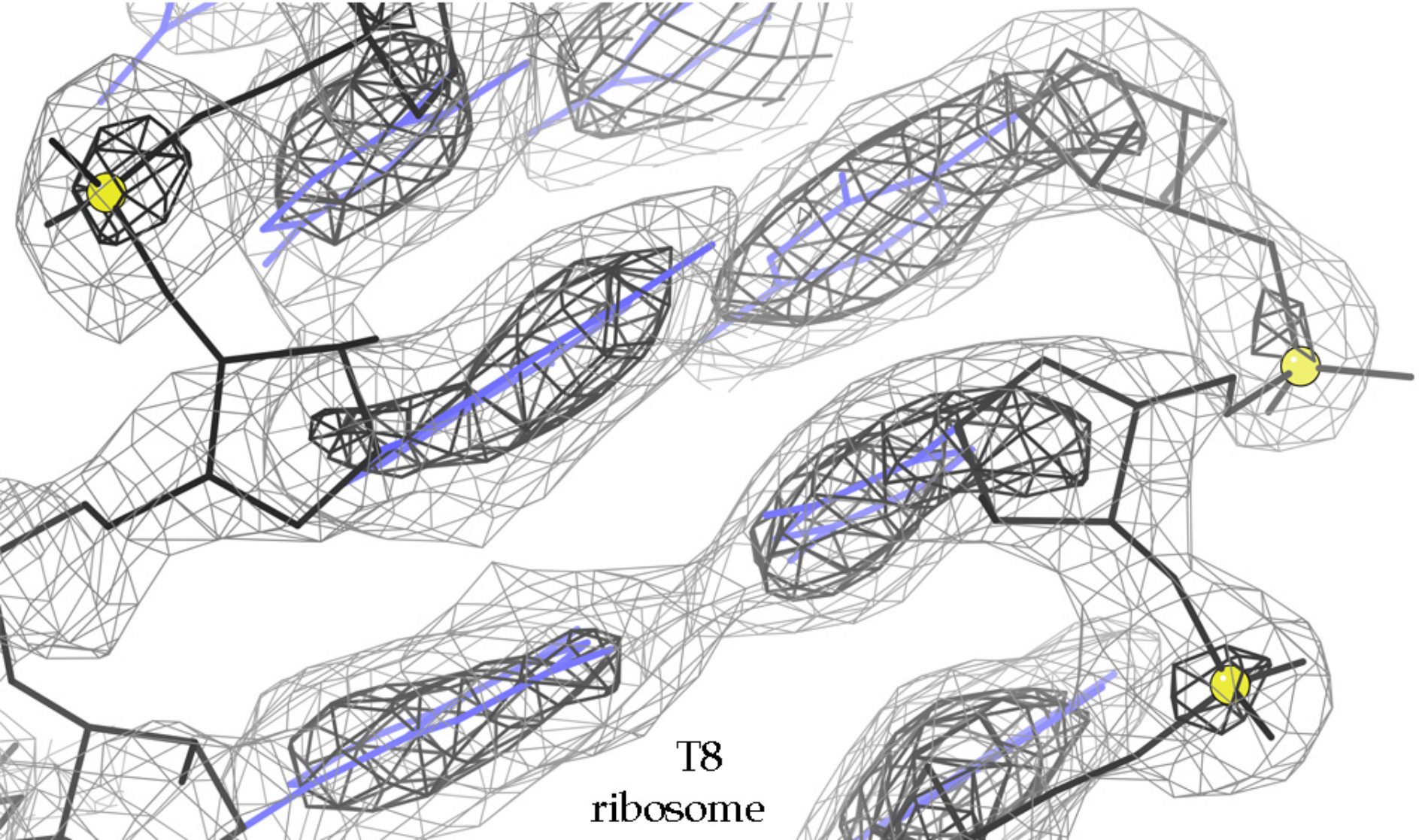


GG995 *trans* nonPro, clear

So -- a cryo EM map at 2.2Å can distinguish between cis and trans non-prolines, and trans should always be tried. CaBLAM helps, but density cannot justify cis at 3-4Å.


sarcin loop
2Fo-Fc
maps
1.04Å
2.4Å
RNA
backbone
is difficult!

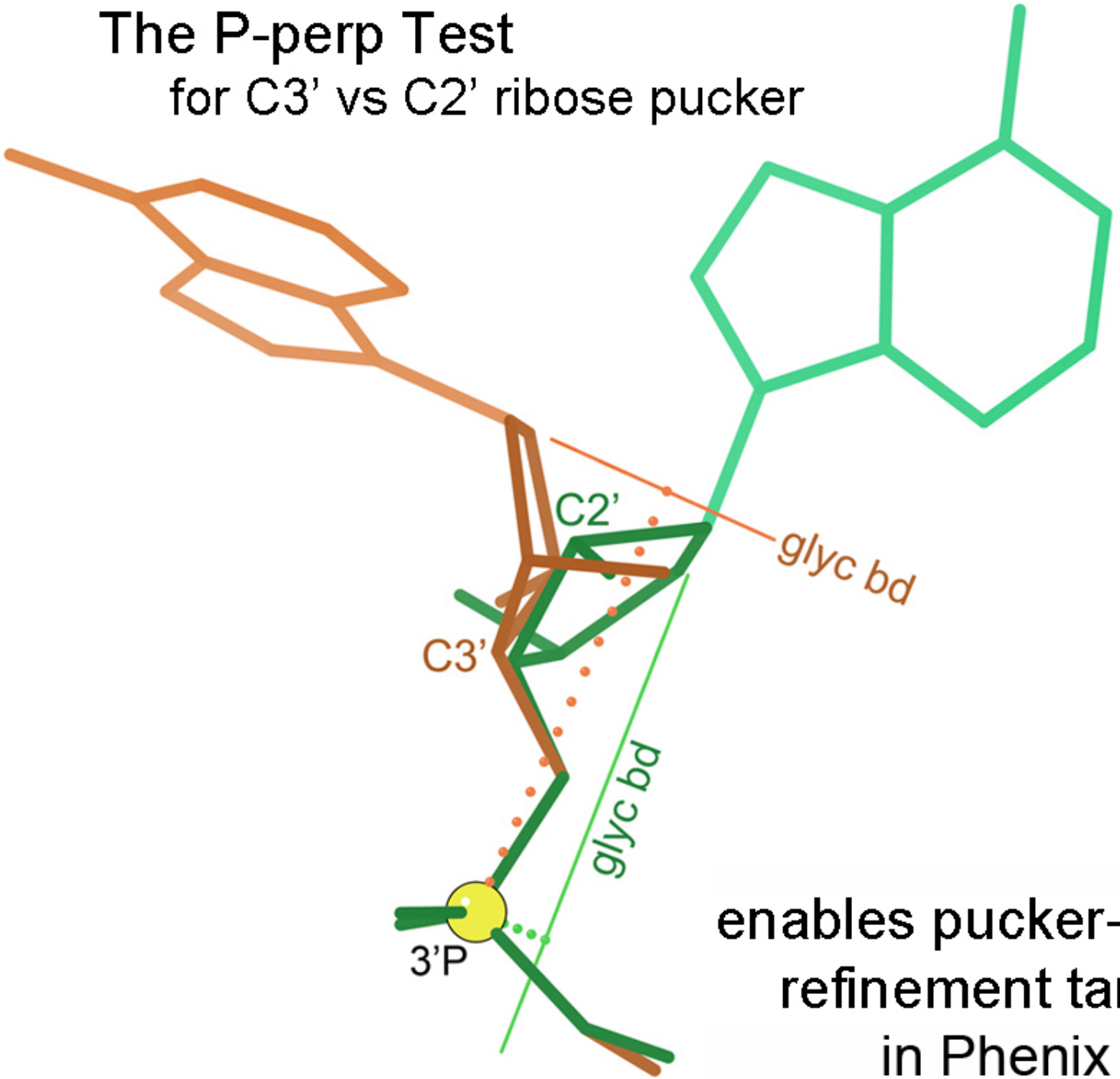
Like negative carboxyls, RNA PO_4s show reduced density. Backbone connectivity is weak, so validation of backbone conformation & ribose pucker is especially needed.



T8
ribosome

The P-perp Test

for C3' vs C2' ribose pucker



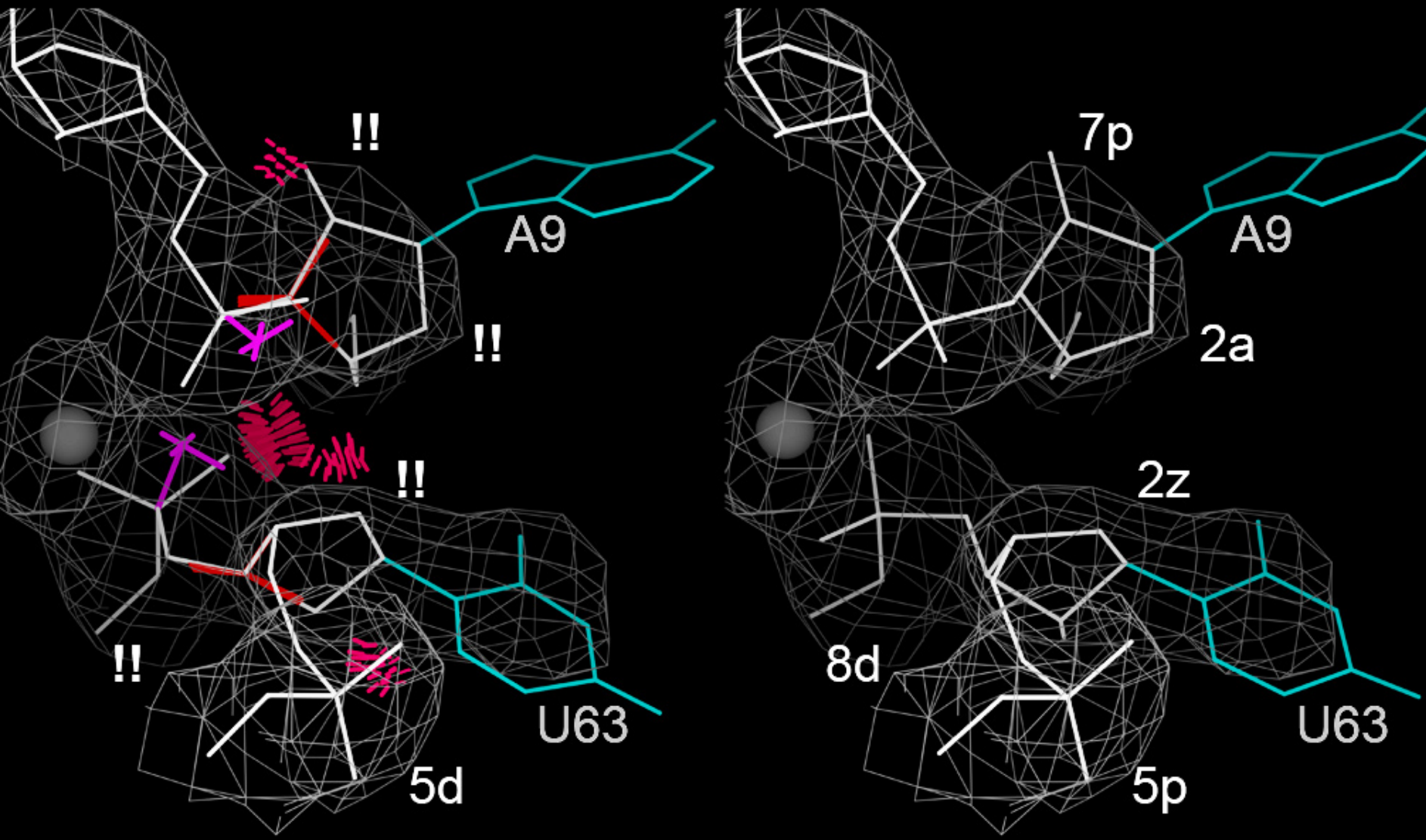
enables pucker-specific refinement targets in Phenix

ERRASER =

Phenix refinement,
MolProbity on RNA,
Rosetta relax,
Rosetta Step-Wise Assembly
Rosetta relax,
Phenix

Rhiju Das, Fang Chou

Stanford

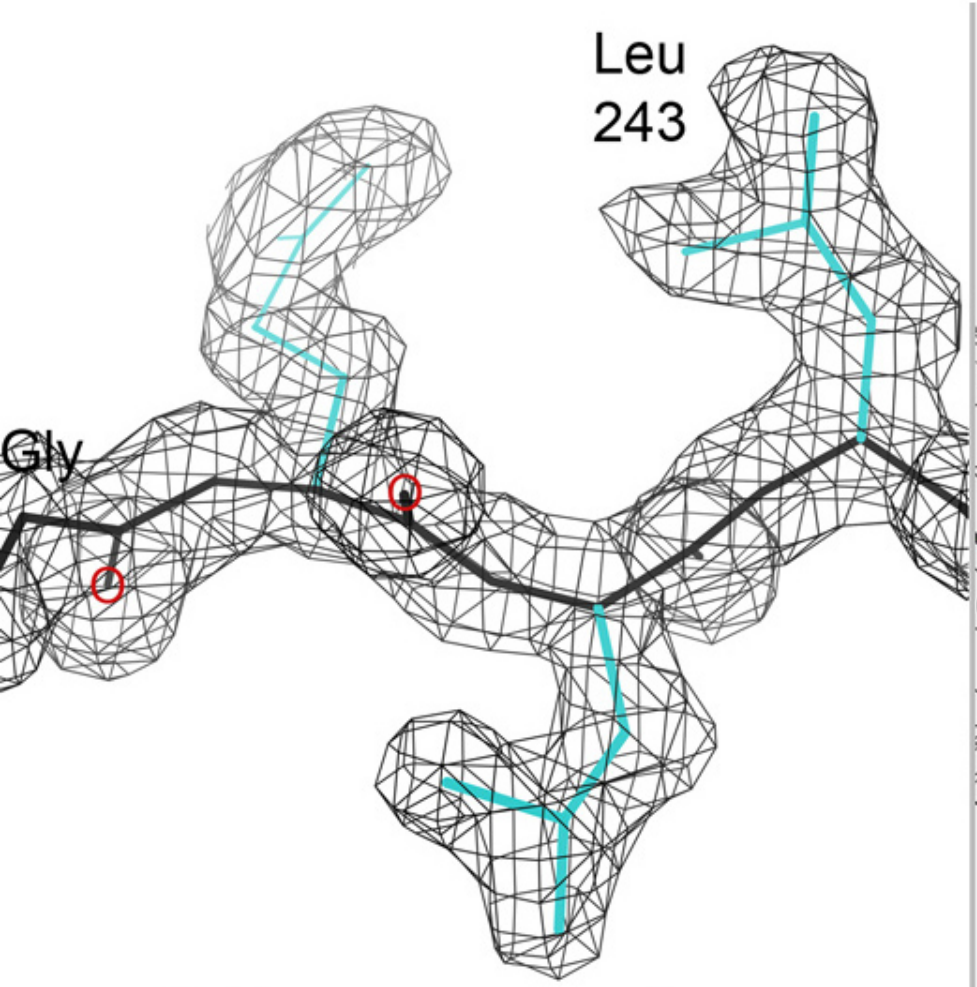


original PDB

ERRASER run 2

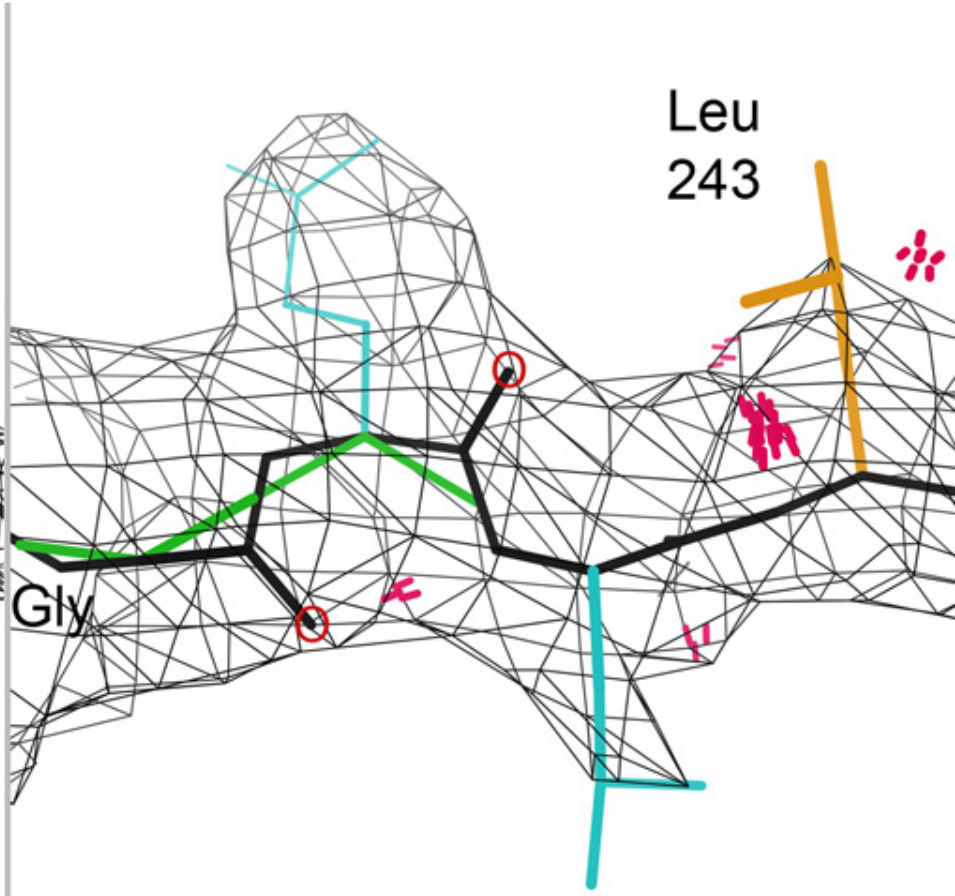
2GIS: A9-G63 contact with Mg++

Tackling Low Resolution



CO groups clear

3fpn 1.8Å

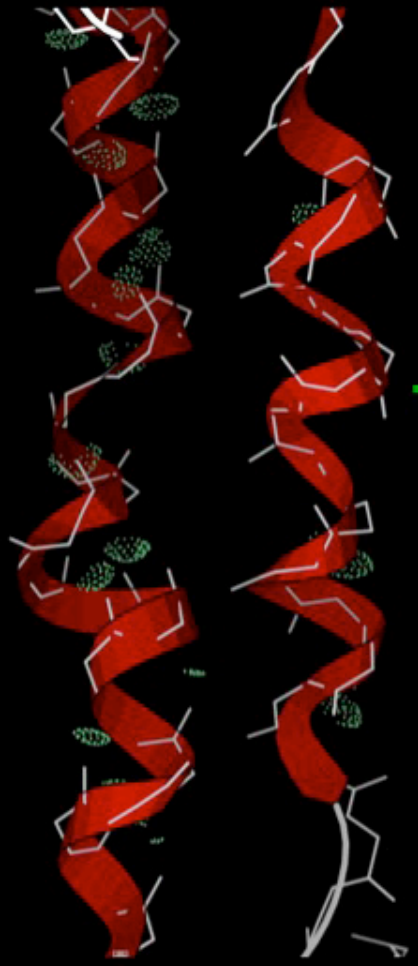


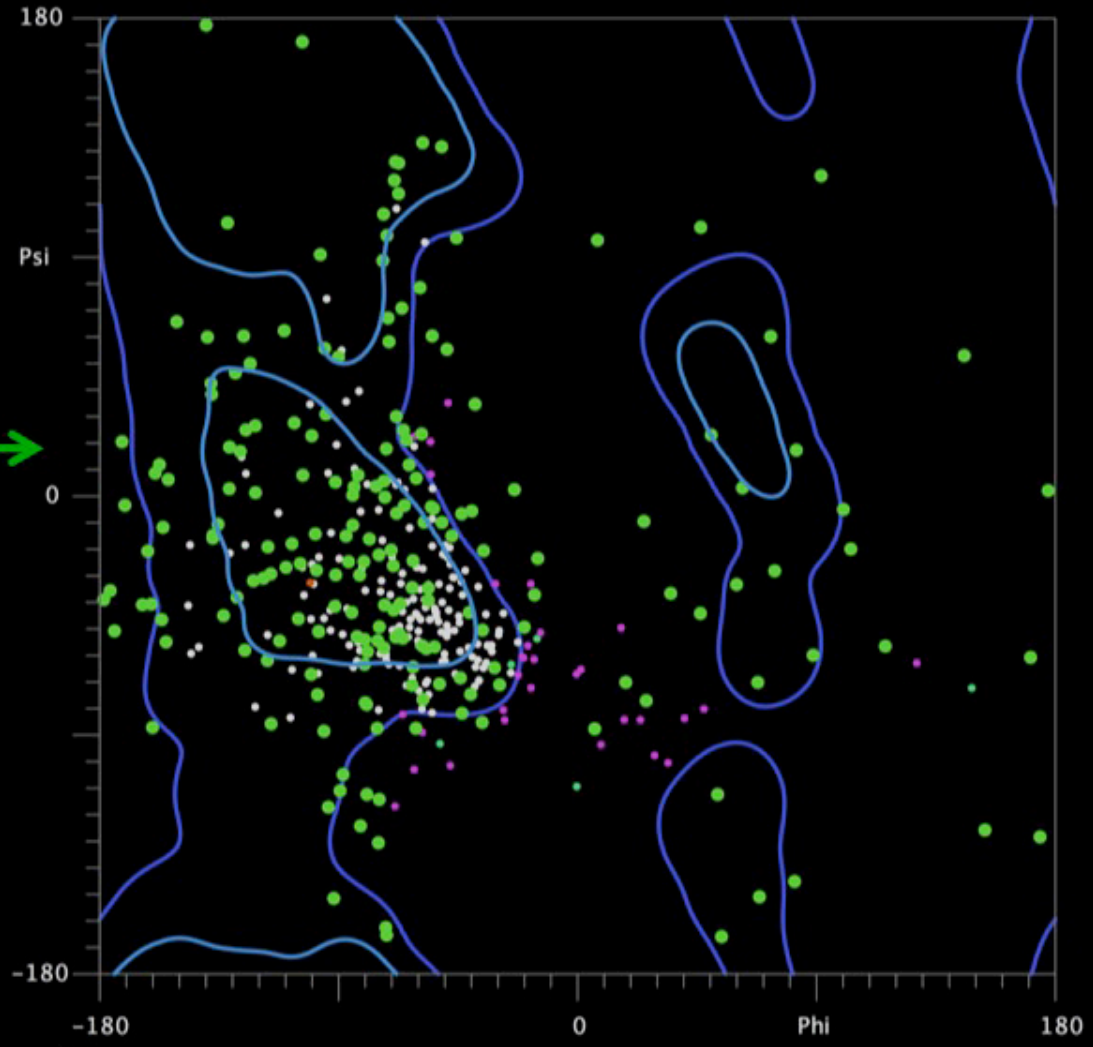
CO direction lost !

2r6f 3.2Å

Secondary Structure Diagnosis

DSSP or Ramachandran fail at low resolution

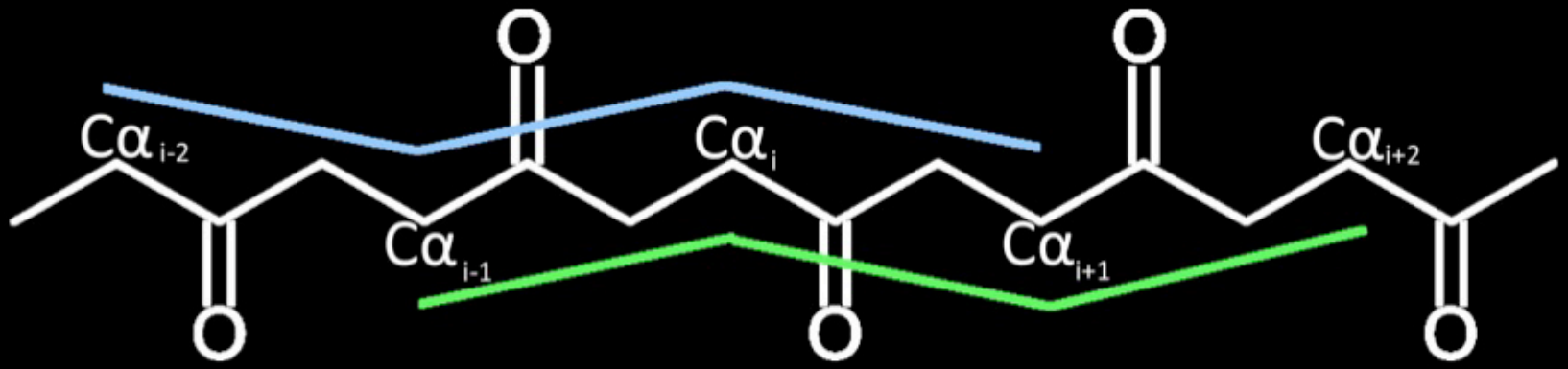


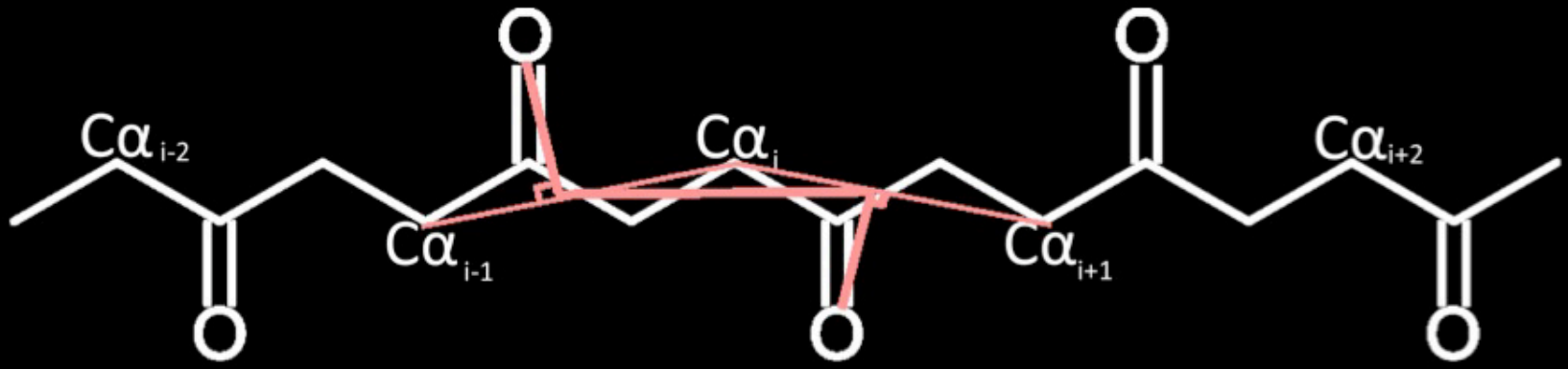


2o01: Photosystem I, 3.4Å

CaBLAM Parameter Space

A minimalist alternative

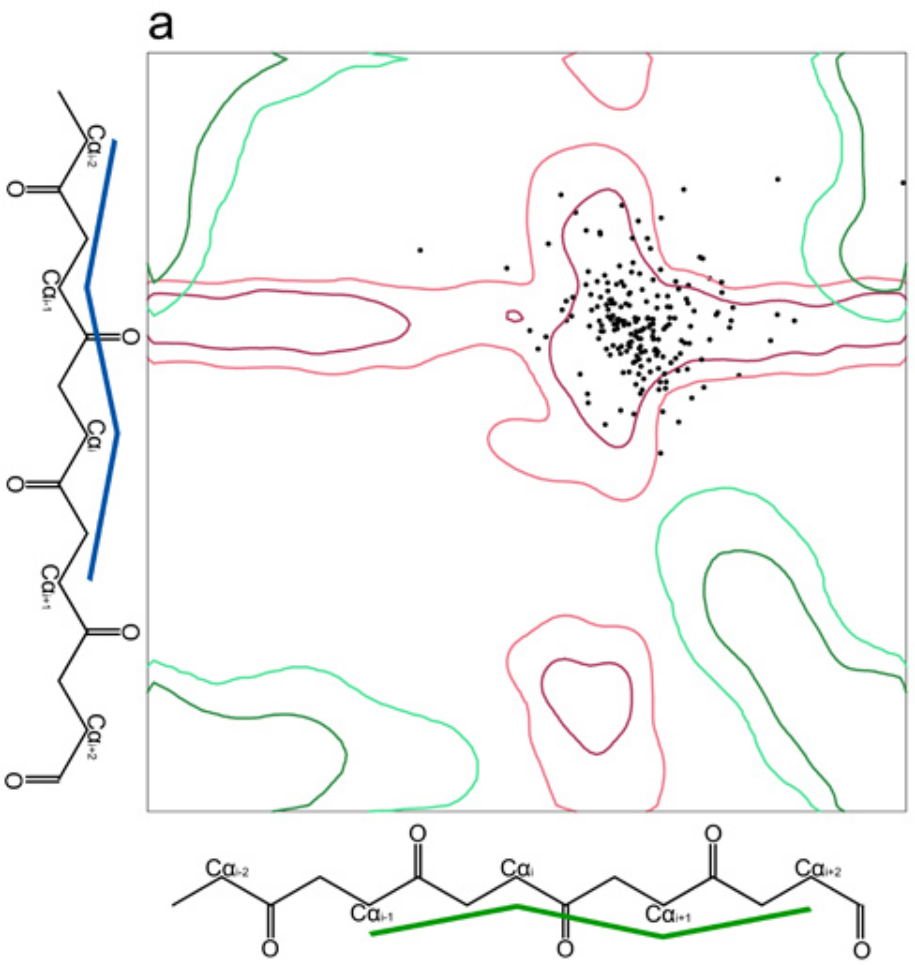


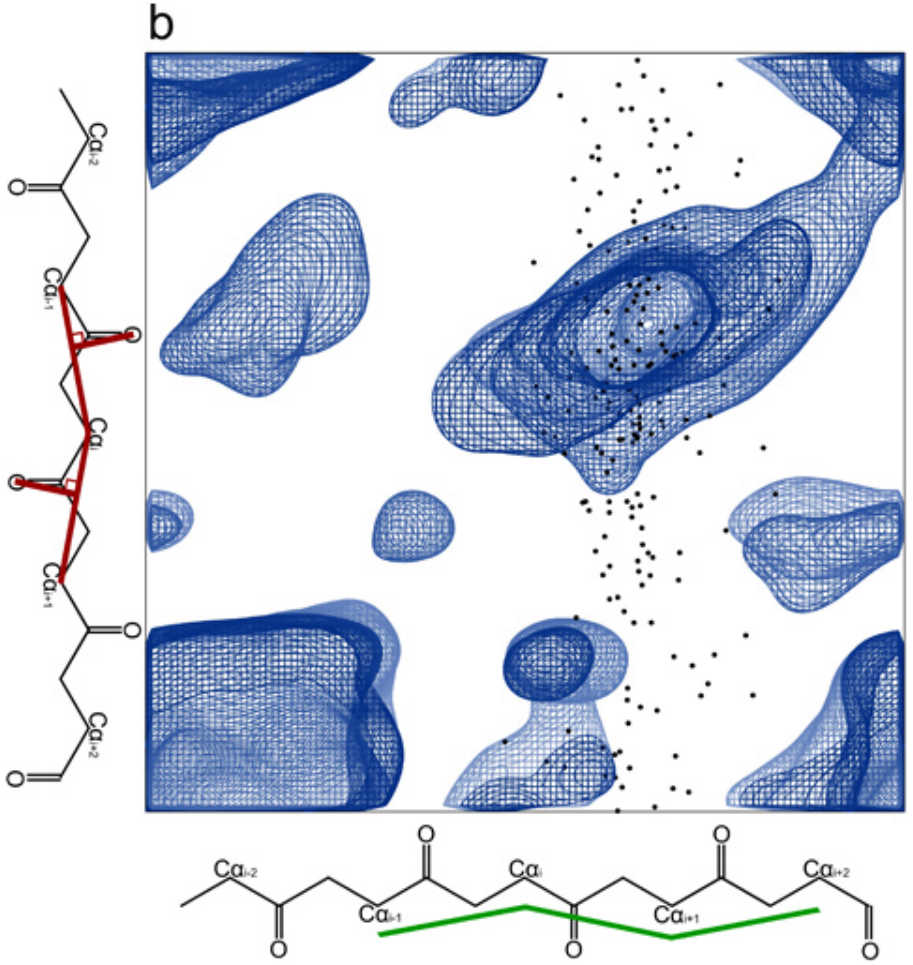


CaBLAM α-helix diagnosis

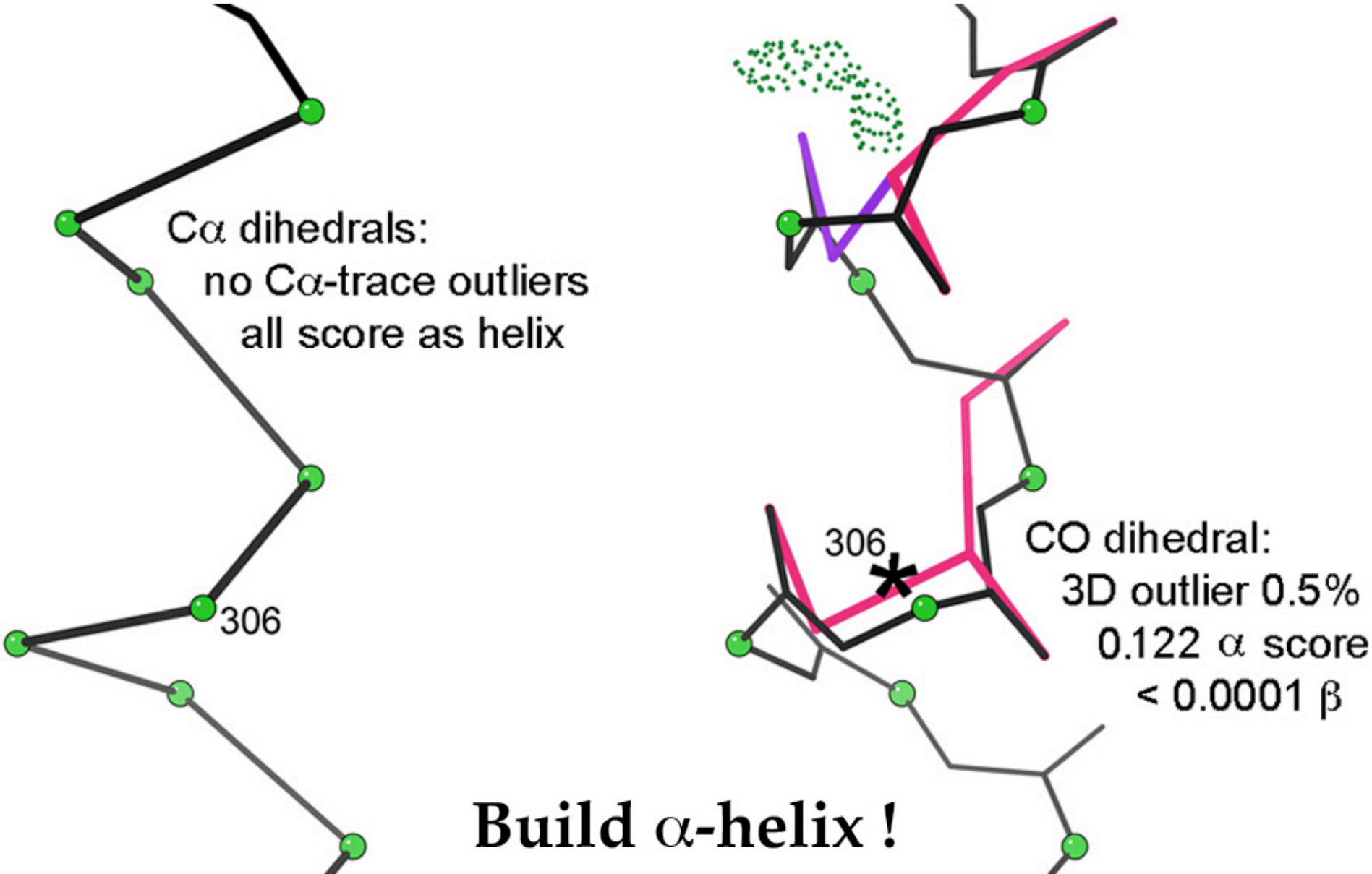
Cα–Cα dihedral datapoints:
~all in α region (red)

CO dihedral datapoints:
many outliers





CaBLAM secondary-structure assignment



Build α-helix !

Comprehensive validation (cryo-EM) (Project: jsr)

Preferences Help Run Abort Ask for help

Input/Output **ValidationCryoEM_3**

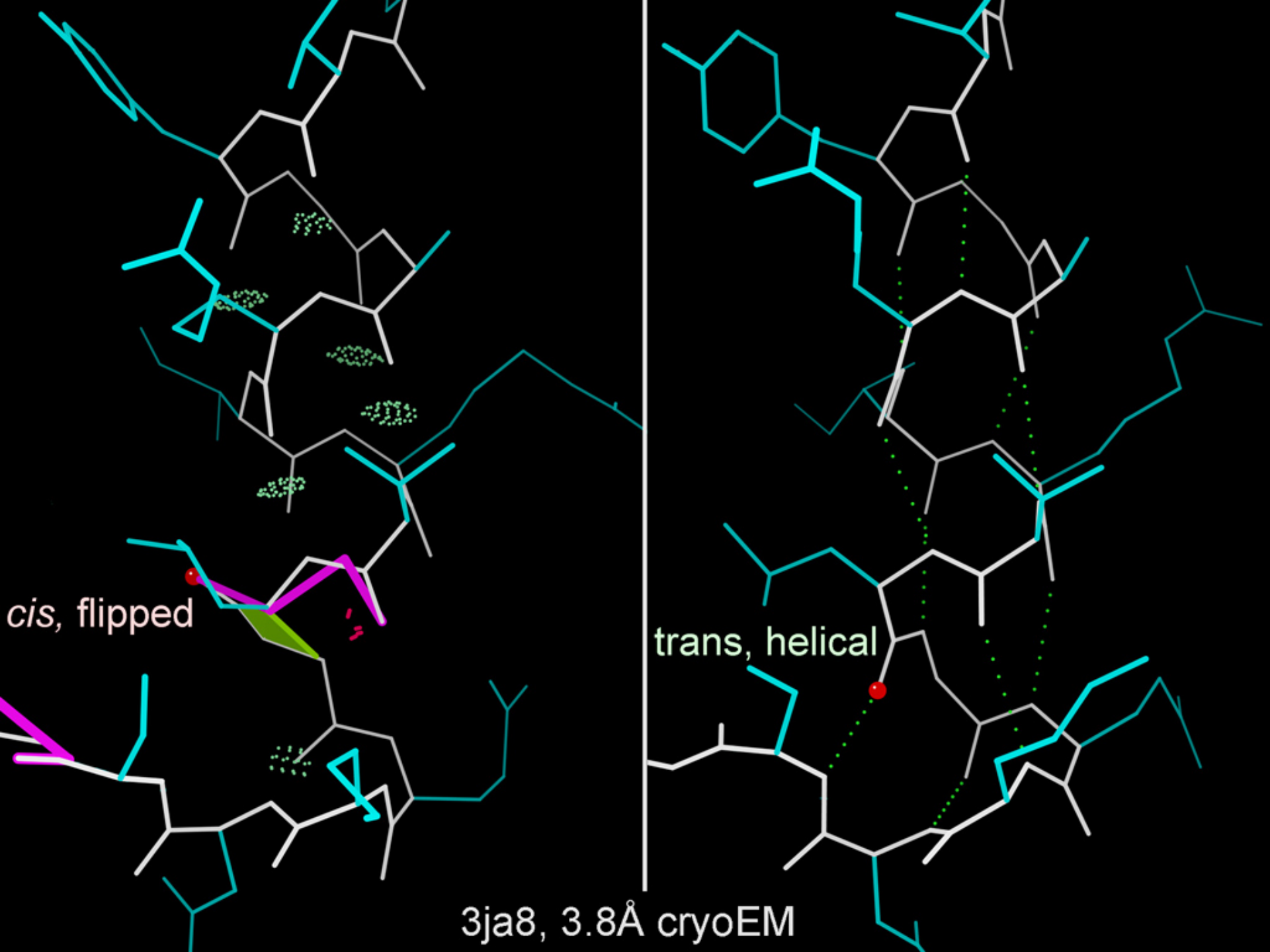
Run status Summary **MolProbity**

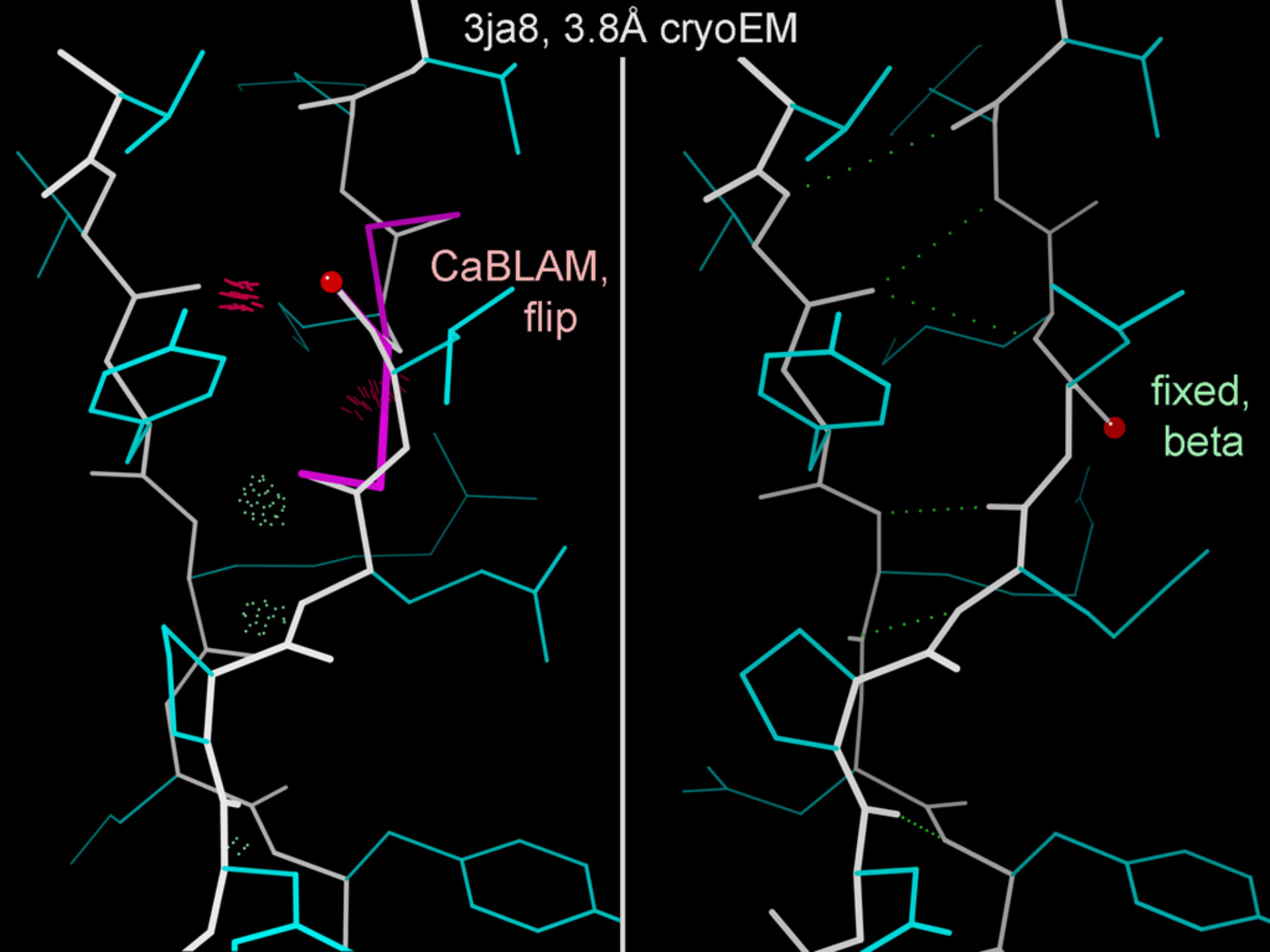
Clashes **CaBLAM** Cβ Cis/Twisted Rotamers Ramachandran Geometry Restraints

CaBLAM

Outliers (%): 2.56 **Disfavored (%):** 9.15 **Cα outliers (%):** 0.62

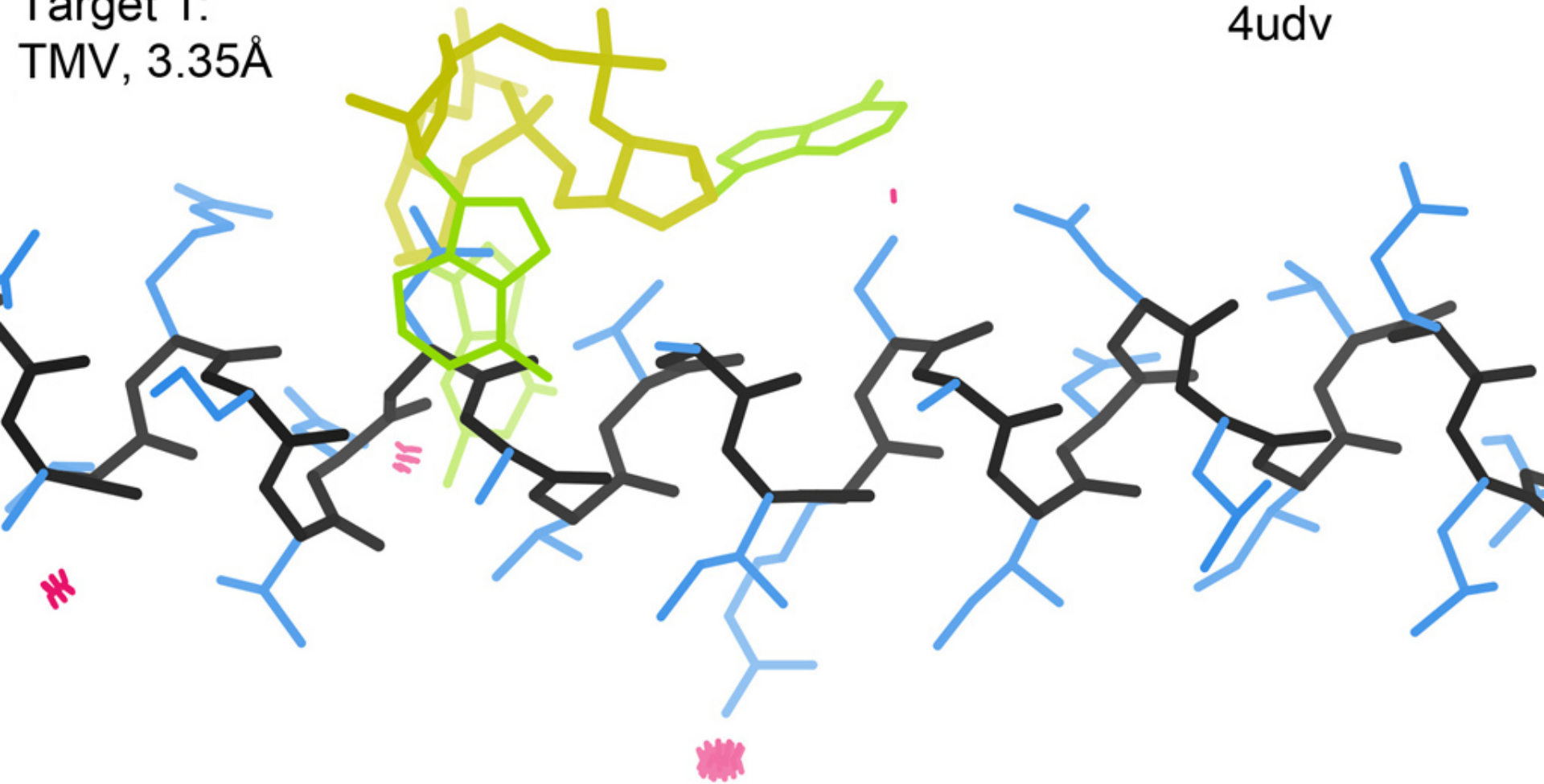
Chain	Residue	Evaluation	CaBLAM Score	CA Geometry Score	Secondary Structure	Helix Sco
C	VAL 605	CaBLAM Disfavored	0.03392	0.12811		0.00060
C	ASP 624	CaBLAM Outlier	0.00769	0.01827		0.00000
C	GLY 625	CaBLAM Disfavored	0.02562	0.03299		0.00822
C	LYS 626	CaBLAM Disfavored	0.04033	0.12037		0.00452
C	ALA 653	CaBLAM Outlier	0.00529	0.10339		0.00086
C	LYS 654	CaBLAM Disfavored	0.03471	0.02375		0.00000
C	ASN 655	CaBLAM Outlier	0.00076	0.02041		0.00121
C	LEU 657	CaBLAM Disfavored	0.02920	0.45221	try beta sheet	0.00000
C	GLN 667	CaBLAM Outlier	0.00223	0.01572		0.00032
C	LEU 721	CaBLAM Disfavored	0.02489	0.10874		0.00038
C	THR 730	CaBLAM Outlier	0.00003	0.30073		0.00000
C	SER 740	CaBLAM Disfavored	0.03416	0.07829		0.00000
C	ILE 752	CaBLAM Disfavored	0.03395	0.10709	try beta sheet	0.00000
C	TYR 785	CA Geom Outlier	0.01132	0.00444		0.00054
C	VAL 802	CaBLAM Disfavored	0.01307	0.10644		0.00000
C	GLY 803	CaBLAM Disfavored	0.02224	0.30186		0.00058
C	ARG 808	CaBLAM Disfavored	0.02253	0.14997		0.00060


cis, flipped
trans, helical
3ja8, 3.8Å cryoEM


3ja8, 3.8Å cryoEM
CaBLAM,
flip
fixed,
beta

Target 1:
TMV, 3.35Å

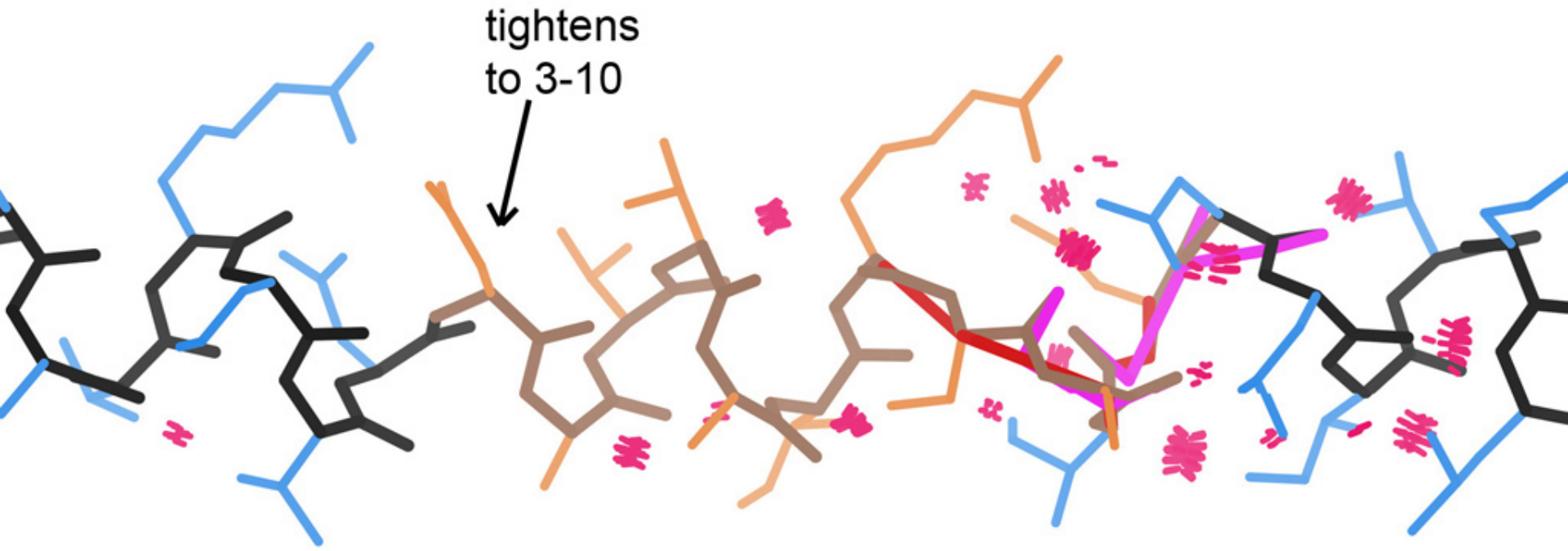
4udv



No Rama or geometry outliers

Target 1:
TMV, 3.35Å

model 181_1
helix 110-130

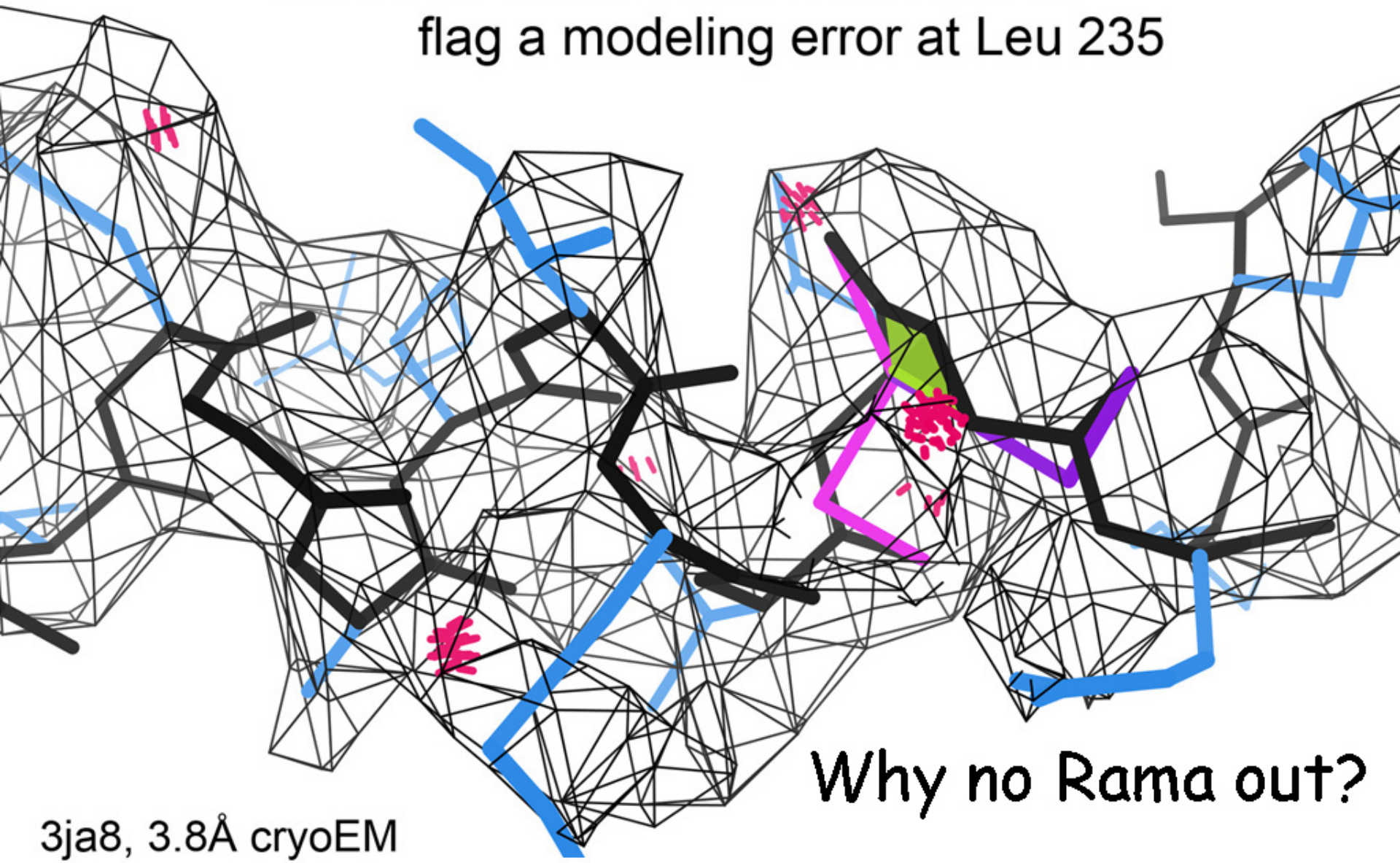


Sequence Misalignment !

No Rama or geometry outliers

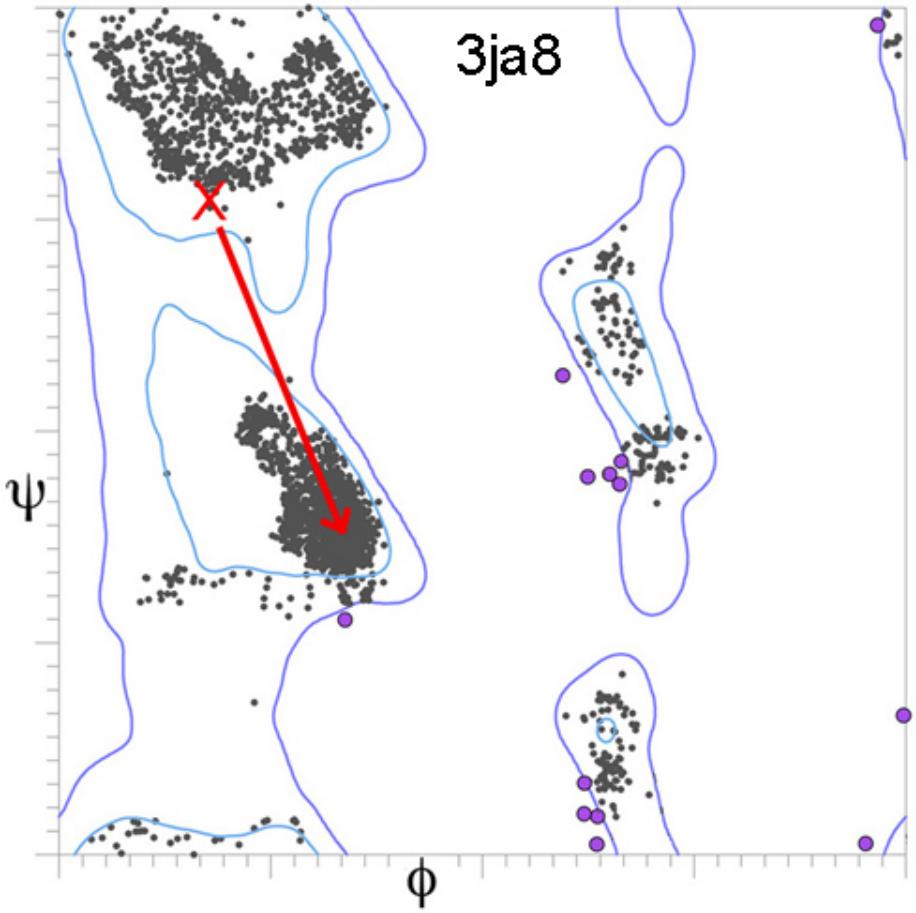
CaBLAM outliers &
more clashes, at end

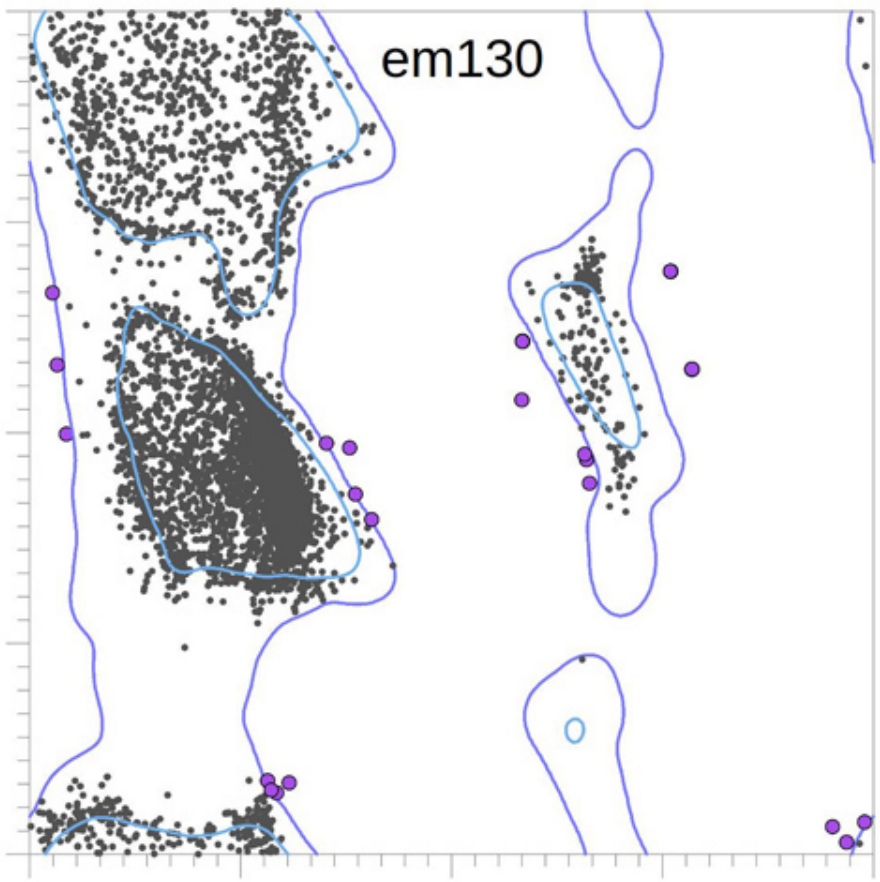
No traditional validation outliers, but
CaBLAM outliers and a *cis*-nonPro
flag a modeling error at Leu 235



Why no Rama out?

3ja8, 3.8Å cryoEM



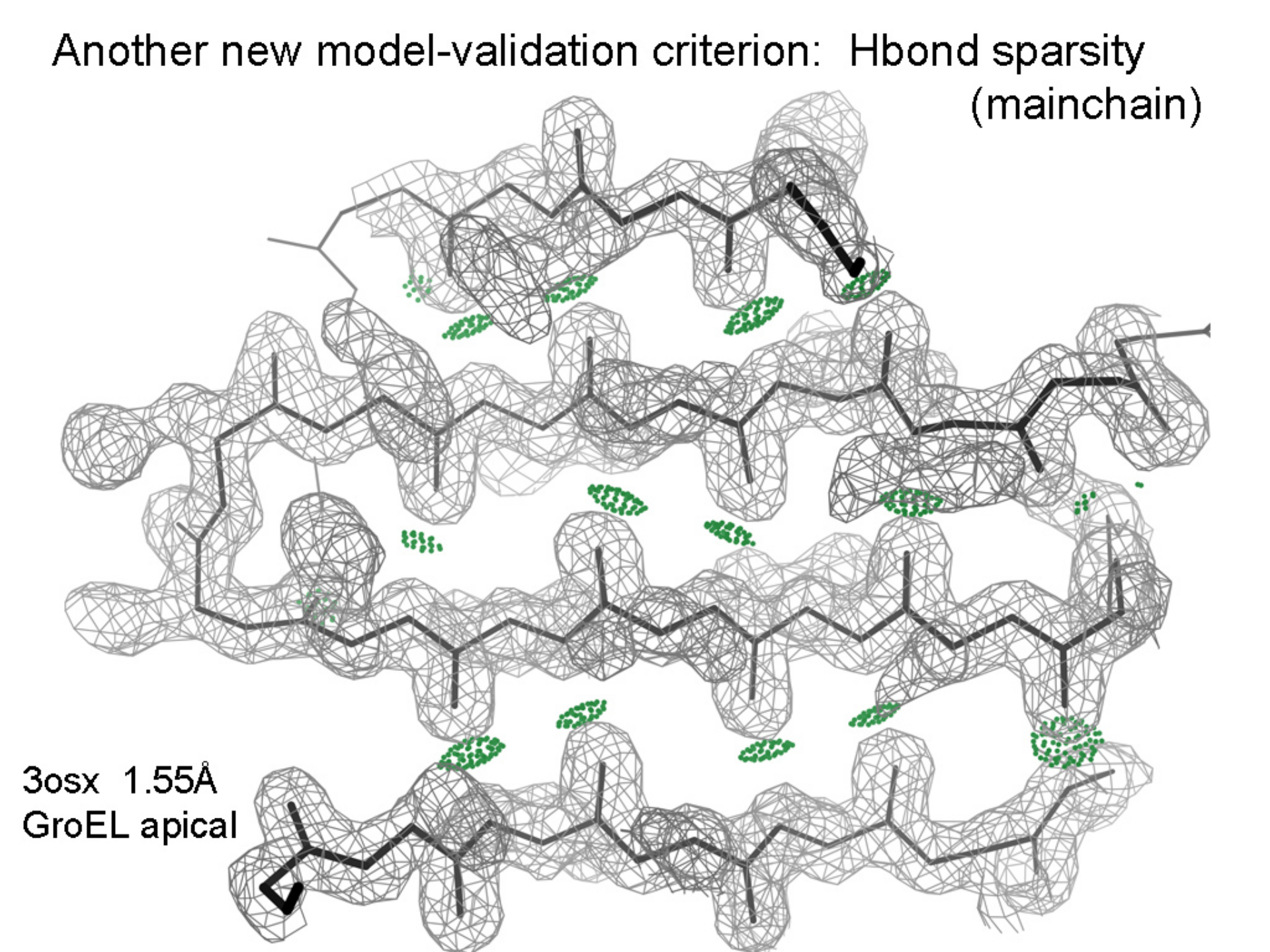


Tight restraints on Ramachandran values

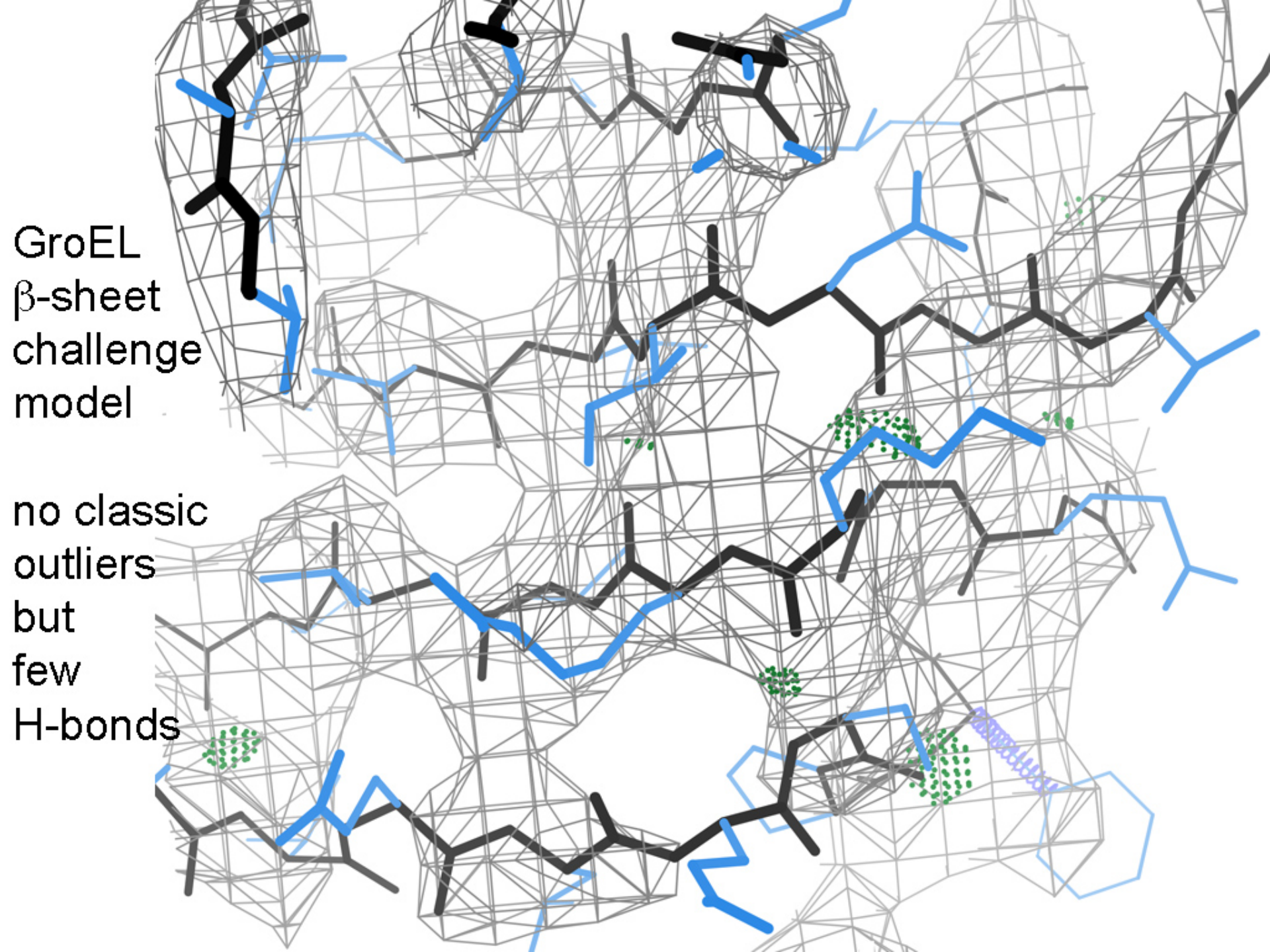
Pileup along high contours, differs among software

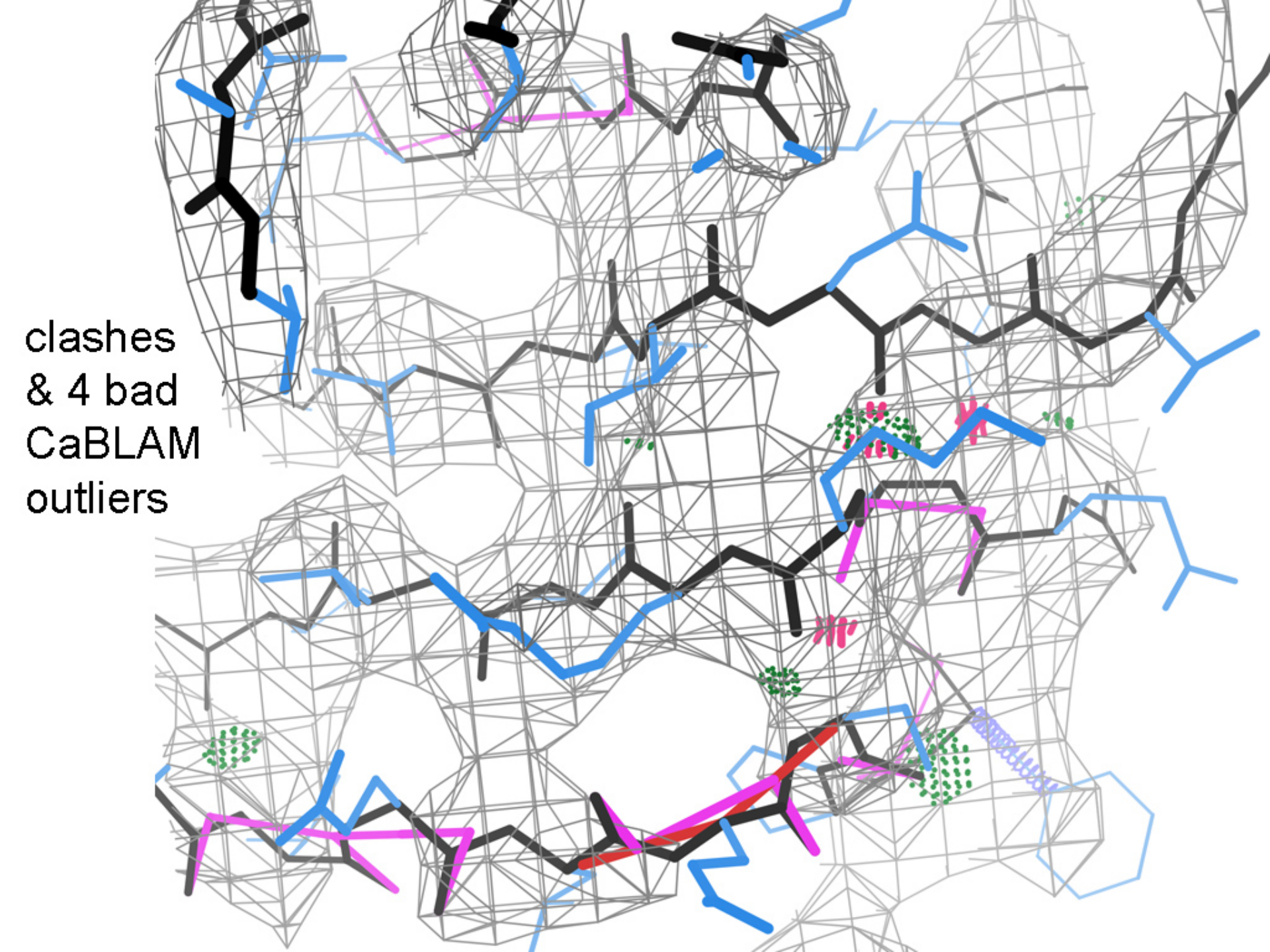
Can shove outliers into wrong local minimum,
uncorrectable by refinement

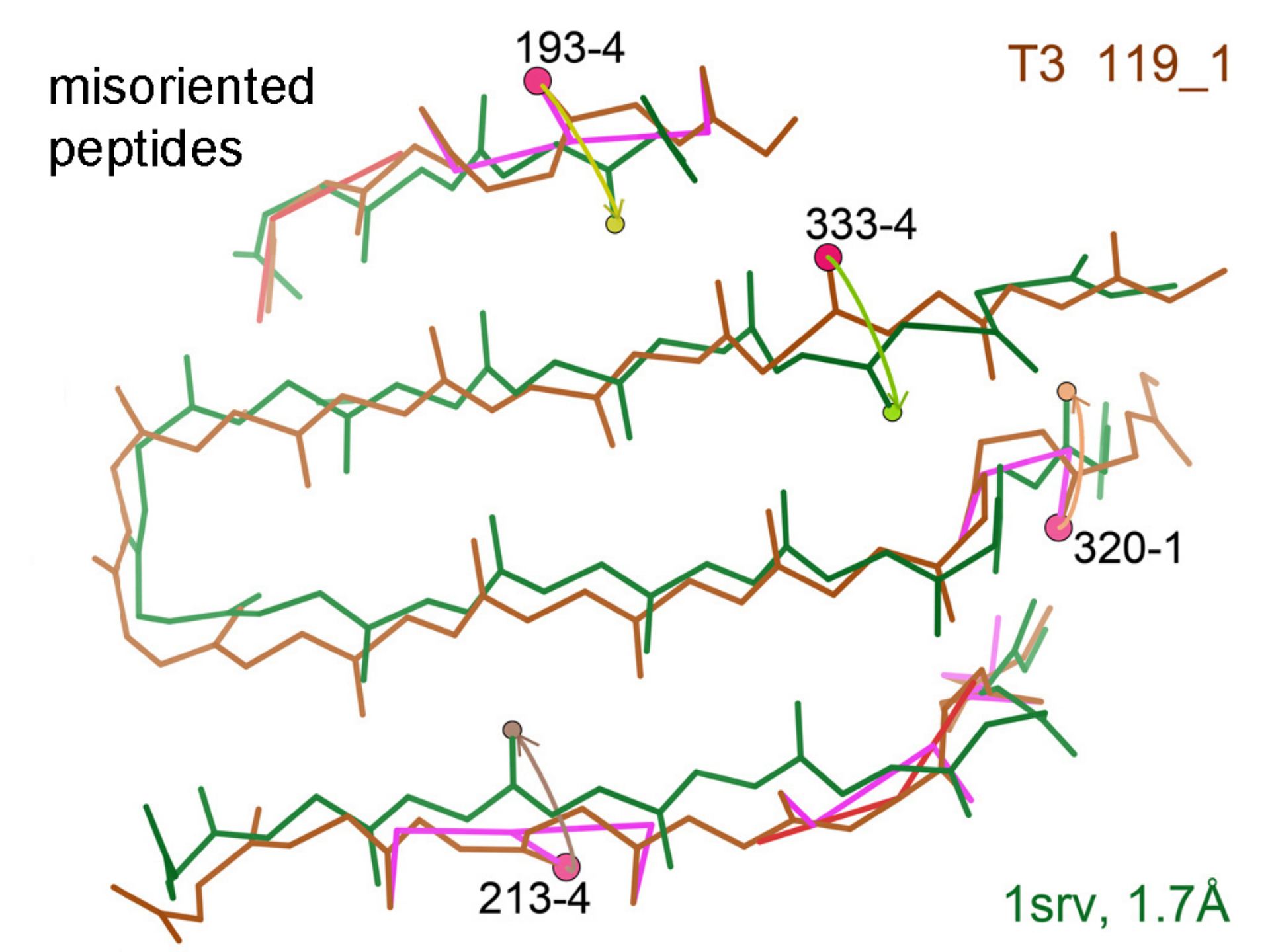
Another new model-validation criterion: Hbond sparsity (mainchain)



3osx 1.55Å
GroEL apical


GroEL
β-sheet
challenge
model
no classic
outliers
but
few
H-bonds

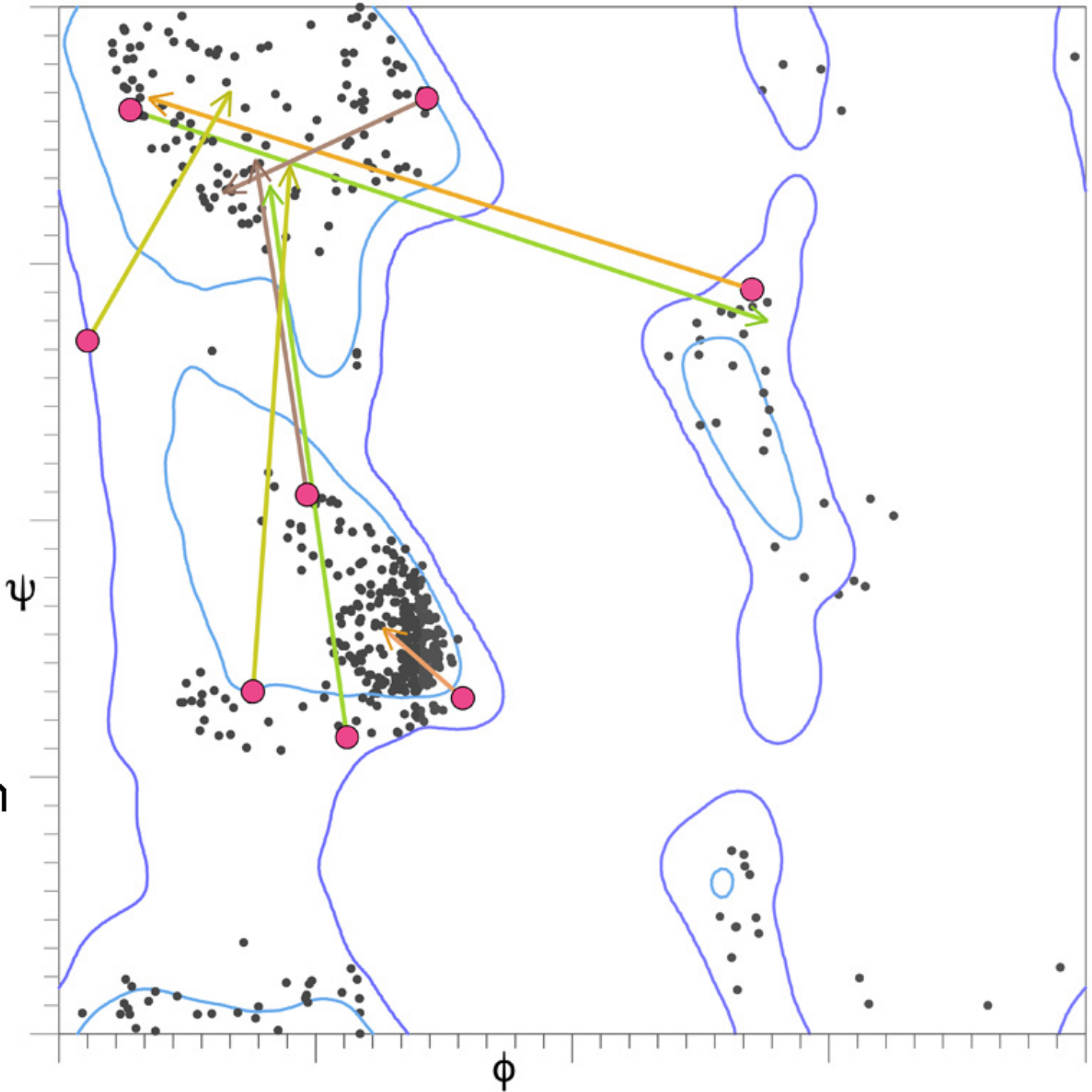

clashes
& 4 bad
CaBLAM
outliers


misoriented
peptides
193-4
333-4
320-1
213-4
T3 119_1
1srv, 1.7Å

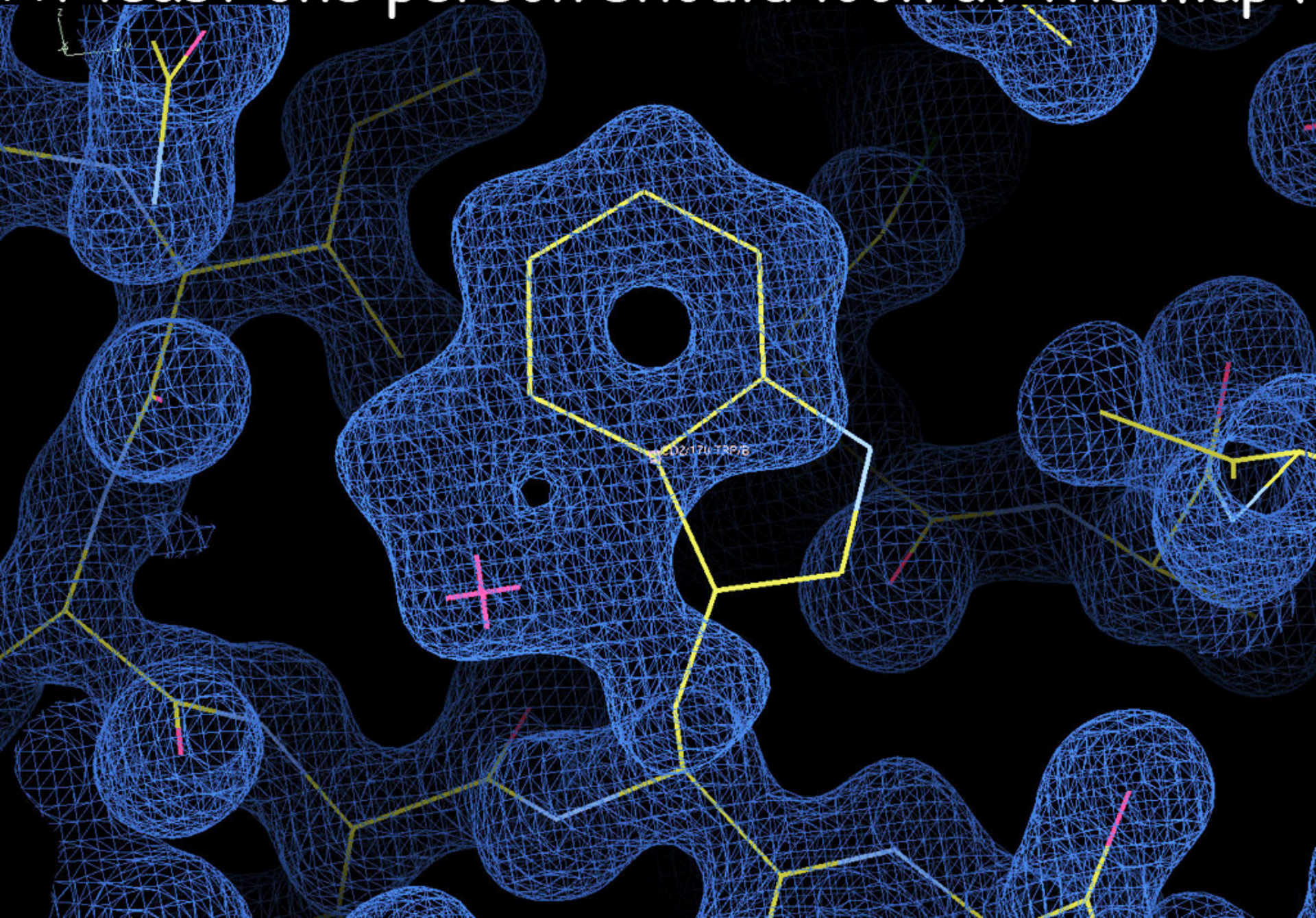
Those flipped peptides have allowed ϕ,ψ values

But --

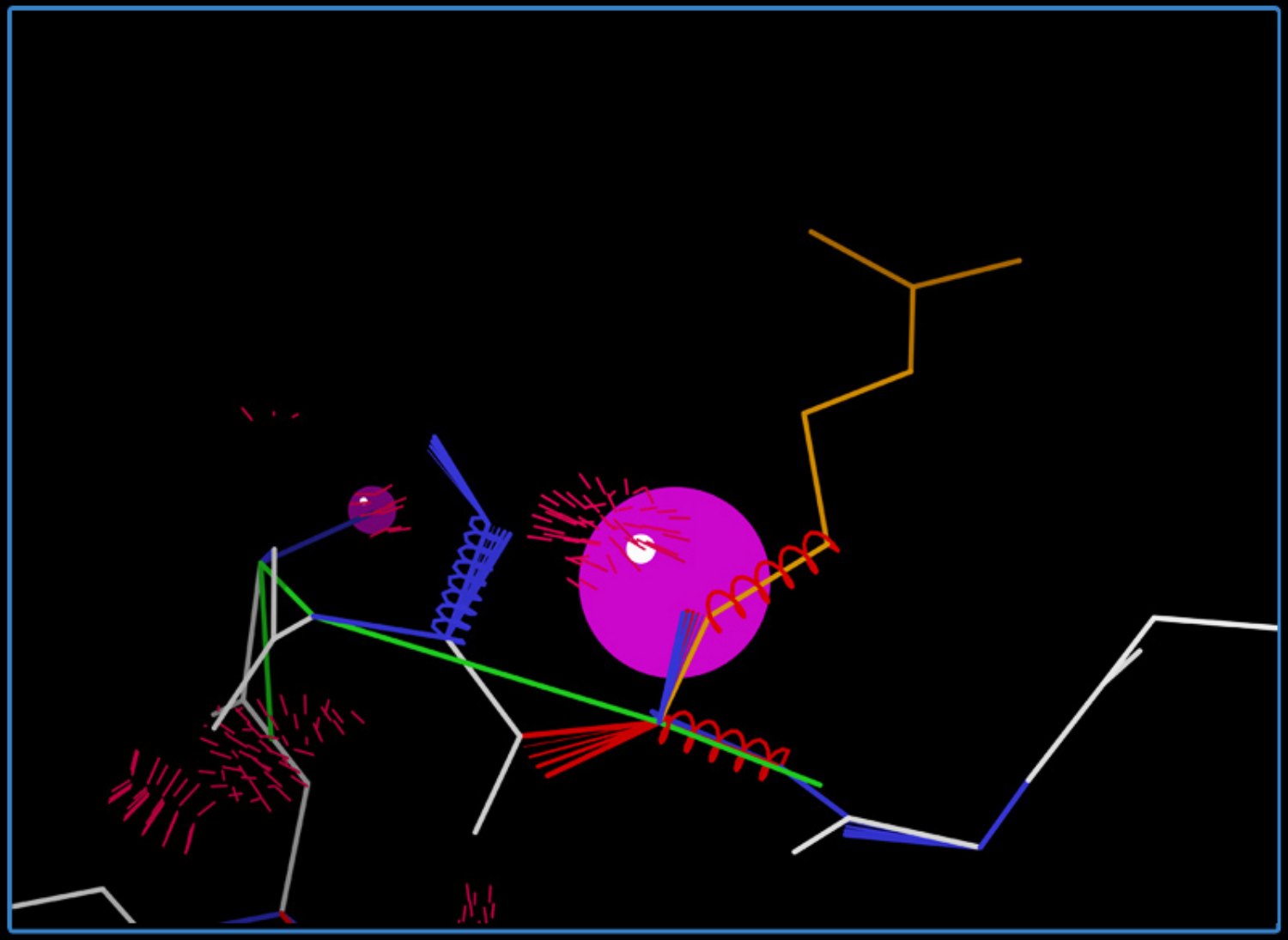
7 of 8 are in the wrong Ramachandran region

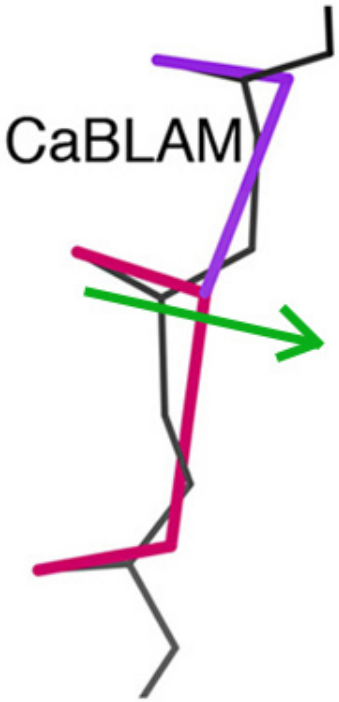


At least one person should look at the map !



. . . and the worst outliers





The most common backbone misfitting at 3-4Å is a misoriented peptide

CaBLAM-diagnose and fix those in the initial model, if possible, before refining with restraints

Easiest cases:

- If within helix or beta, regularize it
- If 2 successive outliers, try rotating middle CO