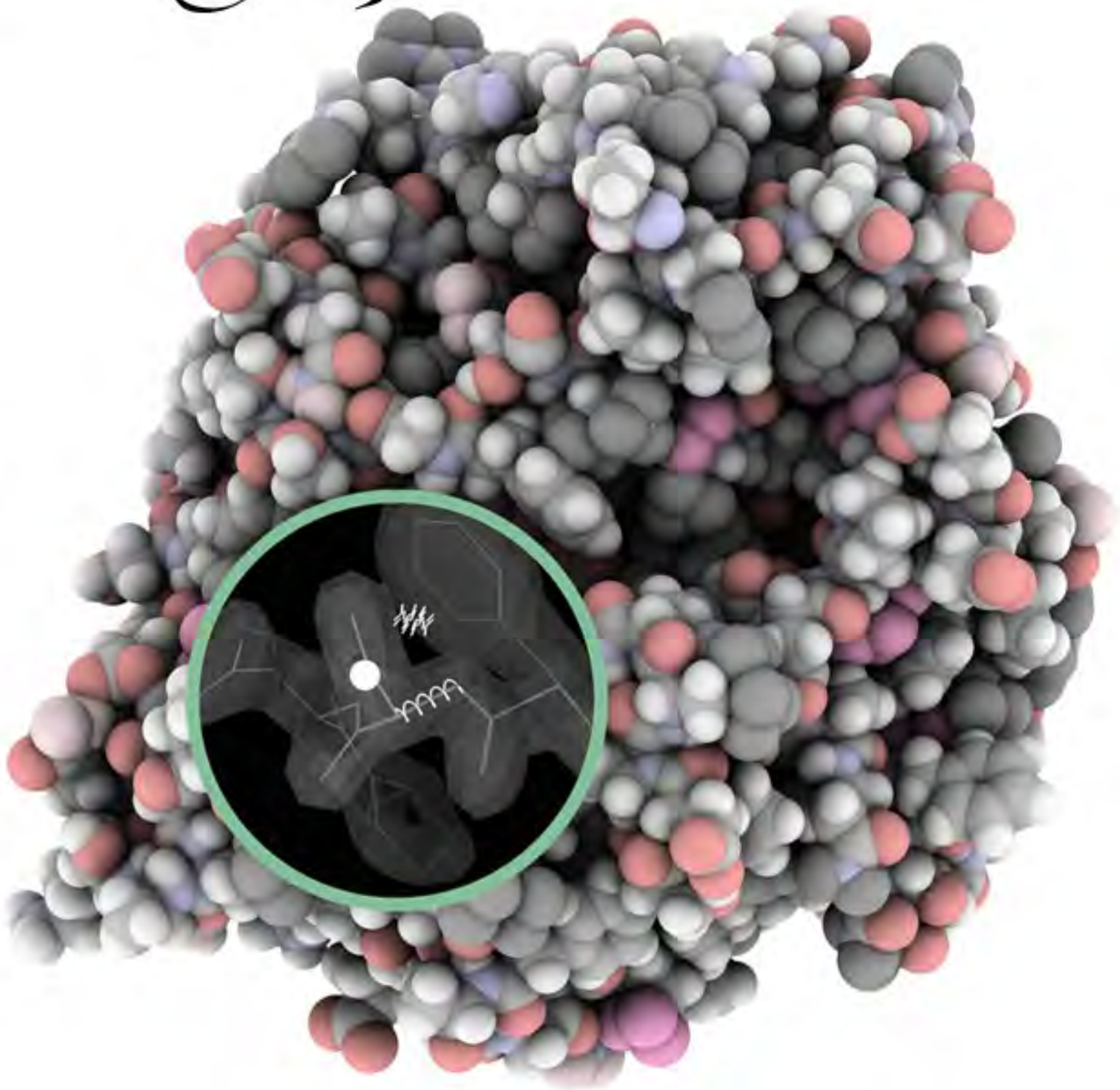
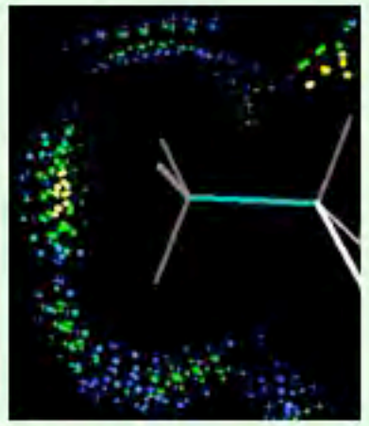
Validating Crystallographic Models – from Diagnosis to Healing

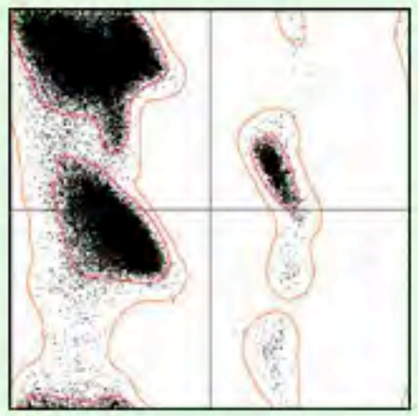


MolProbity Structure Improvement Criteria

All-atom contacts, clashscore



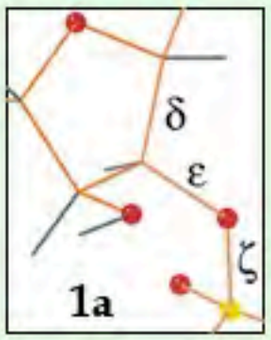
Ramachandran criteria



Sidechain rotamers

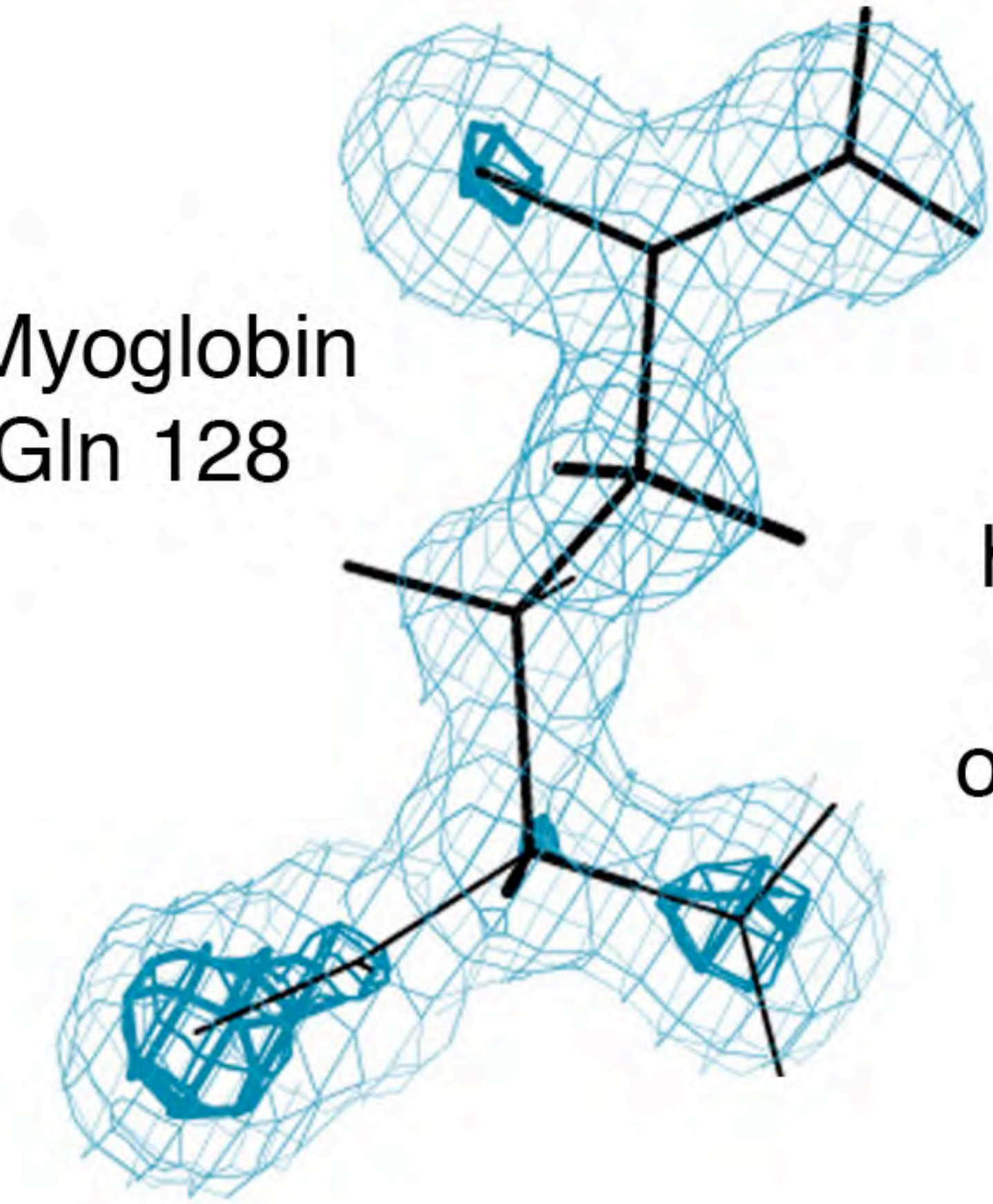
Geometry

RNA bb



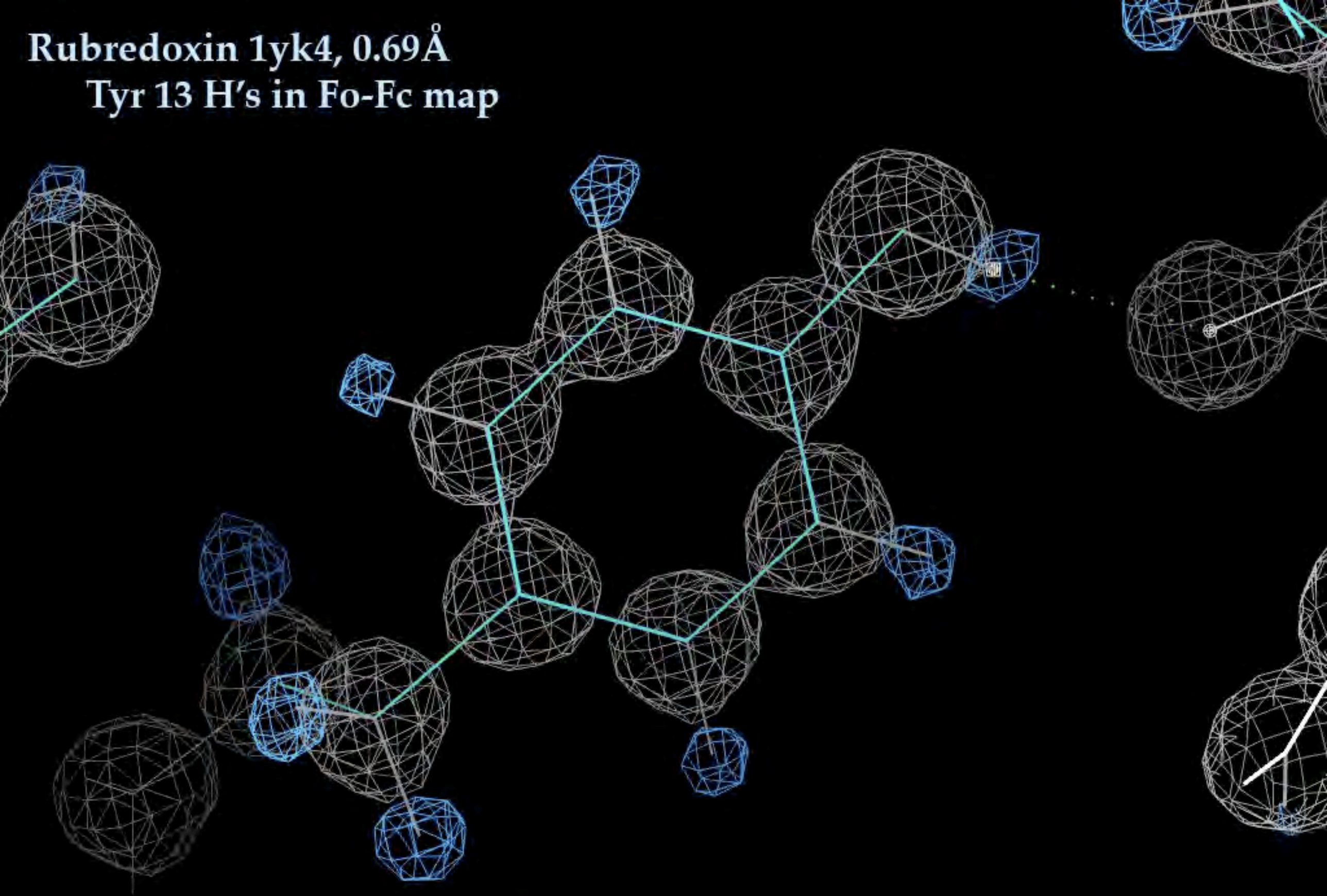
Crystallographic : R_{free}, electron density fit

Myoglobin
Gln 128



hydrogens don't diffract well -- outside of density

Rubredoxin 1yk4, 0.69Å
Tyr 13 H's in Fo-Fc map



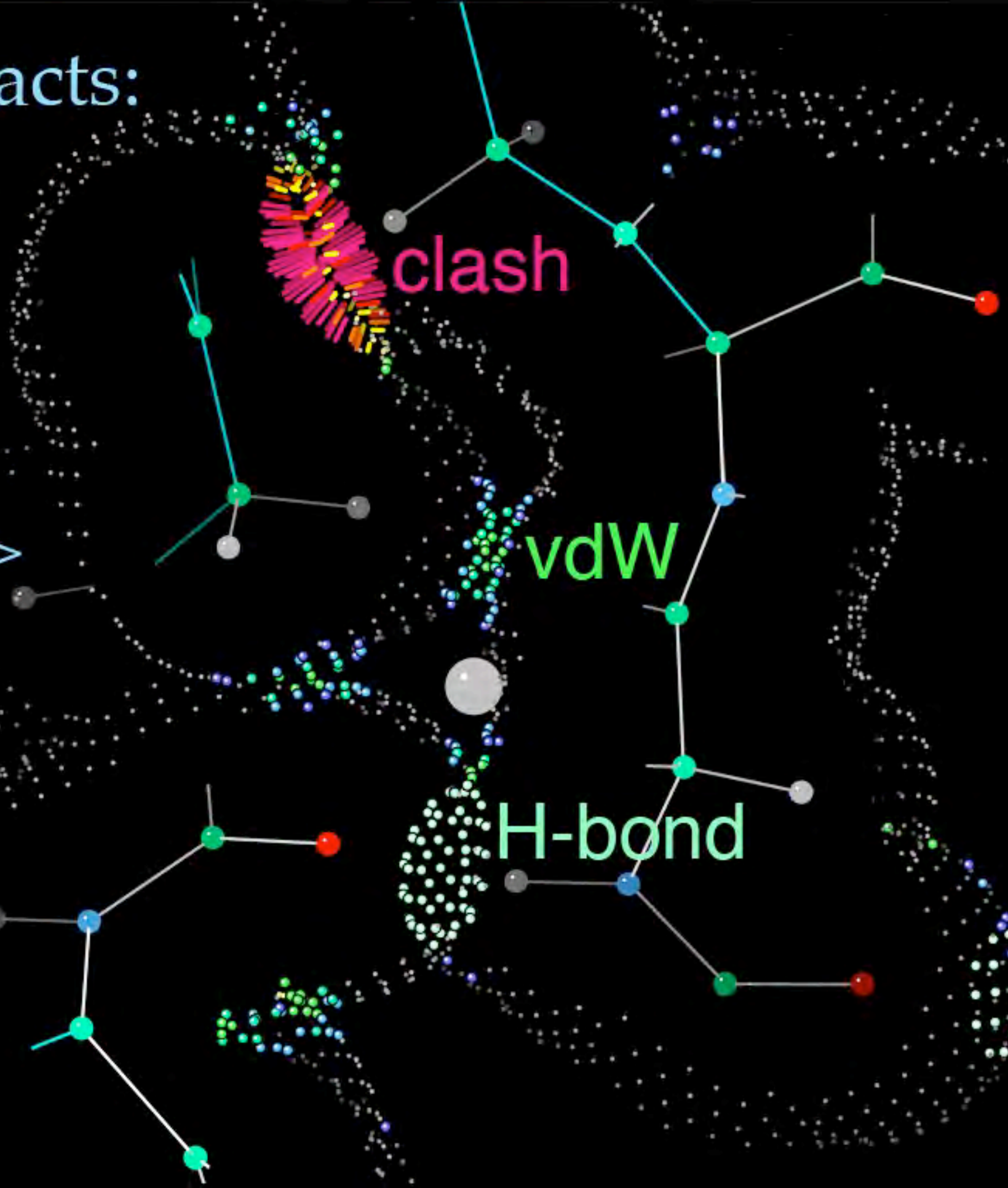
-- but the H atoms are really there!

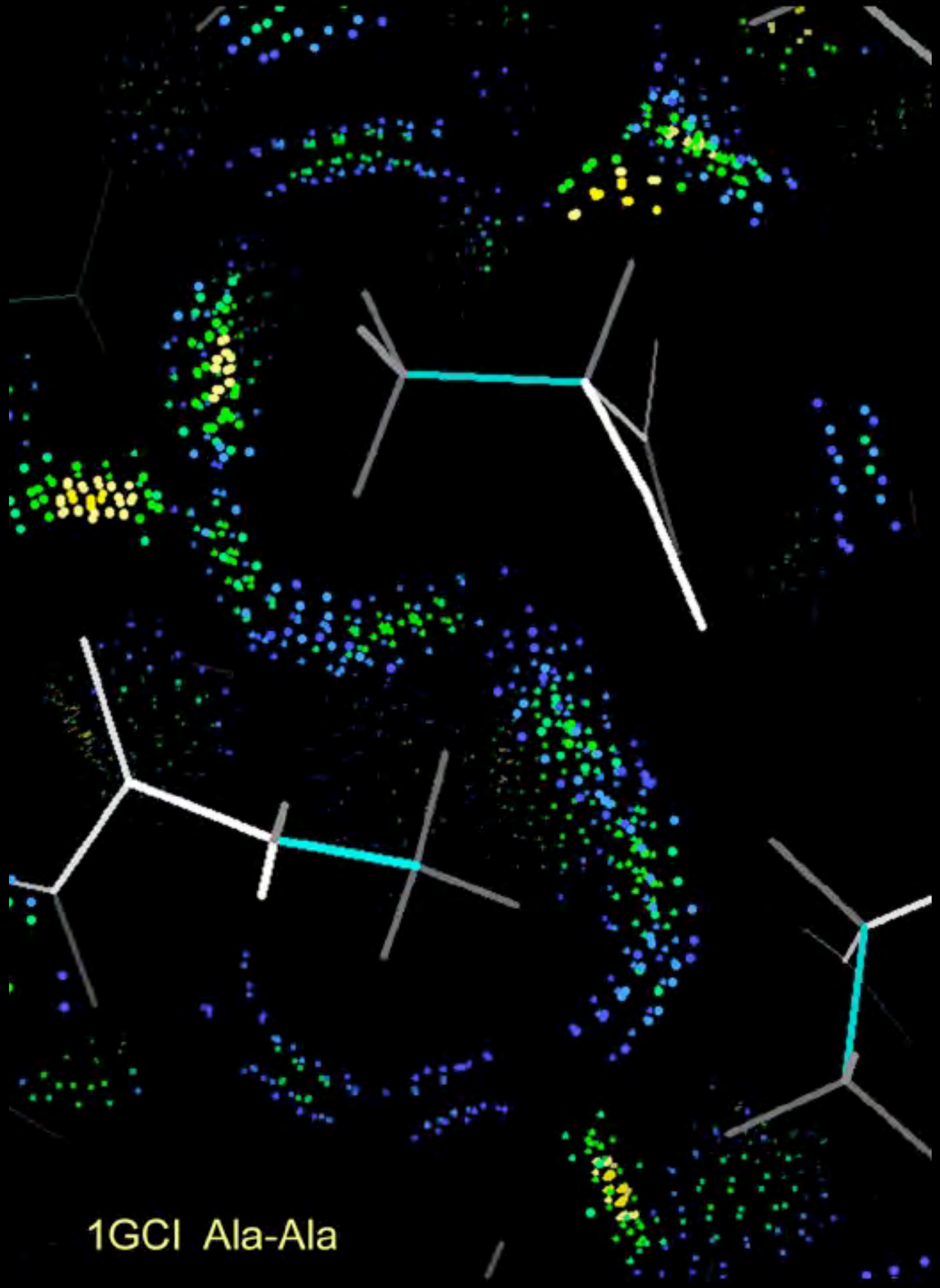
All Atom Contacts:

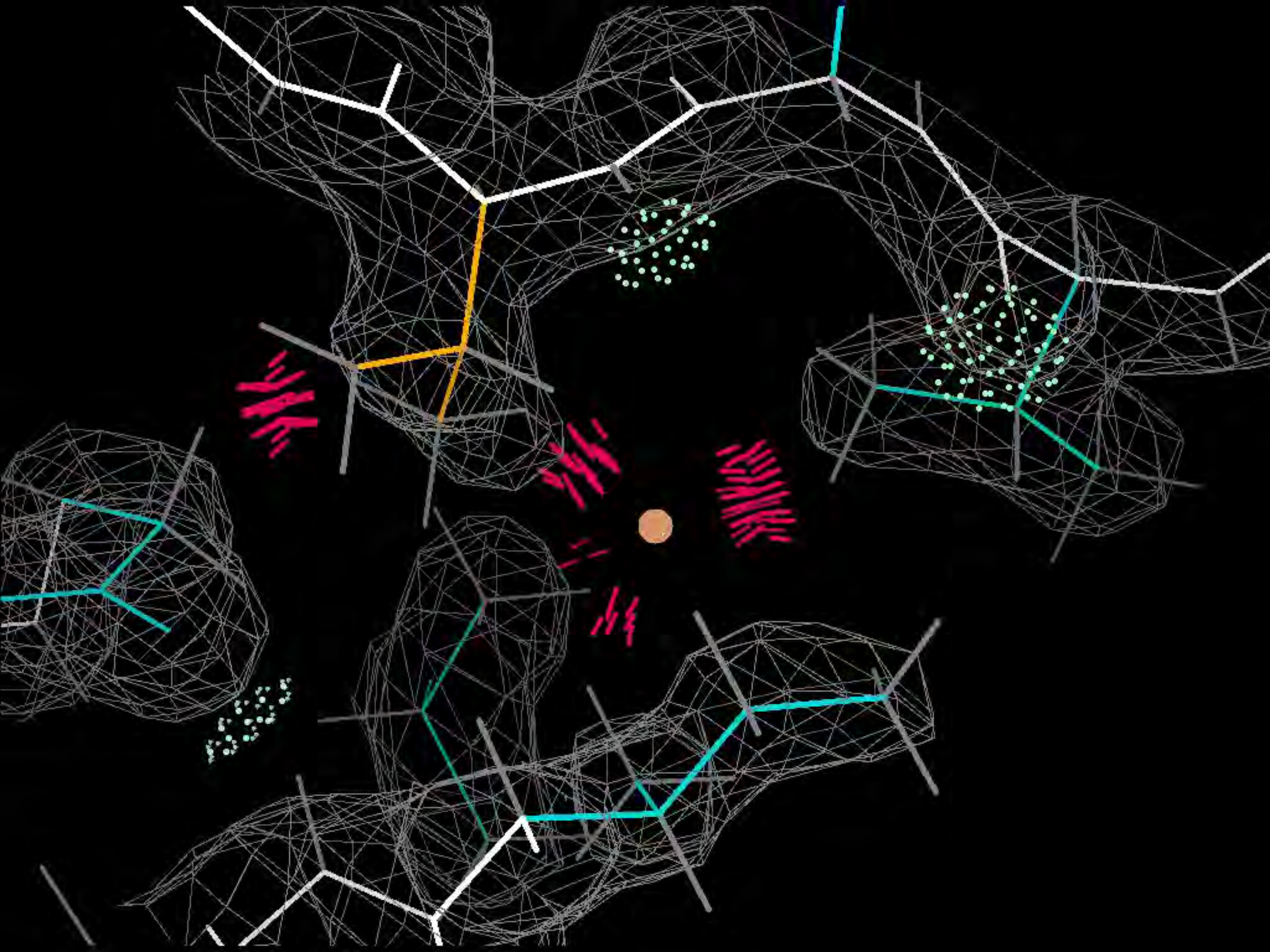
Add H atoms

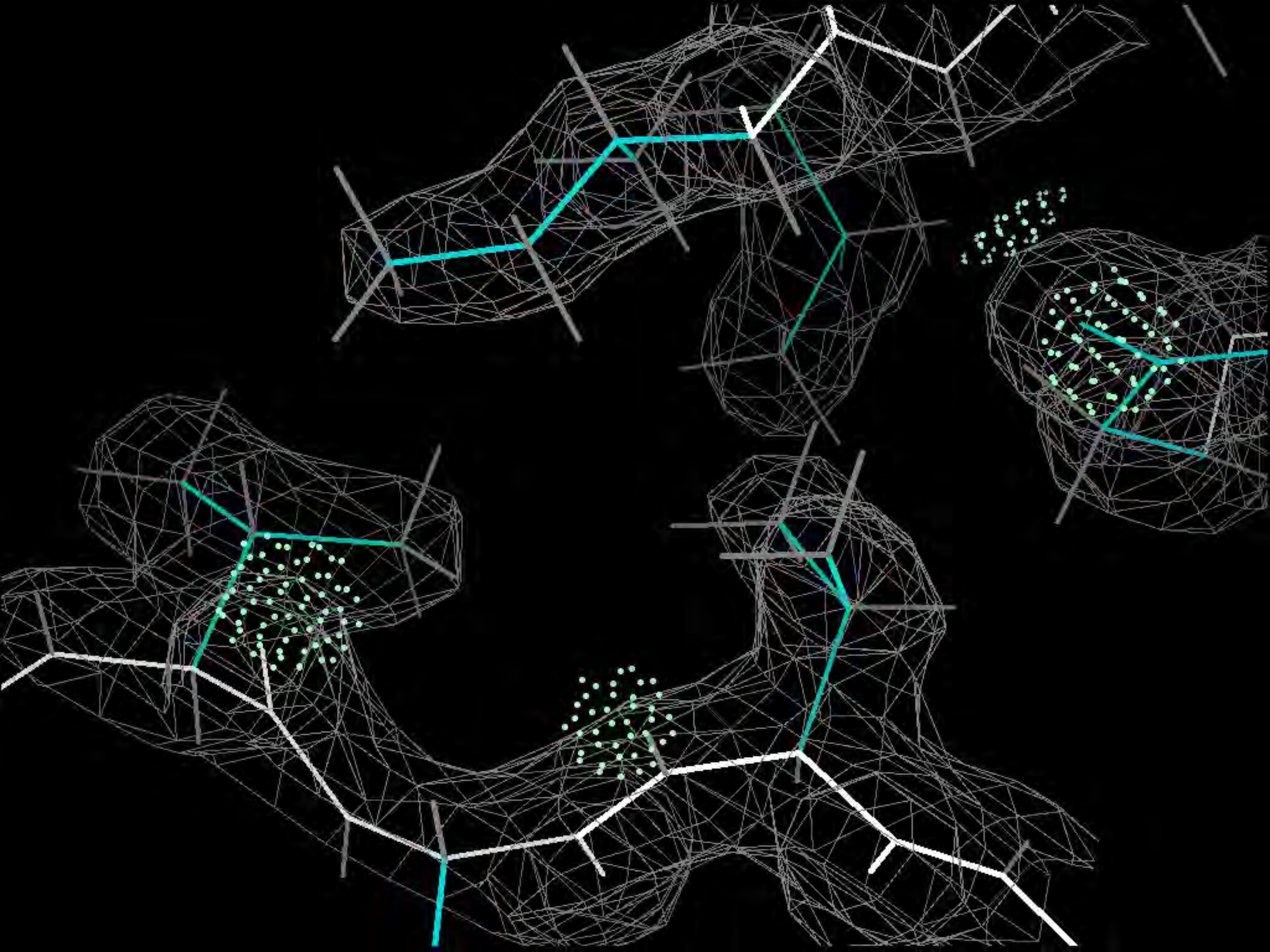
Roll 0.25 Å
radius probe

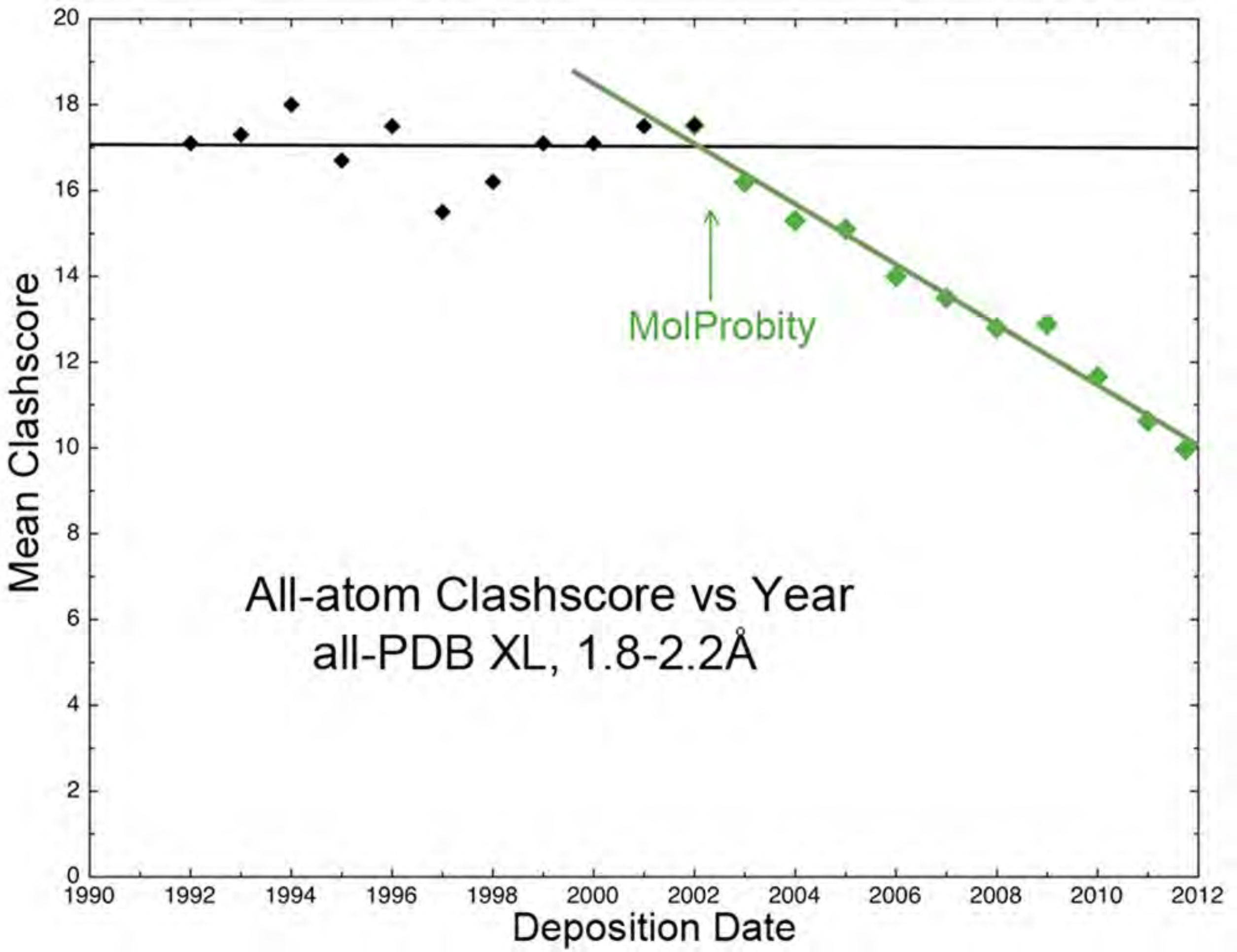
... 3 terms -->



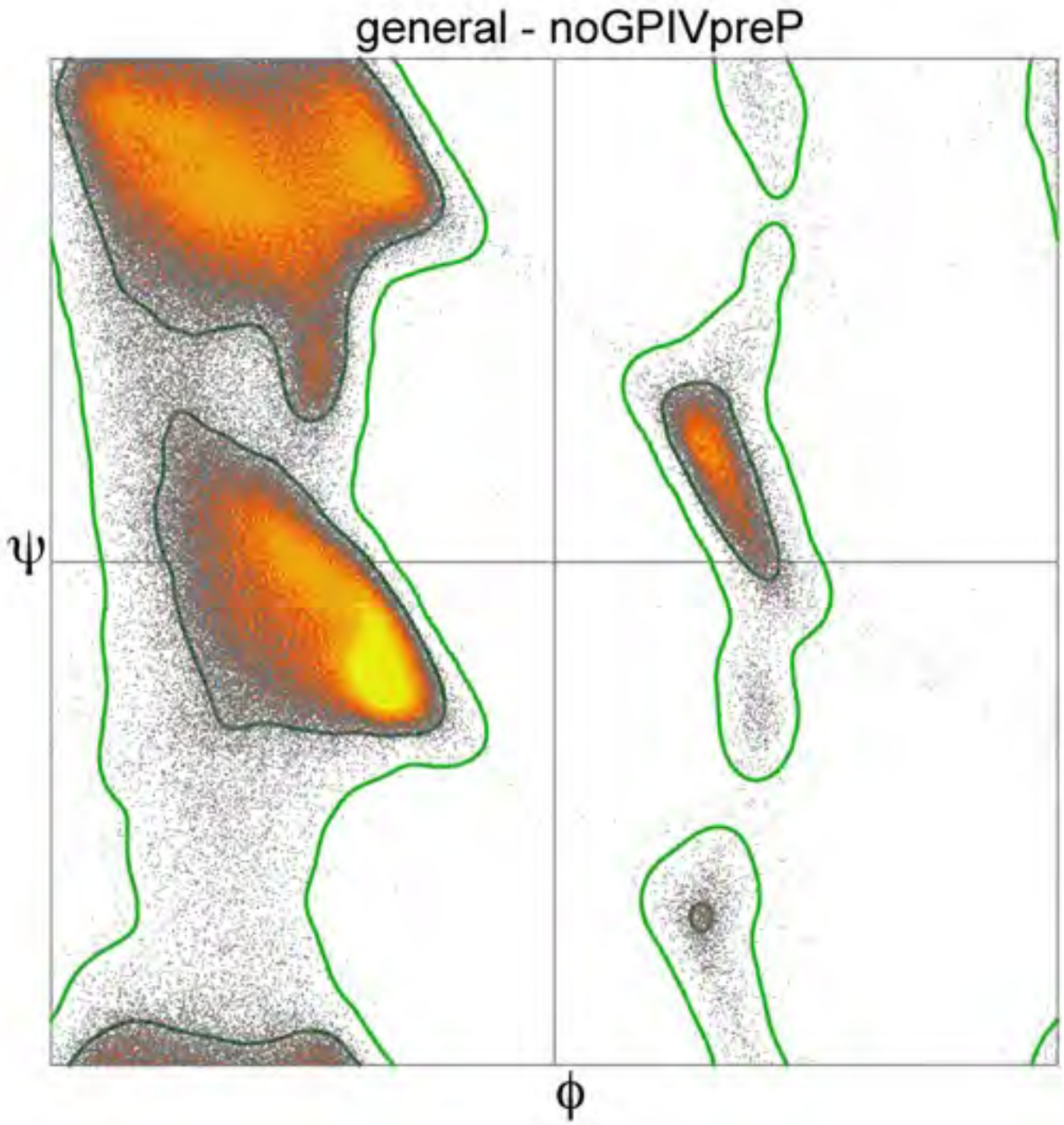

1GCI Ala-Ala

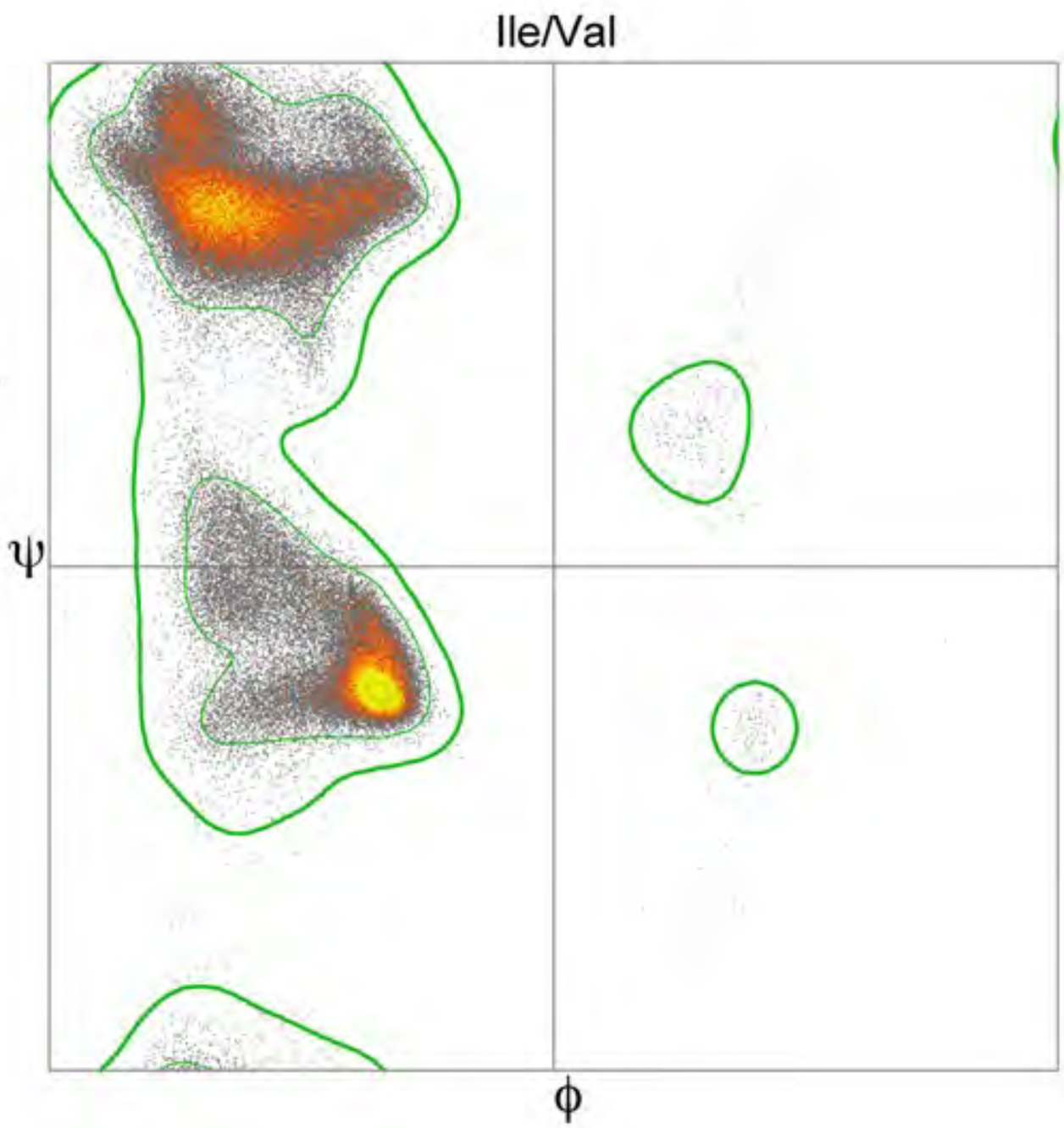





All-atom Clashscore vs Year
all-PDB XL, 1.8-2.2Å
MolProbity
Mean Clashscore
Deposition Date
0
2
4
6
8
10
12
14
16
18
20
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012

Phenix 1.8.2 and MolProbity 4 now out:
with the new H parameters
and 6-category Ramachandran



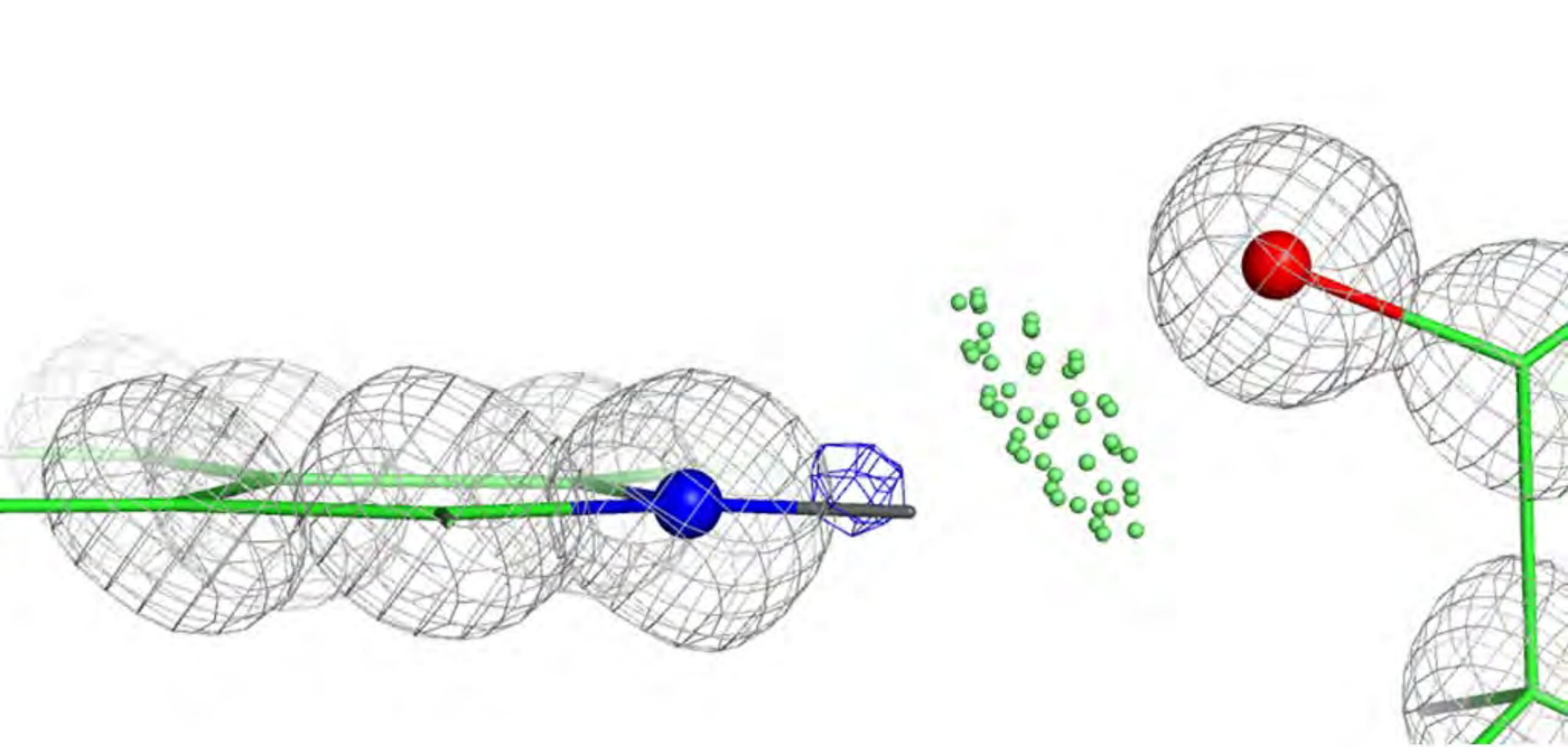


Our starting point is the Hydrogen atom.

H

Consistency for Phenix vs MolProbity

Nuclear vs electron-cloud center positions



H difference peak -- systematic shift toward N

variable shift, sometimes, from environment

Methods -- for x-H positions

(Start from ShelX values)

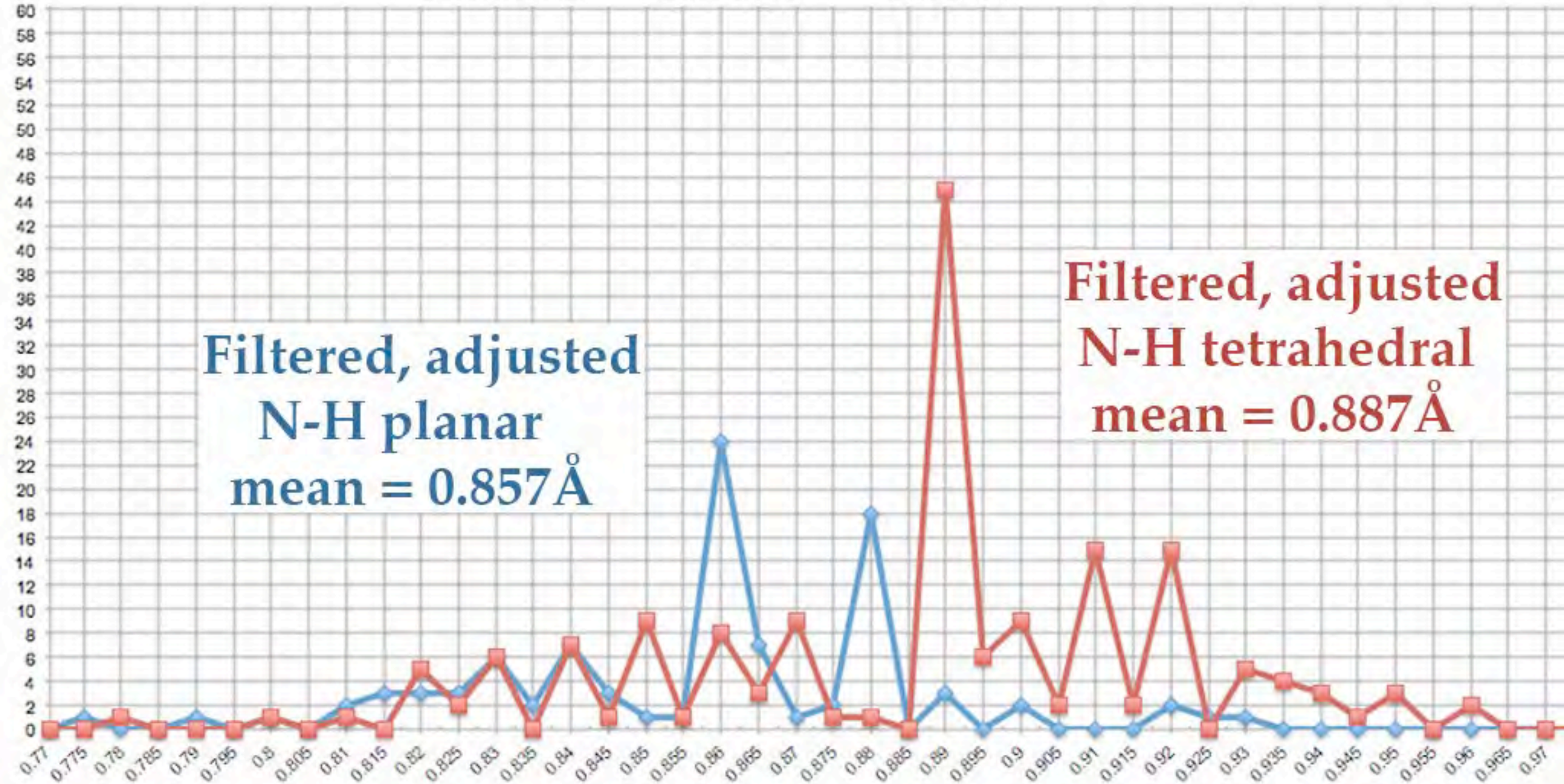
QM density & sphere-fitting
CSD coordinates, x-ray & neutron
PDB high-resolution H difference peaks
COD: with maps, filtered & adjusted

-- for H vdW radii

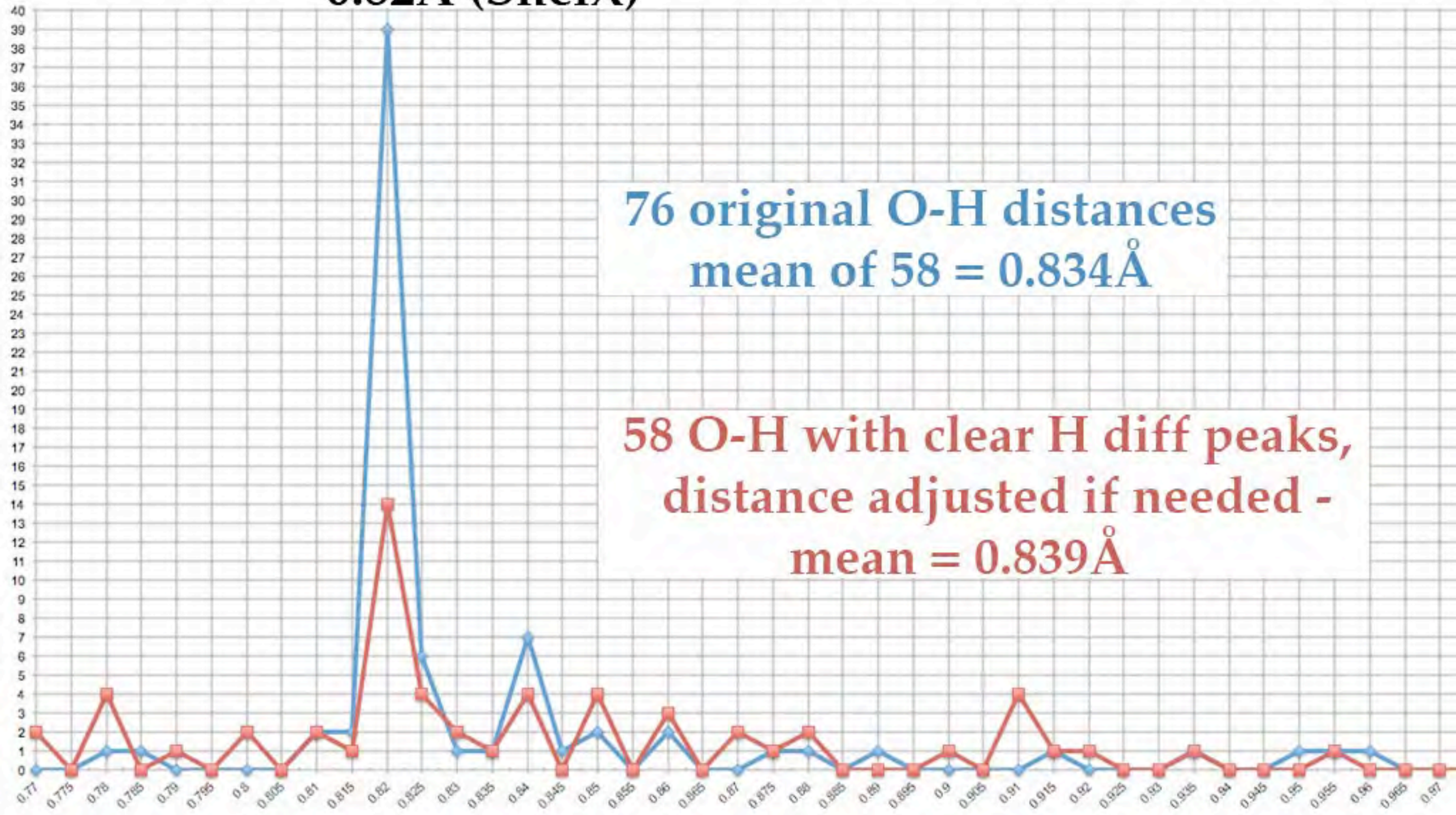
nearest-neighbor atom-pair distributions

COD (Crystallography Open Database):
N-H planar & tetrahedral distances match ShelX

ShelX: 0.86Å 0.89Å



COD (Crystallography Open Database): O-H distances show strong parameter bias



Electron-cloud H

Nuclear H

SHELX

QM

PDB Diff Peaks

CSD X-ray

COD X-ray

CSD Neutron

Old MP

NMR

x-H Distance (Å)

1.4

1.3

1.2

1.1

1

0.9

0.8

0.7

CH tetrahedral

CH planar

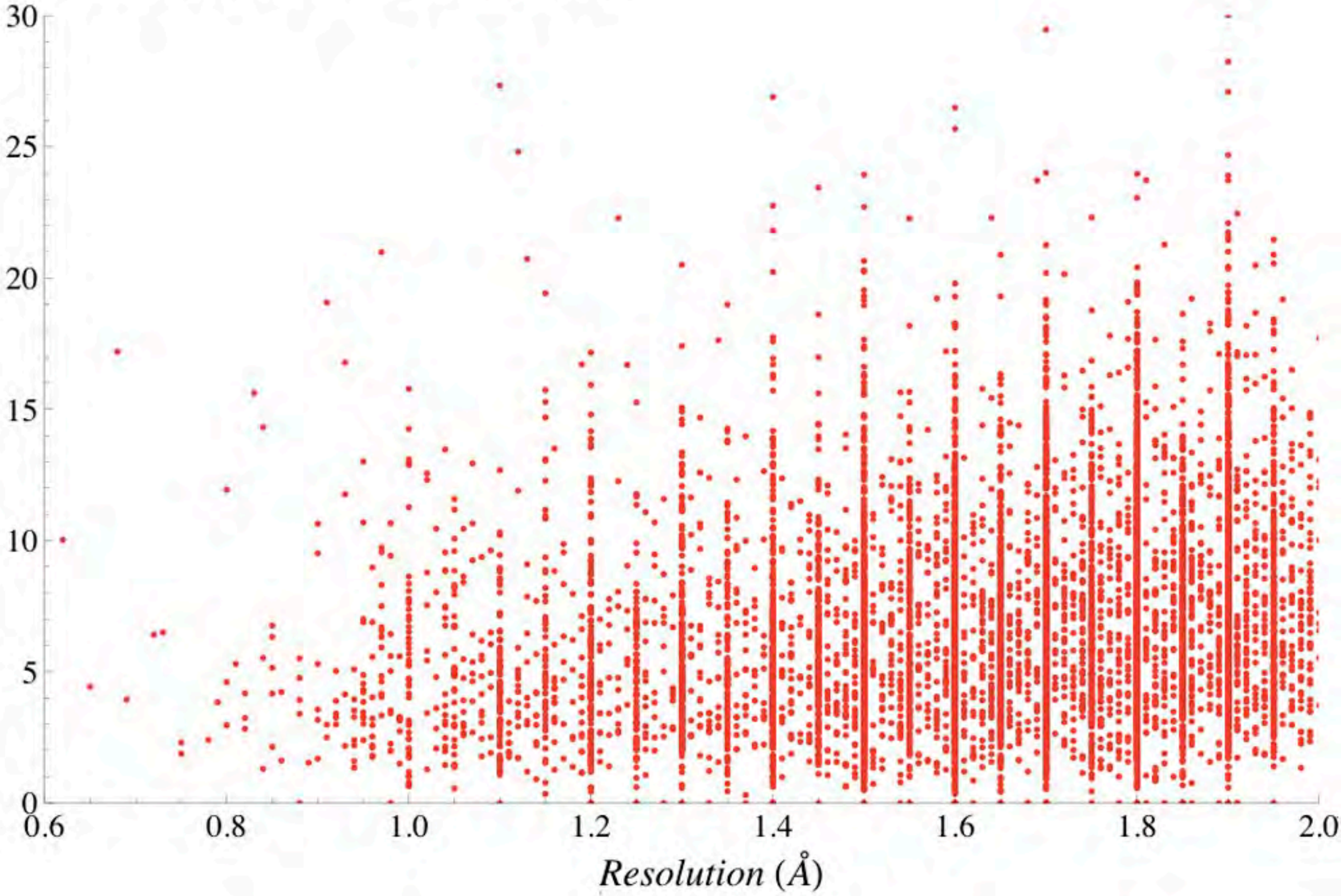
NH tetrahedral

NH planar

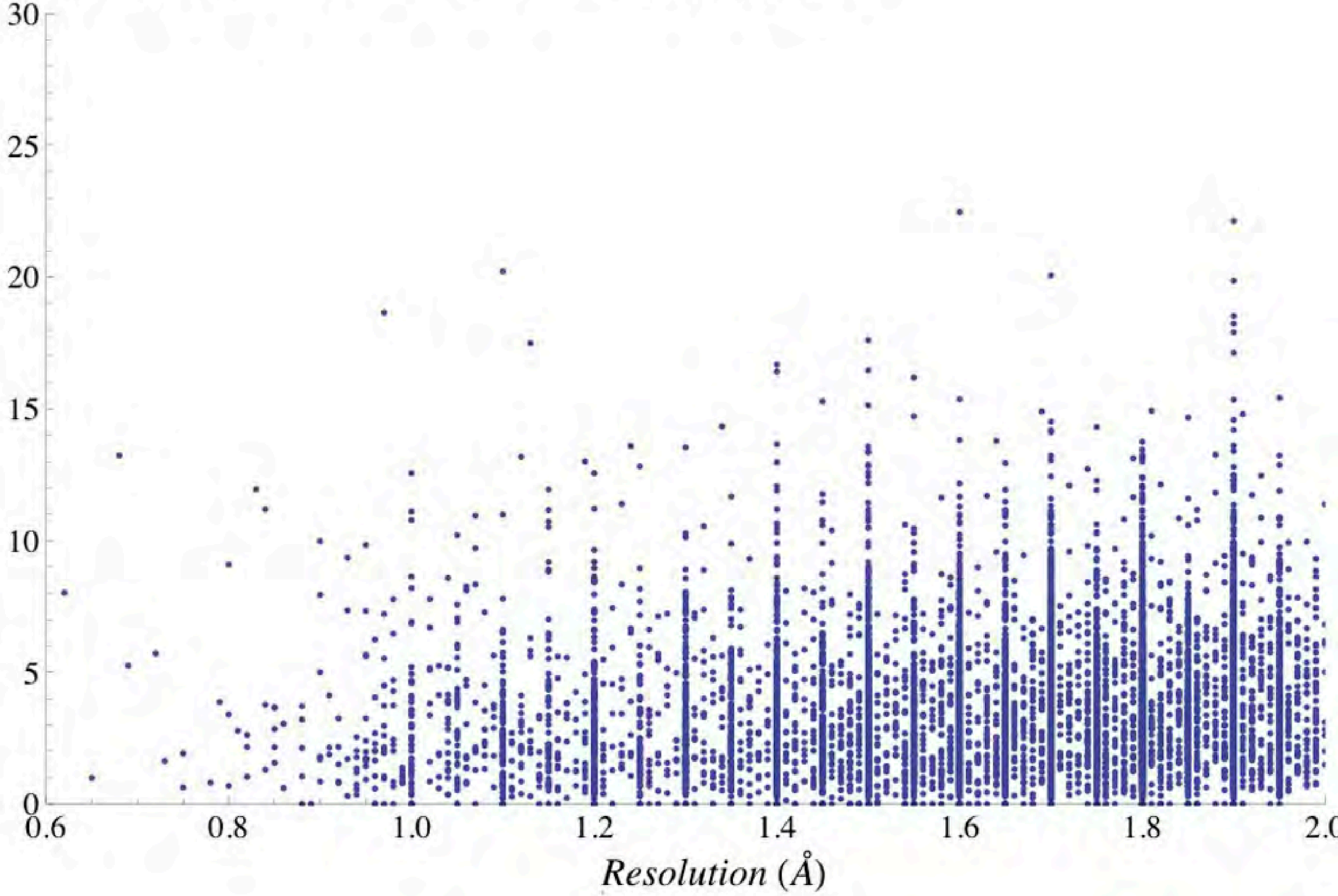
OH

SH

Clashscore, Old -- *too strict?*



Clashscore, New -- better zero asymptote



Outrageous Ambition –

making ALL

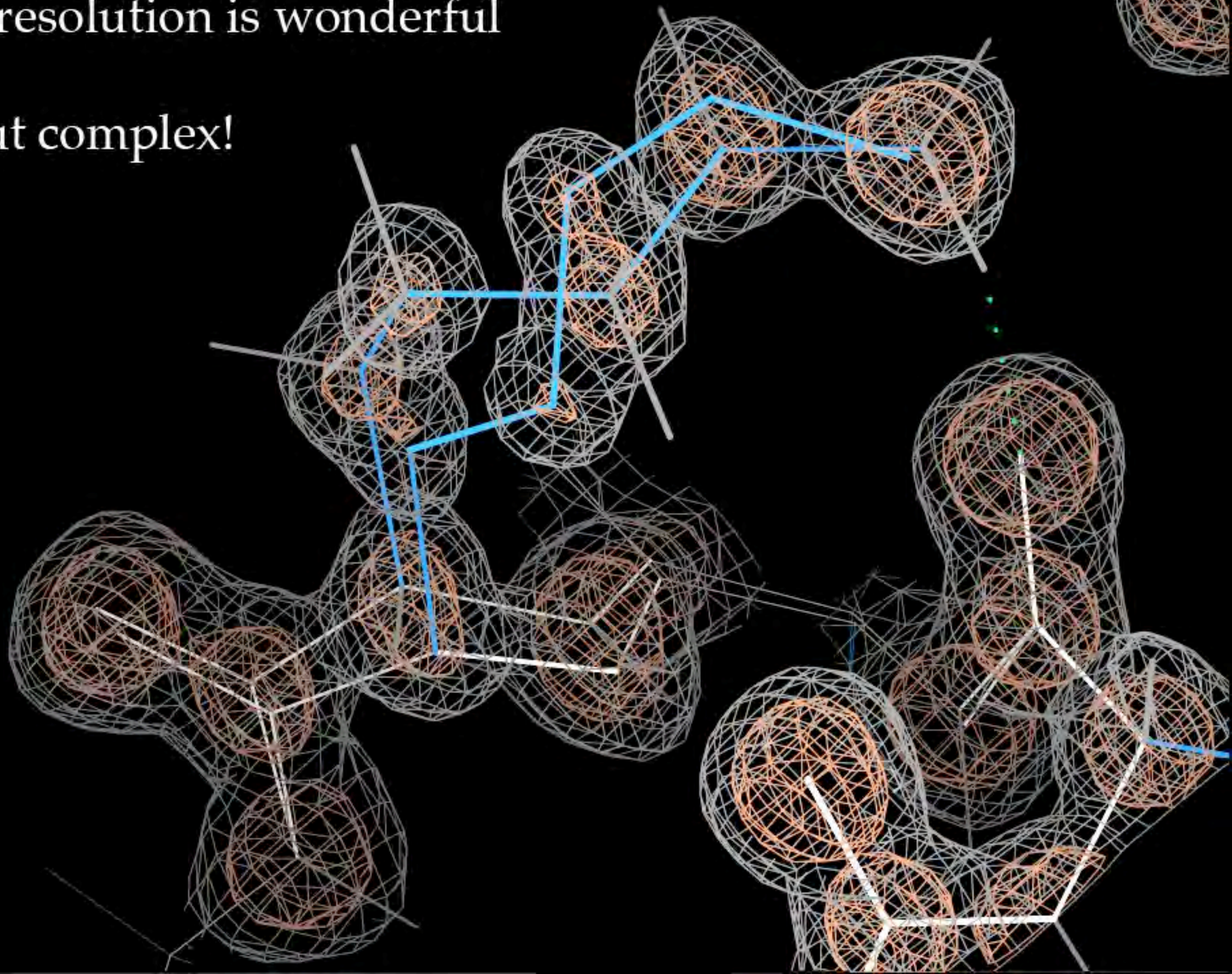
crystal structures much

more ACCURATE:

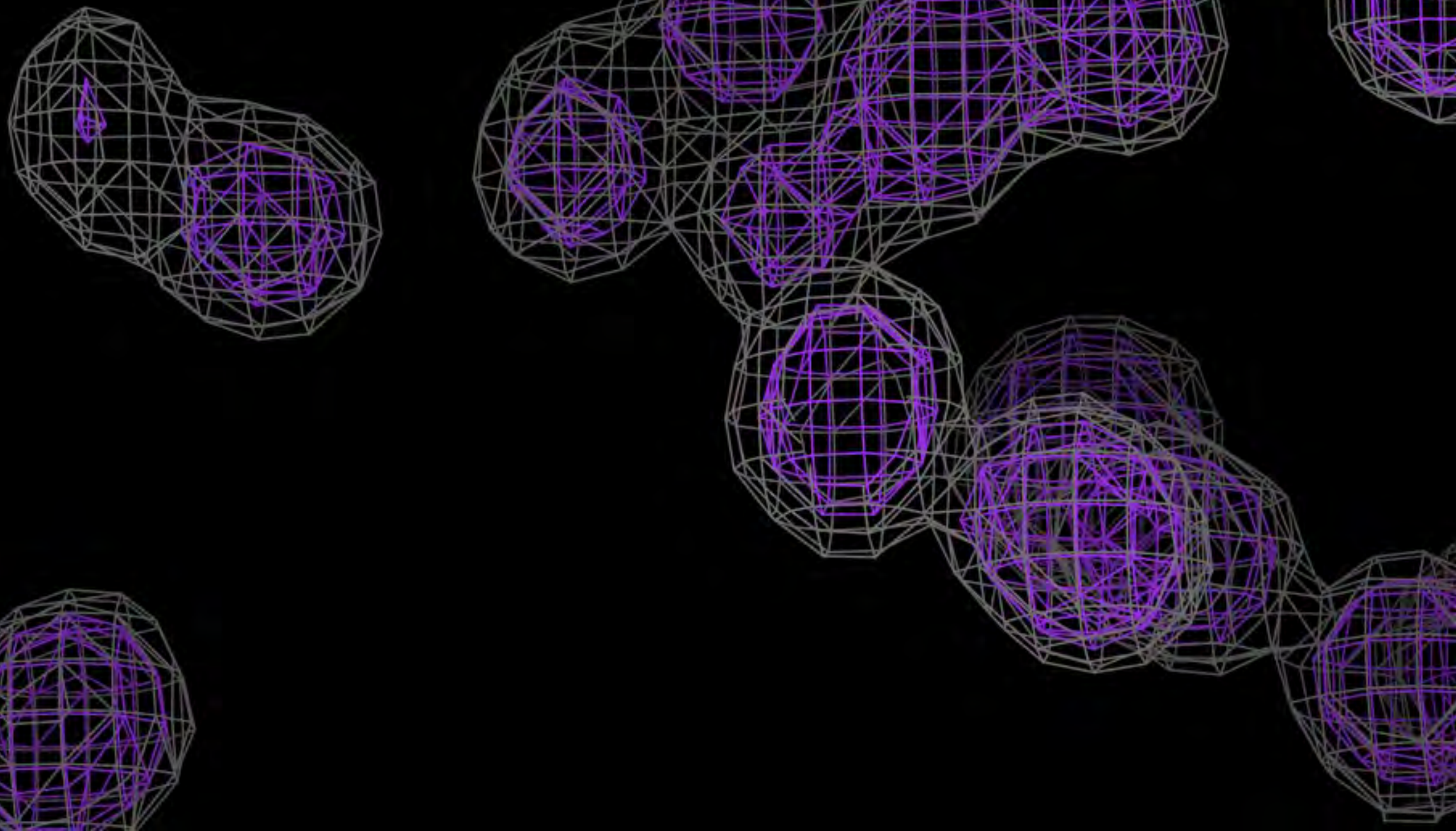
So: low & hi resolution, RNA

High resolution is wonderful

-- but complex!



1gwe, 0.88Å
Leu105 backbone
& waters



1gwe, 0.88Å
Leu105 backbone
& waters

alt A

alt B

1gwe, 0.88Å
Leu105 backbone
& waters

alt B

alt A

alt A

alt B

Low resolution
is ***HARD***

1U87, 3.5Å

Leu
243

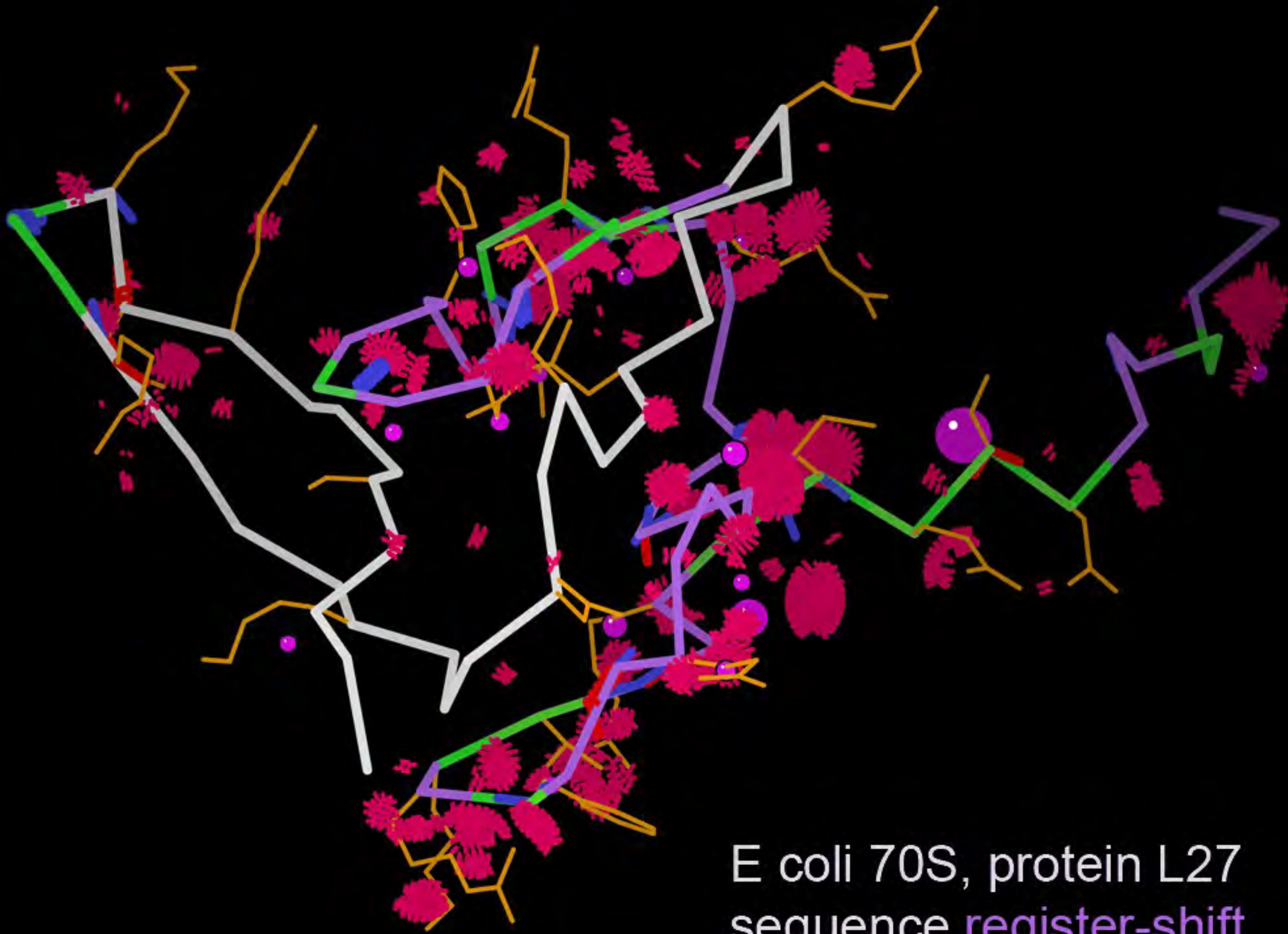
Gly

3fpn 1.8Å

Leu
243

Gly

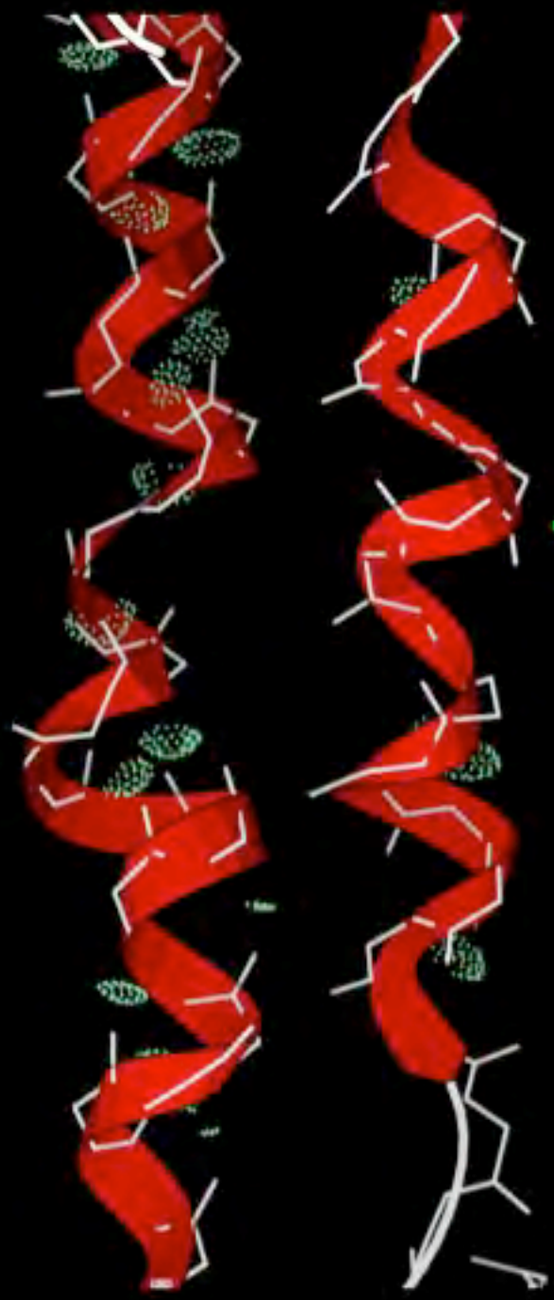
2r6f 3.2Å

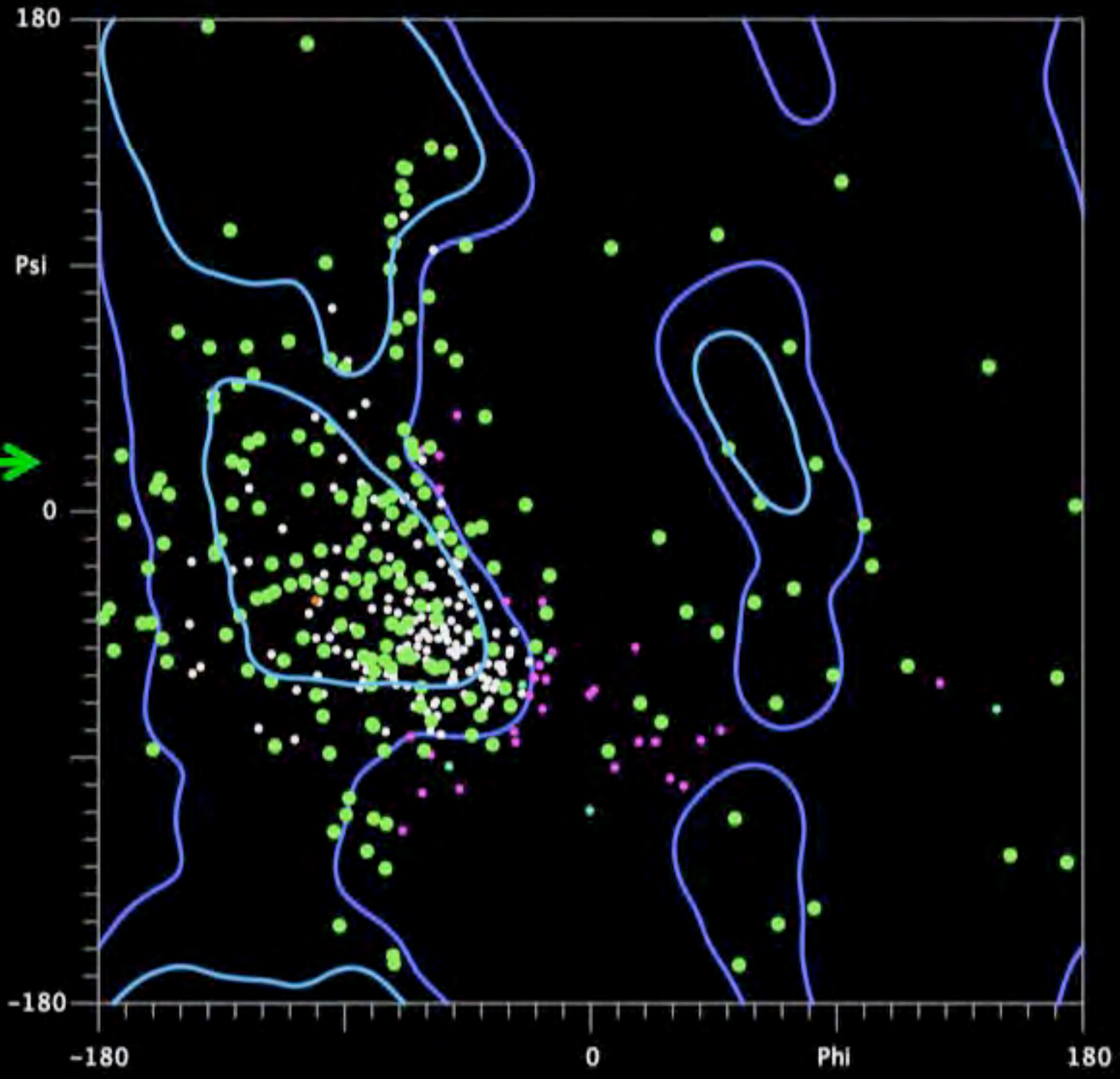

E coli 70S, protein L27
sequence register-shift

E coli 70S, protein L27
after correction: 3r8s

Secondary Structure Diagnosis

Ramachandran is unreliable at low resolution

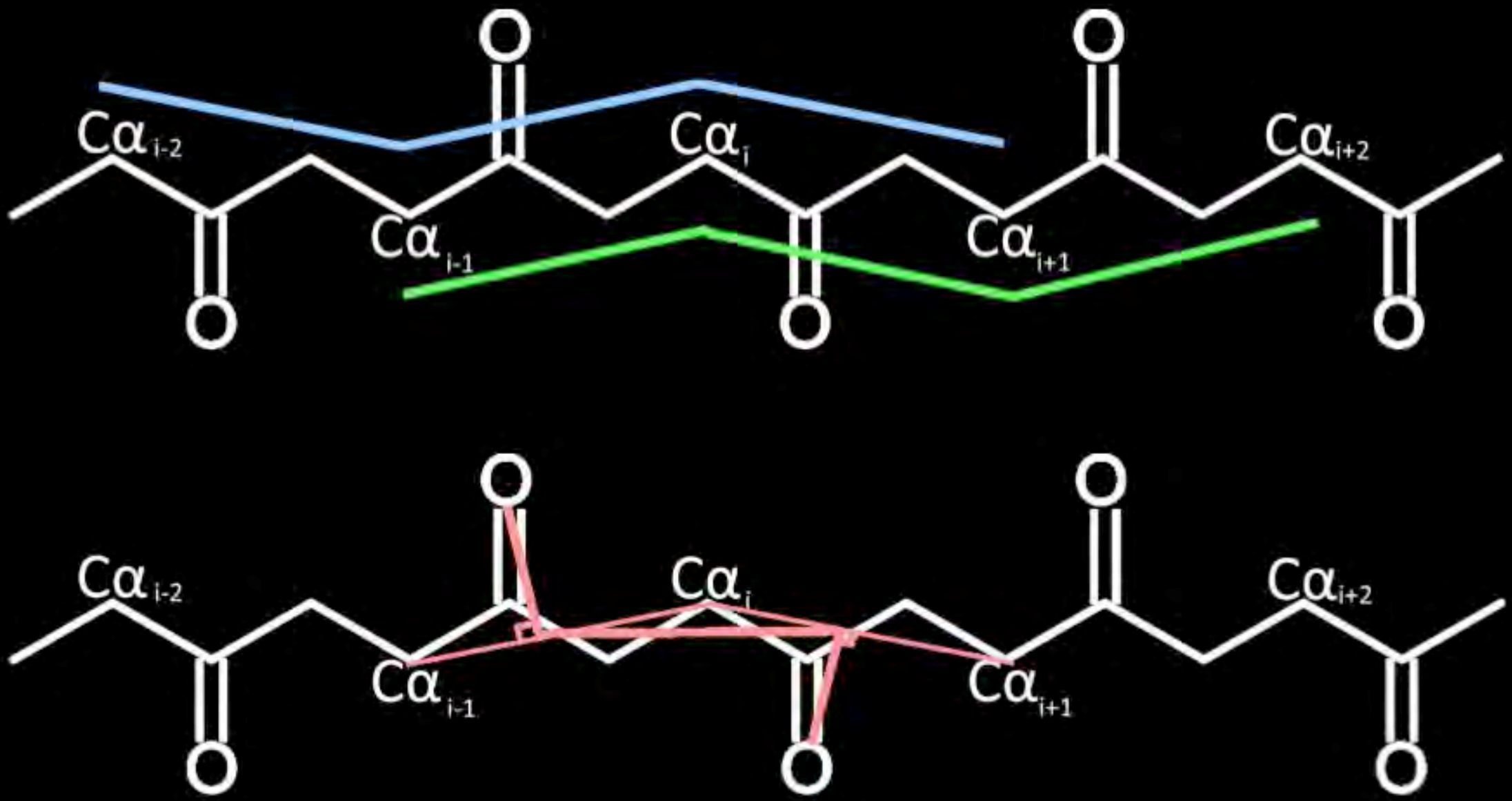




2o01: Photosystem I, 3.4Å

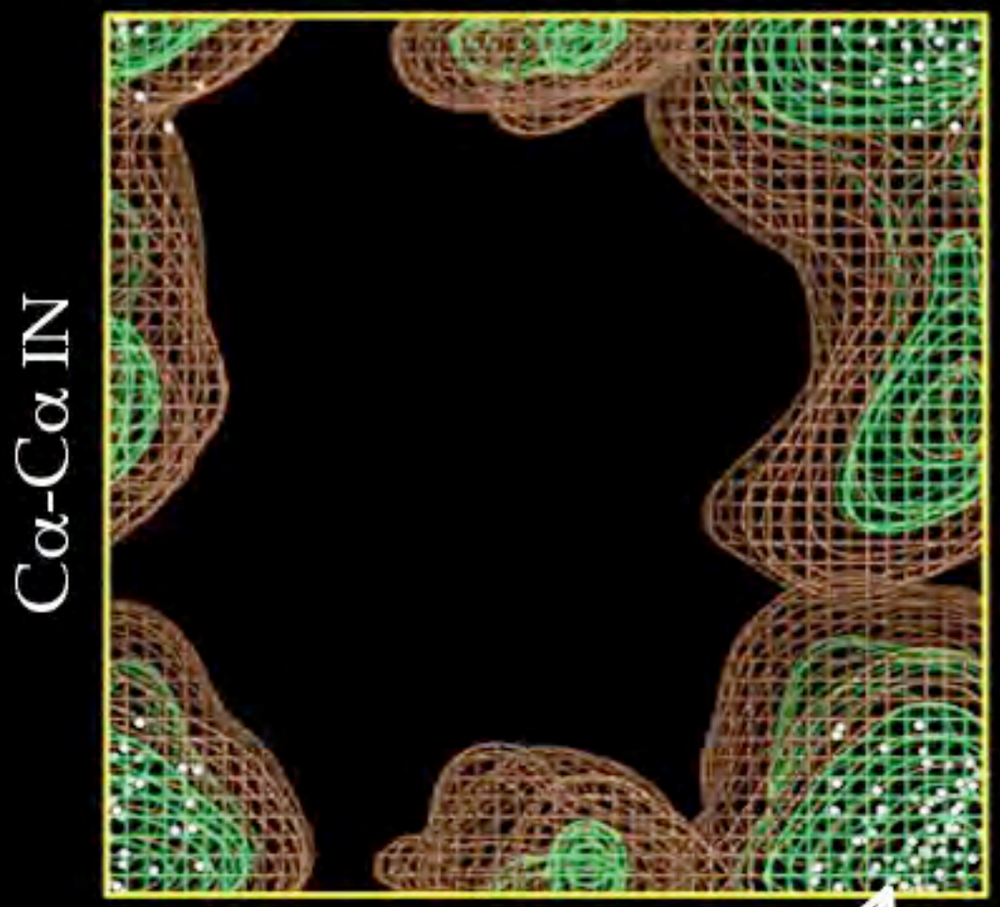
New CaBLAM Parameter Space

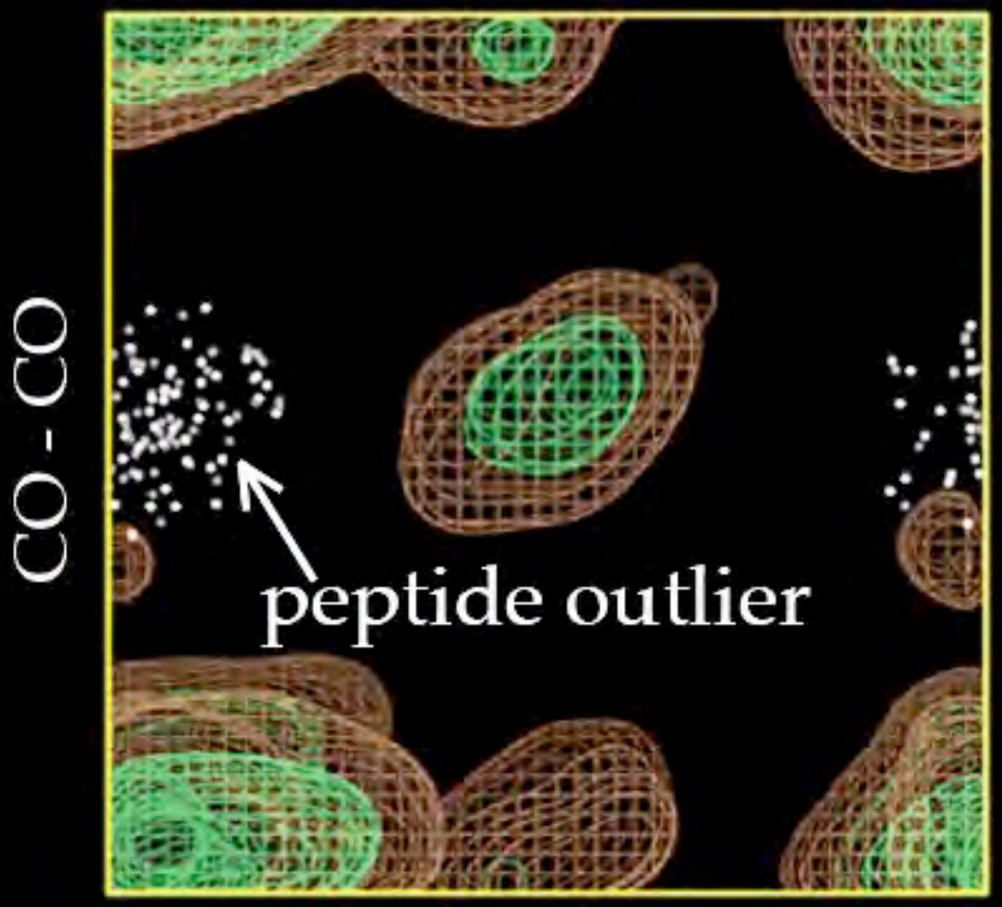
A minimalist alternative



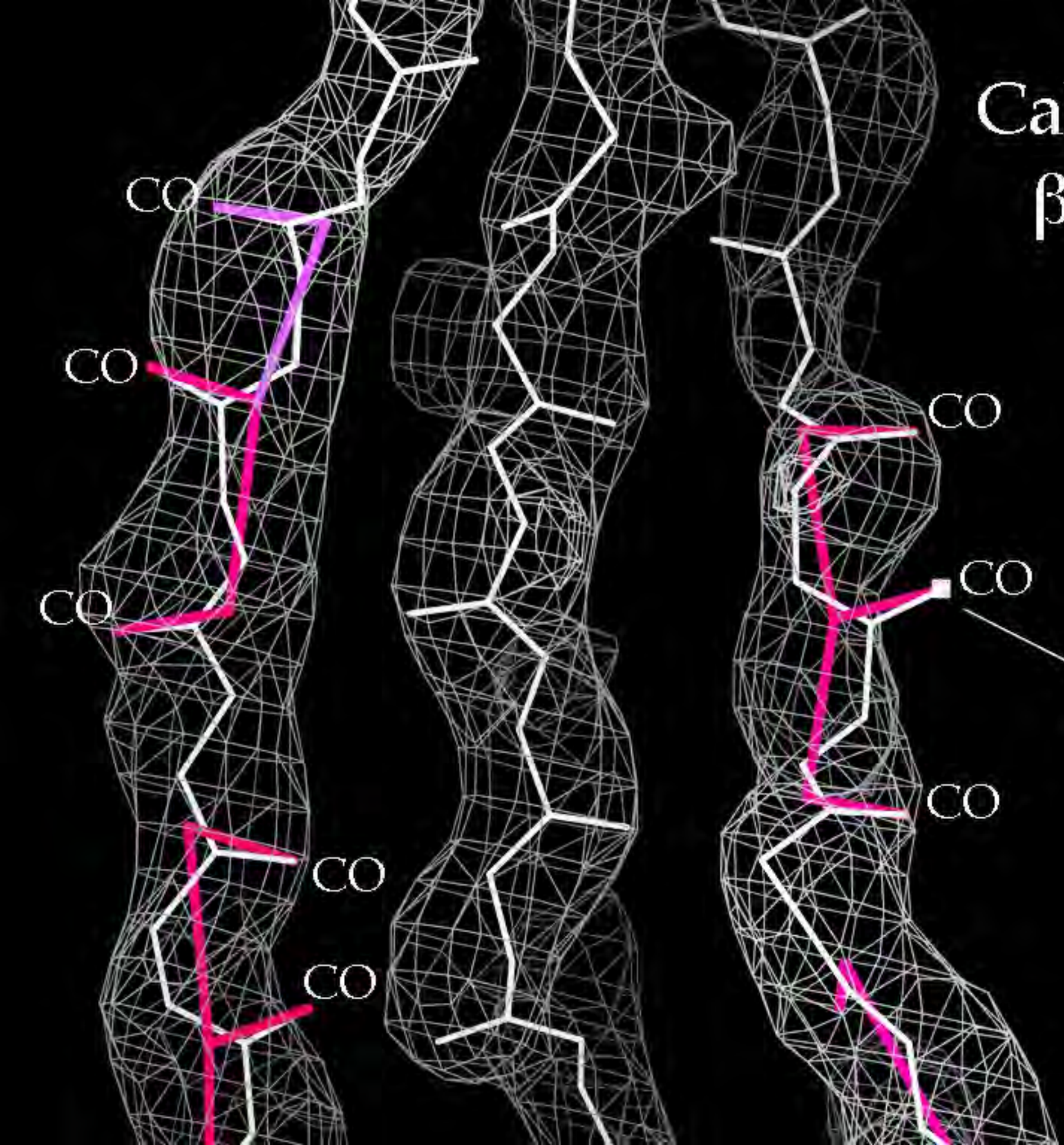
Diagnosing Strands

Pathological strands from 70S Ribosome





CabLAM-space β validation

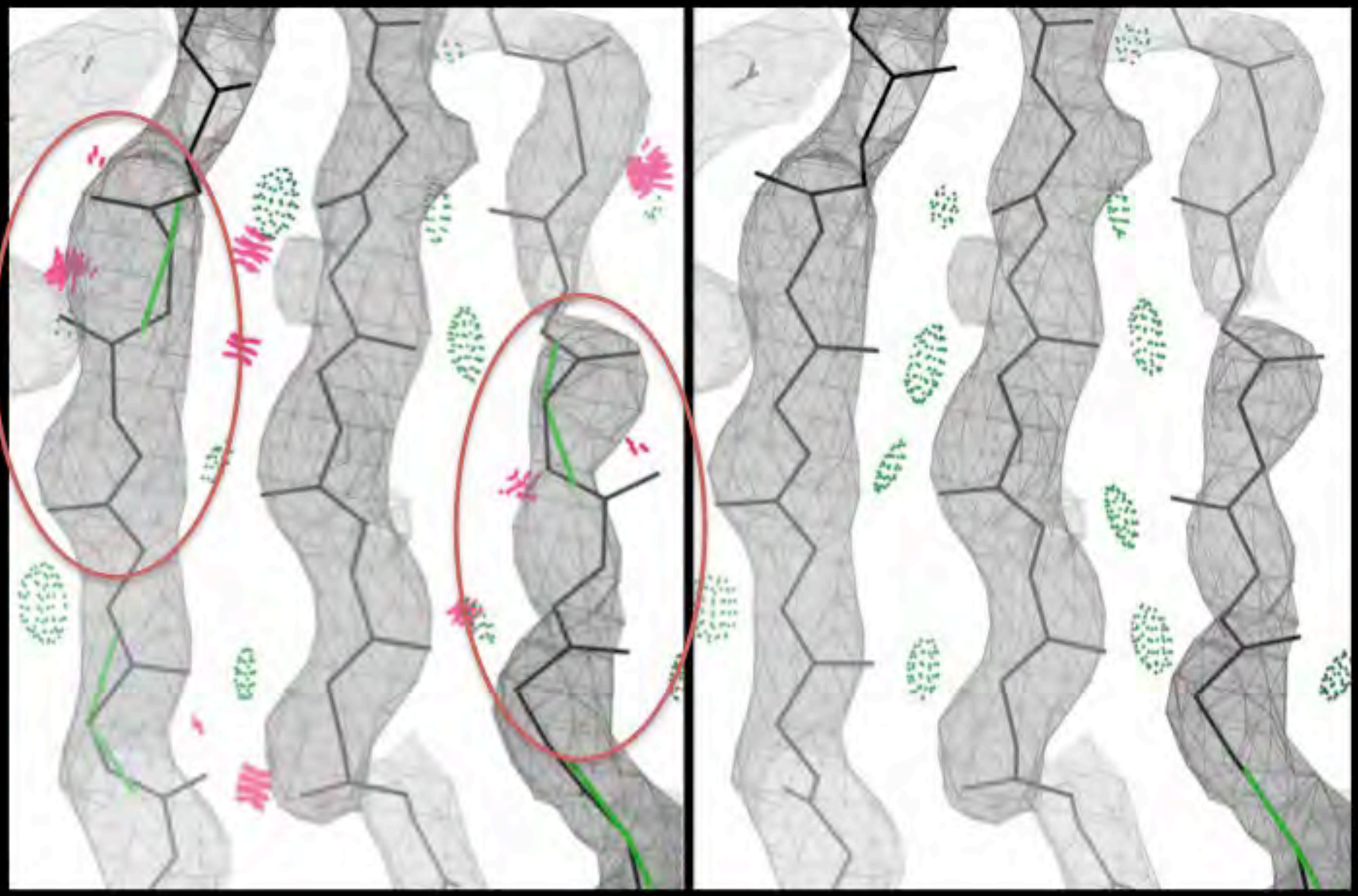


general = 0.7%
(outlier)

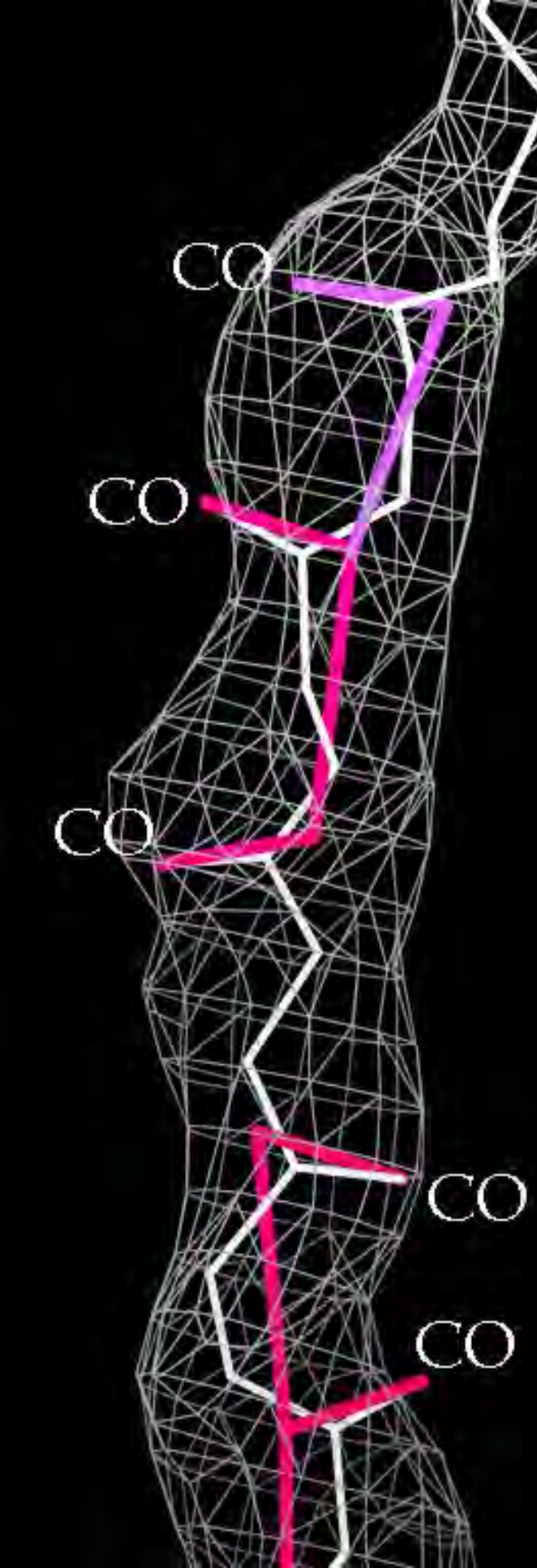
β = 85.8%

α = 0%

Healing a CaBLAM Diagnosis



beta-sheet errors from the E.coli ribosome (L6):
3 carbonyls in the same direction



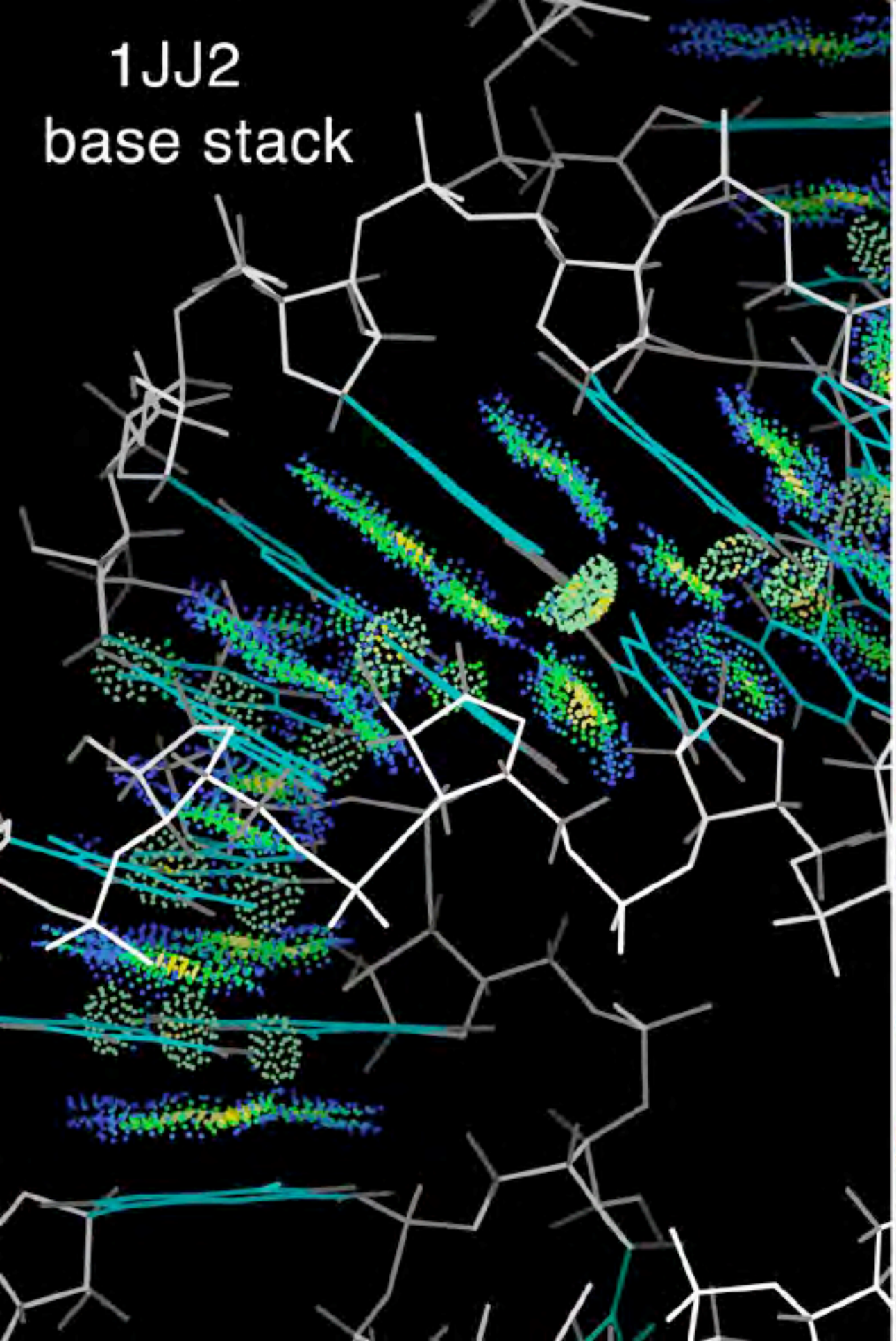
CabLAM validation now available
to try out on command-line
in latest Phenix release

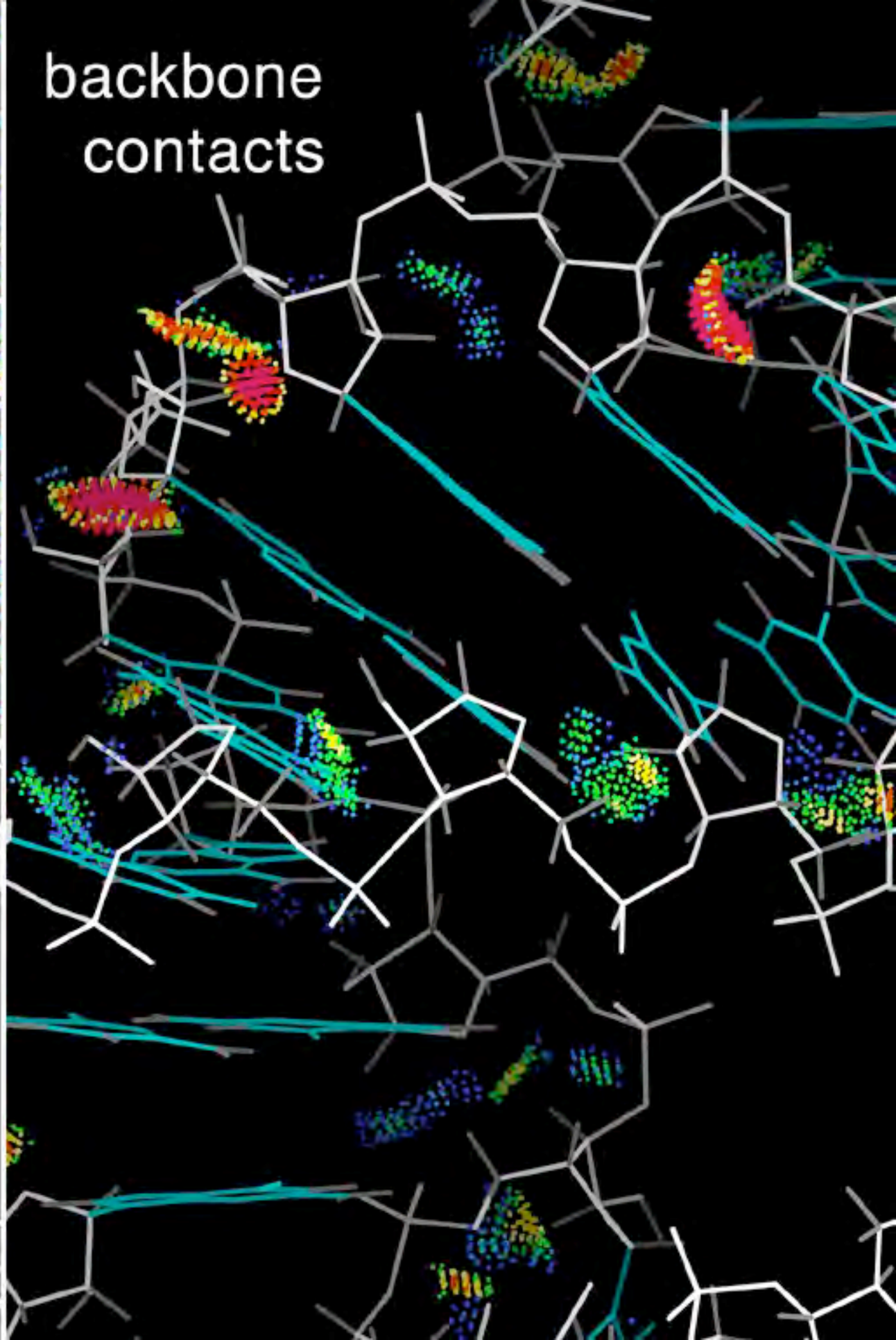
Diagnosis of “disguised”
secondary structure for
manual (not yet automatic) fix-up

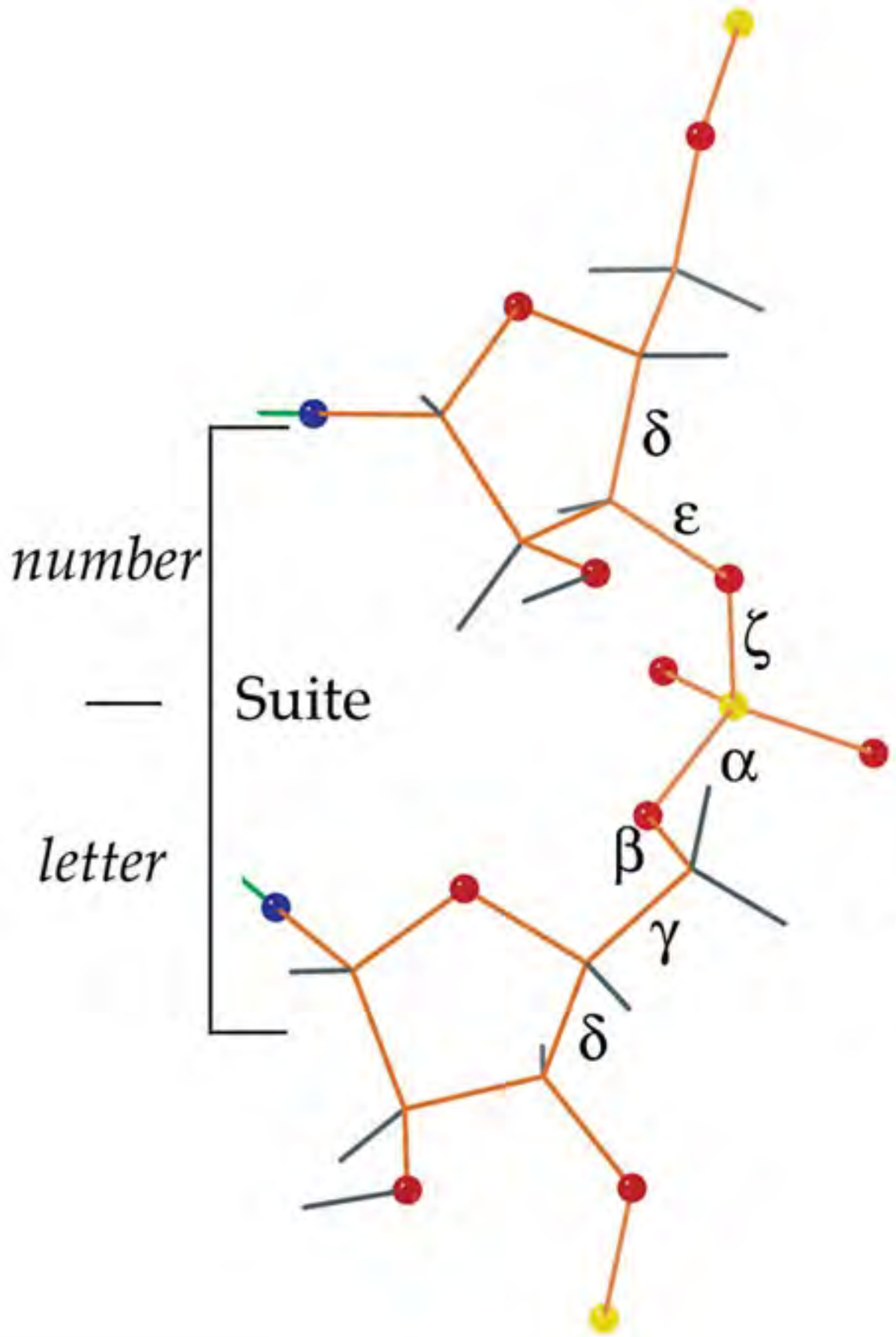
sarcin loop
2Fo-Fc
maps

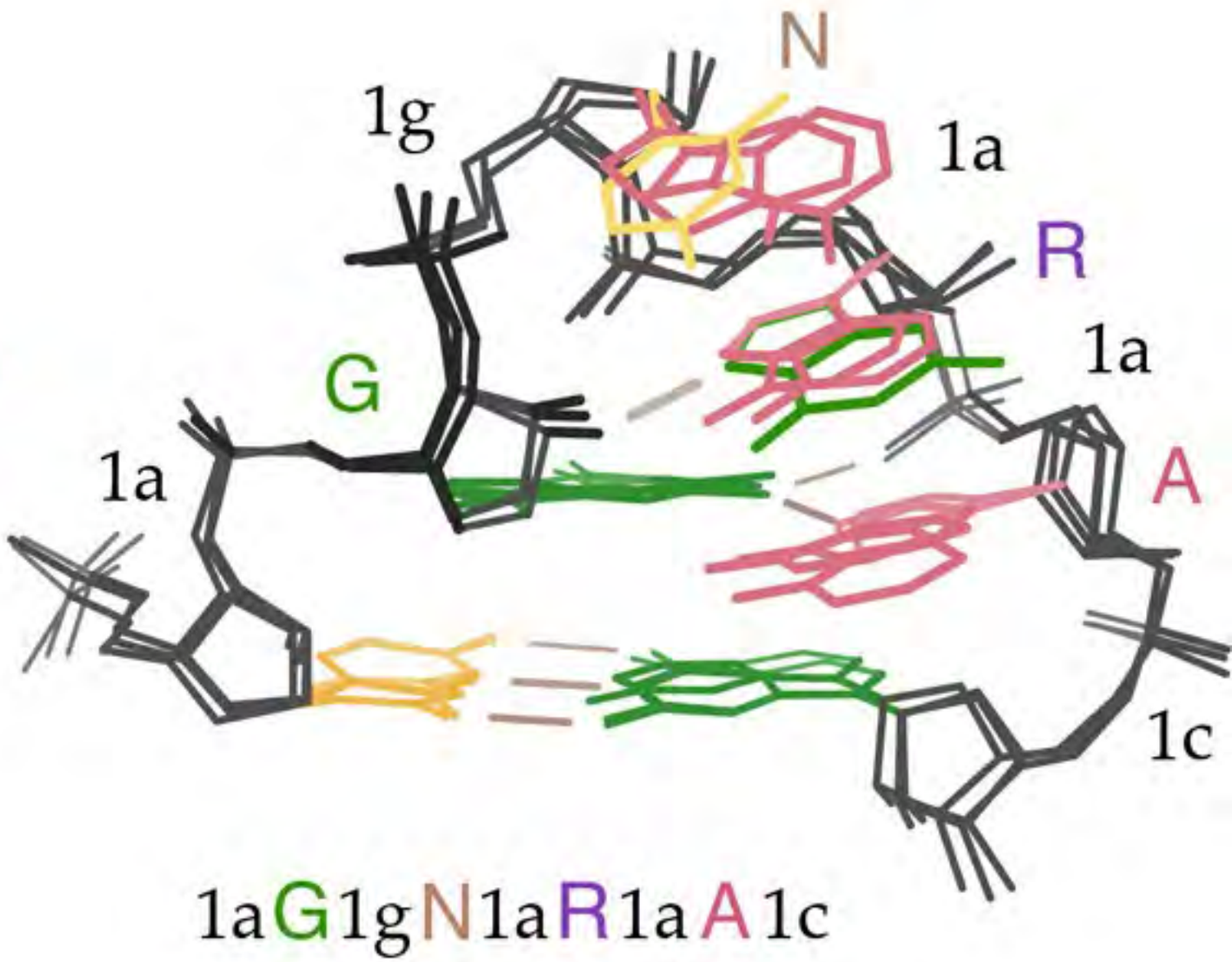
1.04Å

2.4Å

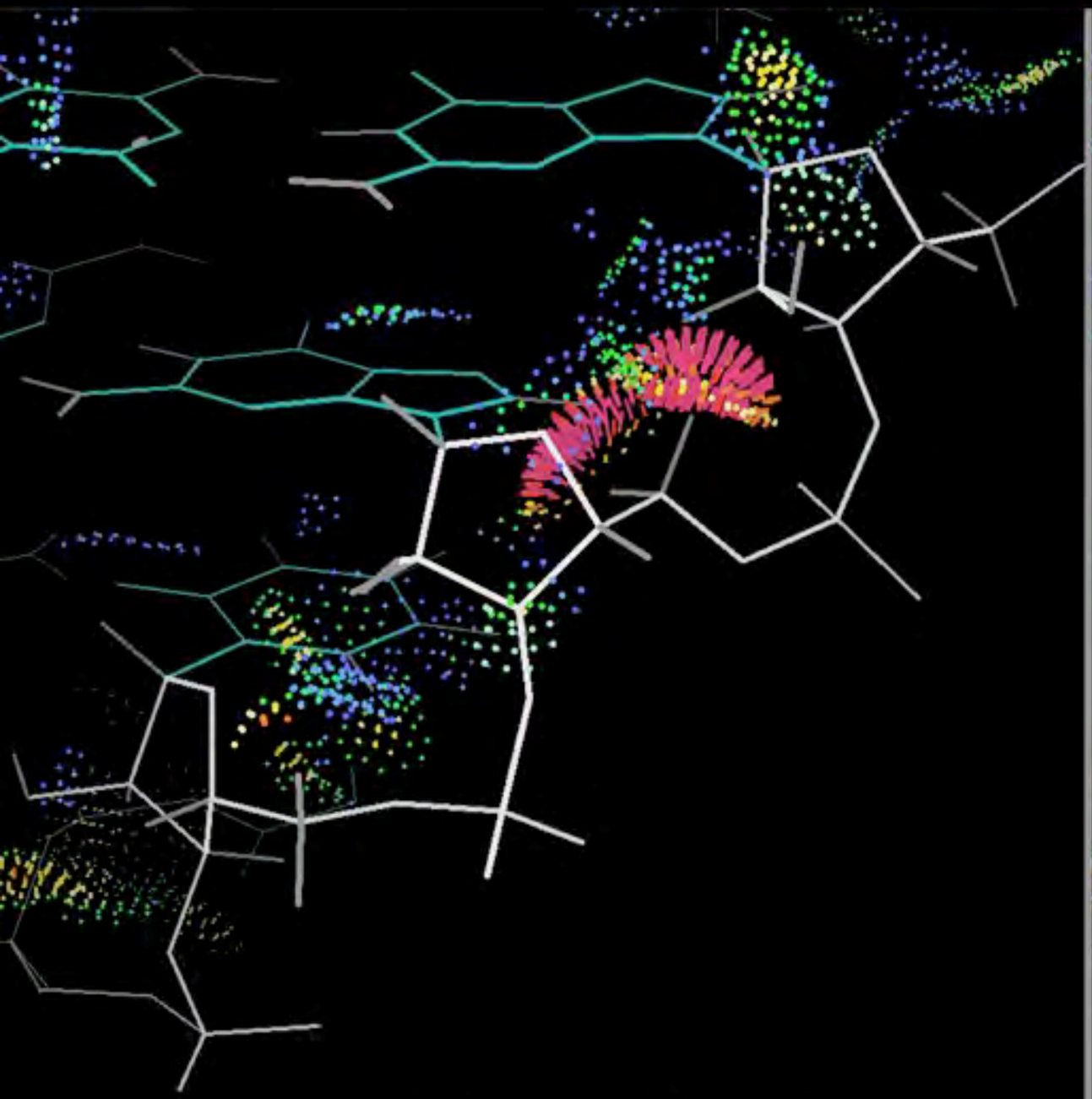

1JJ2
base stack


backbone
contacts

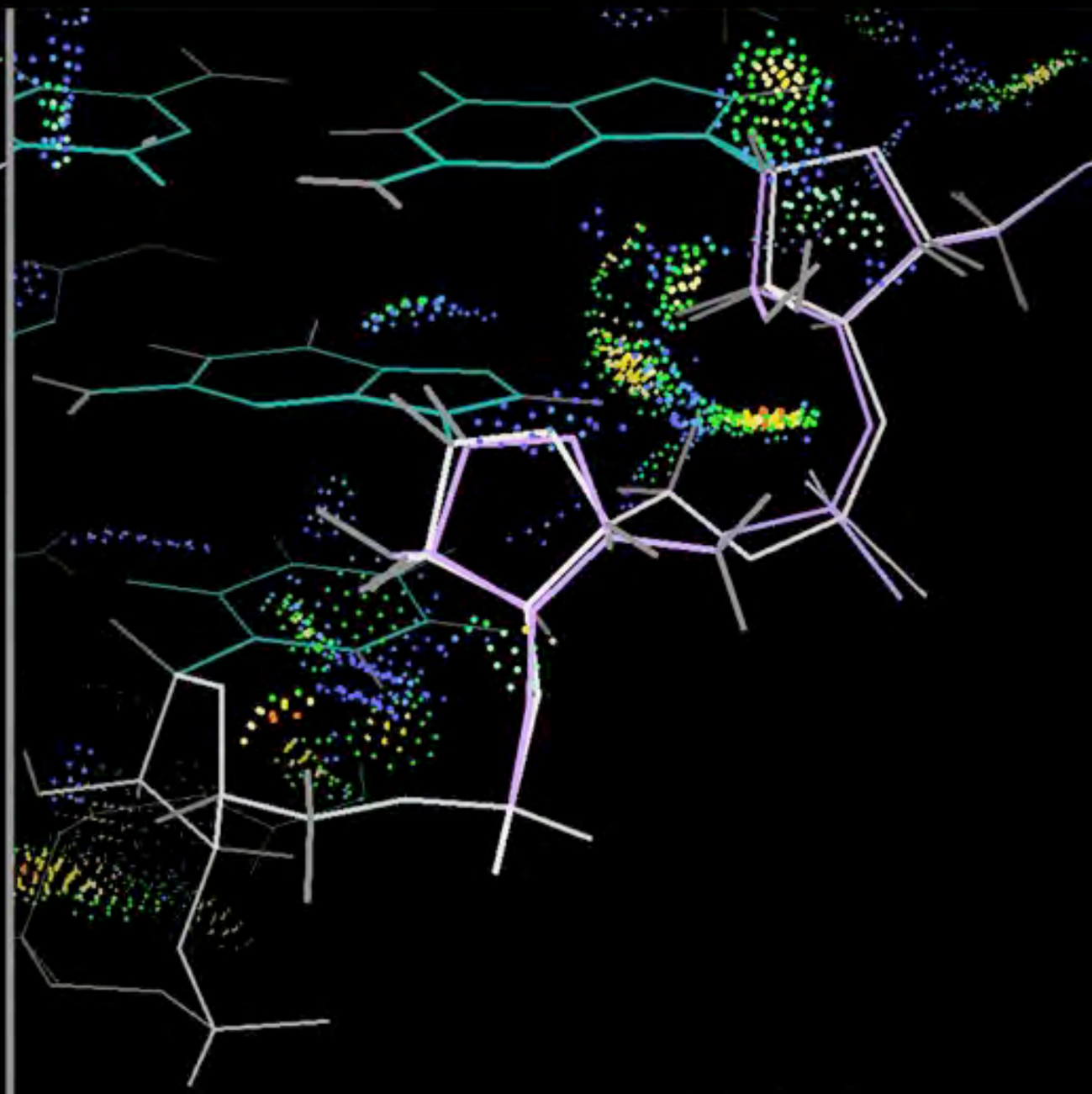




1aG1gN1aR1aA1c

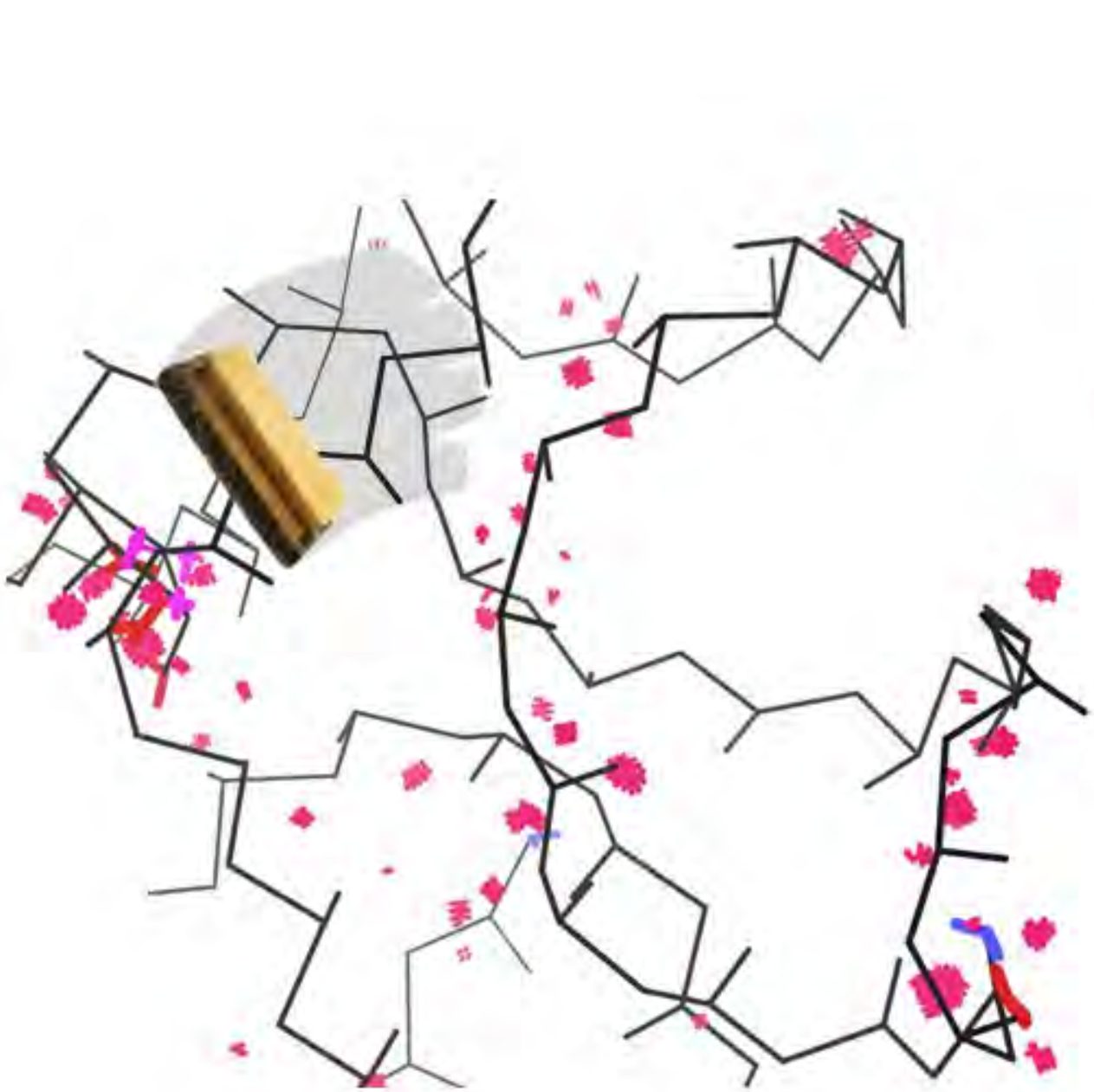


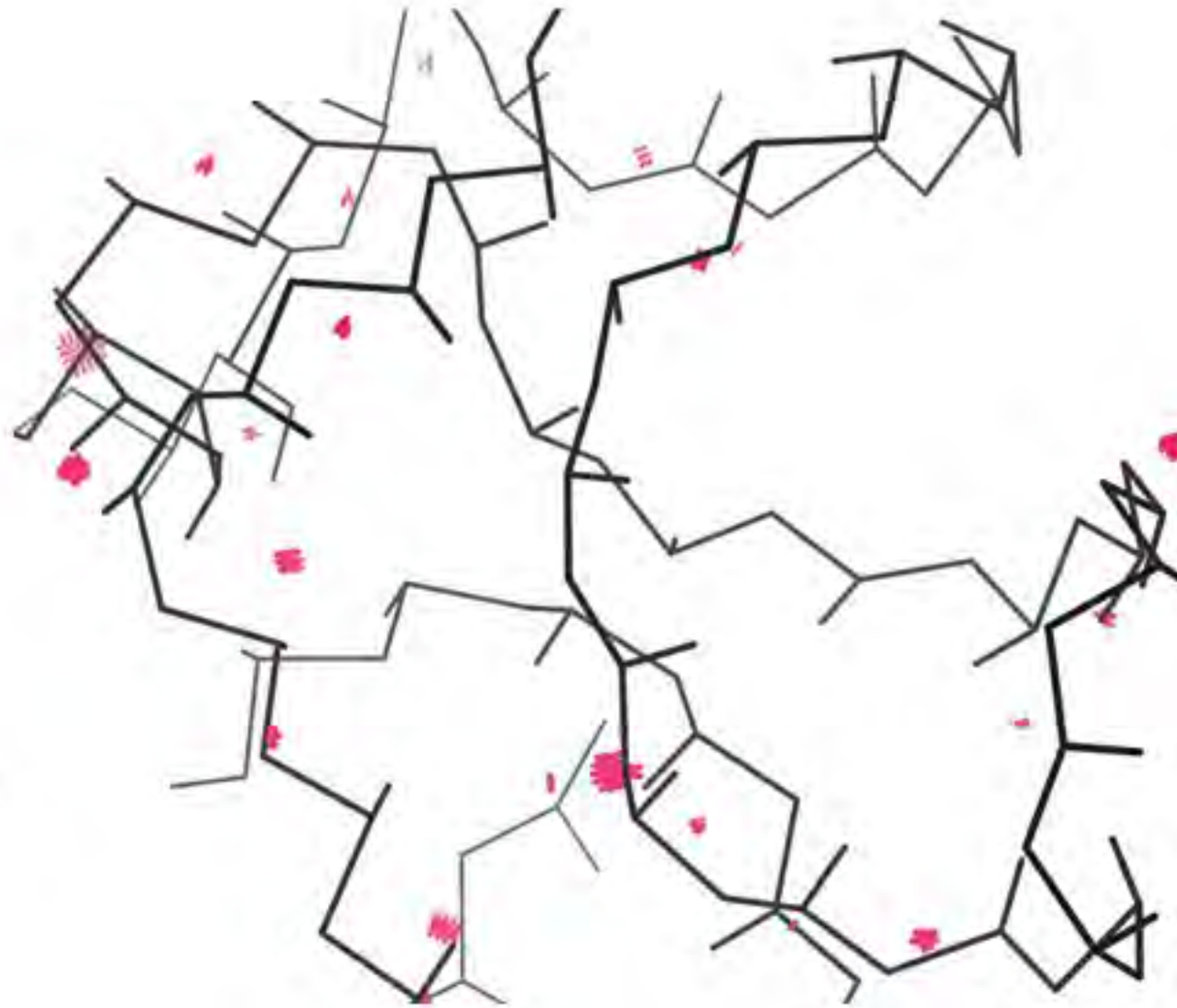
1JJ2, 23S 870-1 original



— repaired

A revolution in ERRASing errors from RNA backbone





Chou *et al* 2012, Nature Methods 10:74

ERRASER =

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

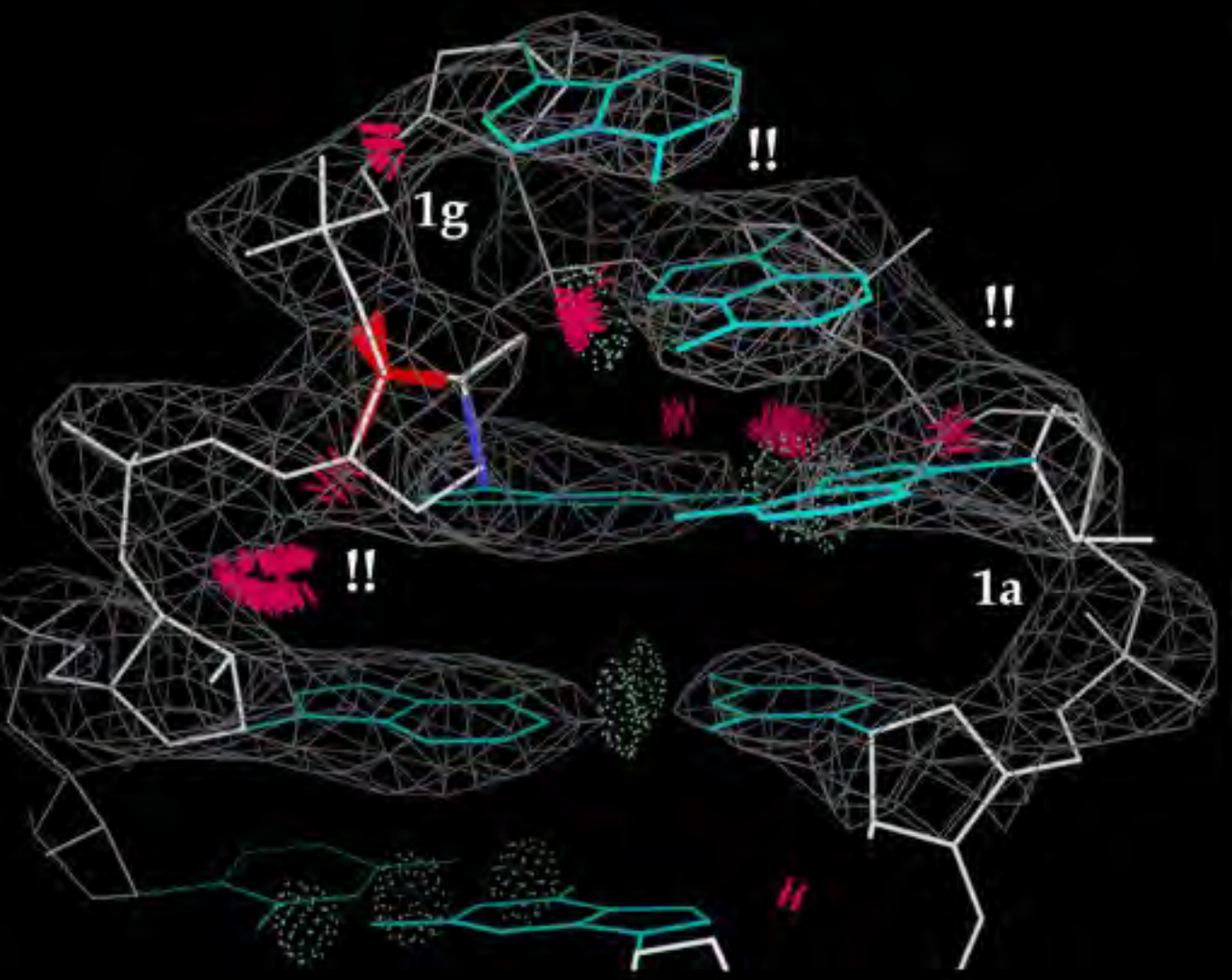
Rhiju Das, Fang Chou
Stanford

MOLPROBITY

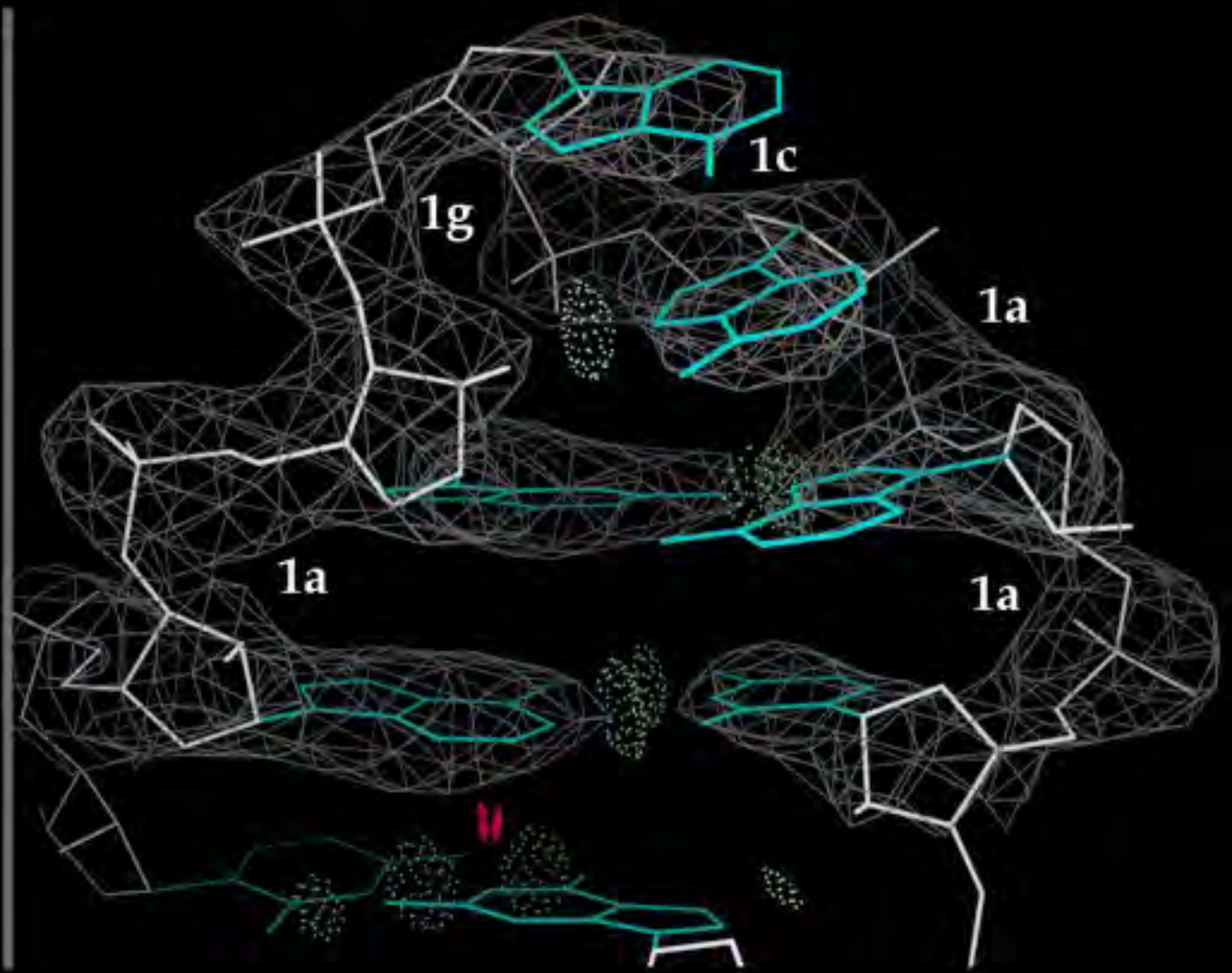
		2GIS: original pdb		after ERRASER run 3	
All-Atom Contacts	Clashscore, all atoms:	43.14	43rd percentile	9.59	98th percentile
		2.90Å		2.90Å	
Nucleic Acid Geometry	Probably wrong sugar puckers:	8	Goal: 0	0	Goal: 0
	Bad backbone conformations#:	21	Goal: 0	4	Goal: 0
	Residues with bad bonds:	0.00%	Goal: 0%	0.00%	Goal: 0%
	Residues with bad angles:	9.57%	Goal: <0.1%	0.00%	Goal: <0.1%

Rfree: 0.269 **Rfree: 0.250**

2gis SAM-I riboswitch, 2.9Å, GNRA 50-54



as deposited



after ERRASER

!!

A9

!!

!!

!!

U63

5d

original PDB

7p

A9

2a

2z

8d

U63

5p

ERRASER run 2

2GIS: A9-G63 contact with Mg++

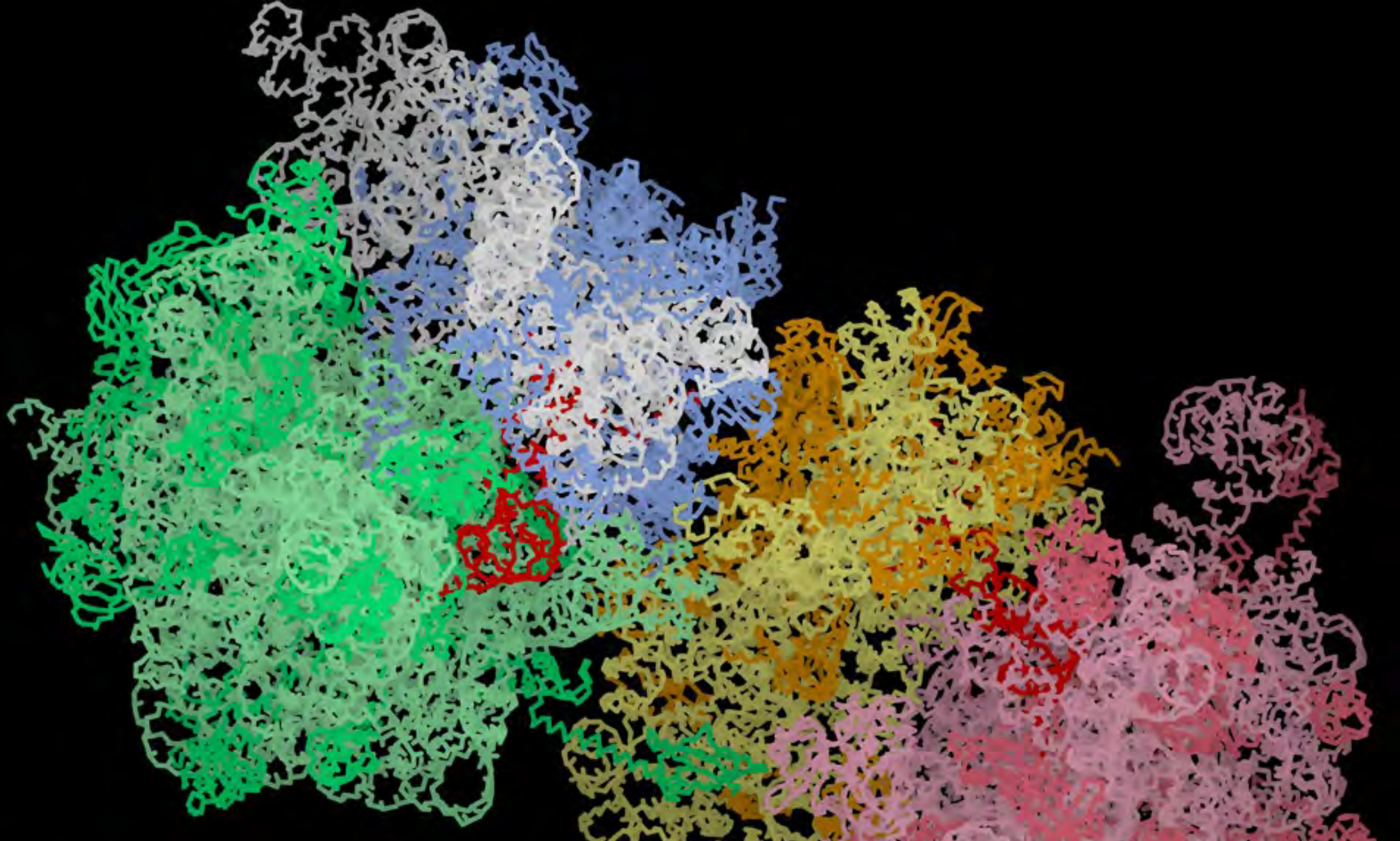
Mg

U 123

1CX0, 2.3Å - original

U 123

1CX0, 2.3Å - rebuilt



We might actually
build detailed-accuracy
ribosome models in our lifetime!

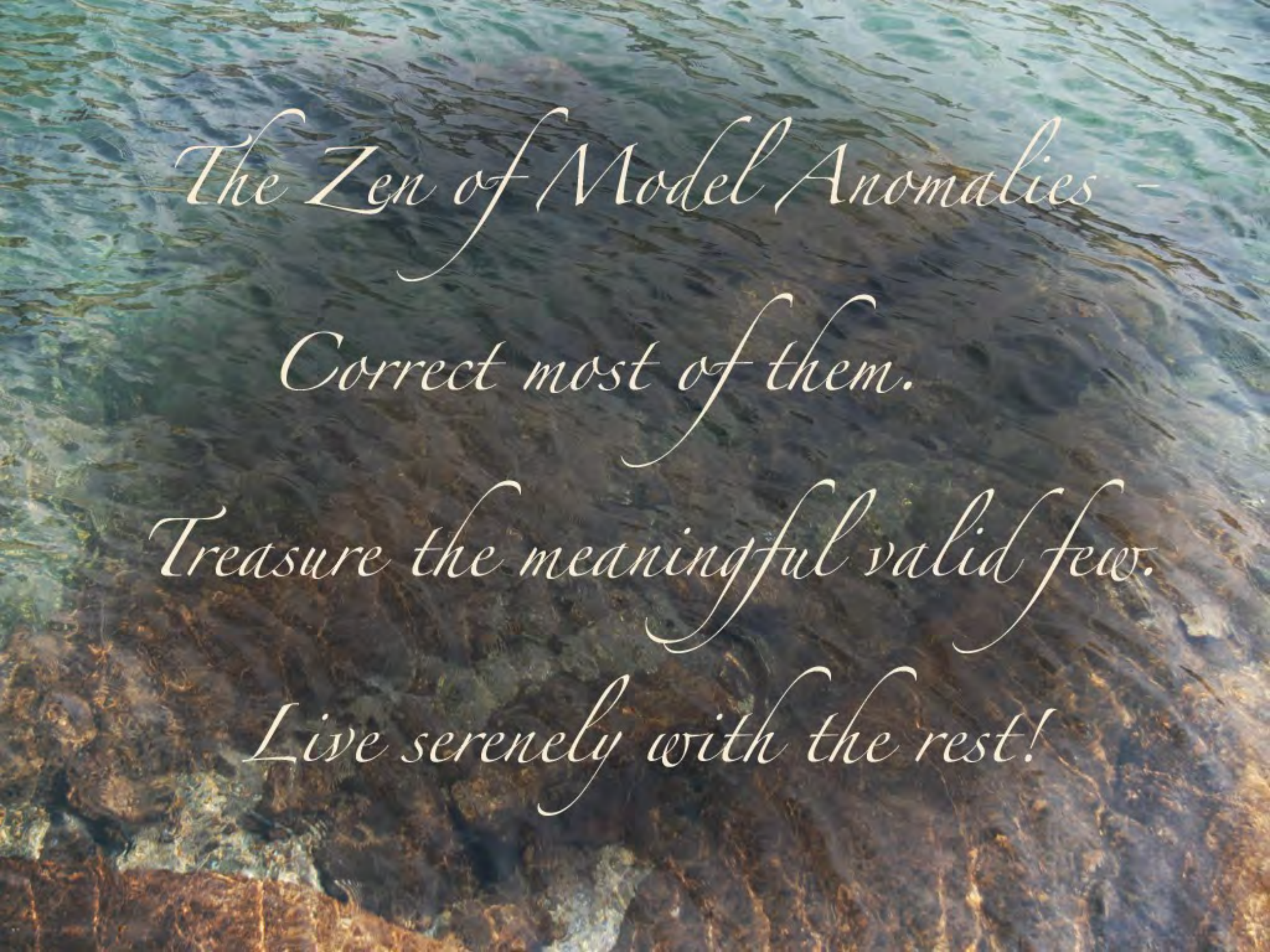
All-Atom Contact Analysis

RLab:

Bryan Arendall
Jeff Headd
Steven Lewis
Gary Kapral
Christopher Williams
Swati Jain
Bradley Hintze
Lindsay Deis
Lizbeth Videau
Michael Prisant
Jane & Dave Richardson

Mike Word
Simon Lovell
Laura Murray
Ian Davis
Vincent Chen
Jeremy Block
Dan Keedy
Jack Snoeyink, UNC
Rhiju Das, Stanford
RNA Ontology Consortium
PHENIX developers!

NIGMS


The Zen of Model Anomalies -
Correct most of them.
Treasure the meaningful valid few.
Live serenely with the rest!