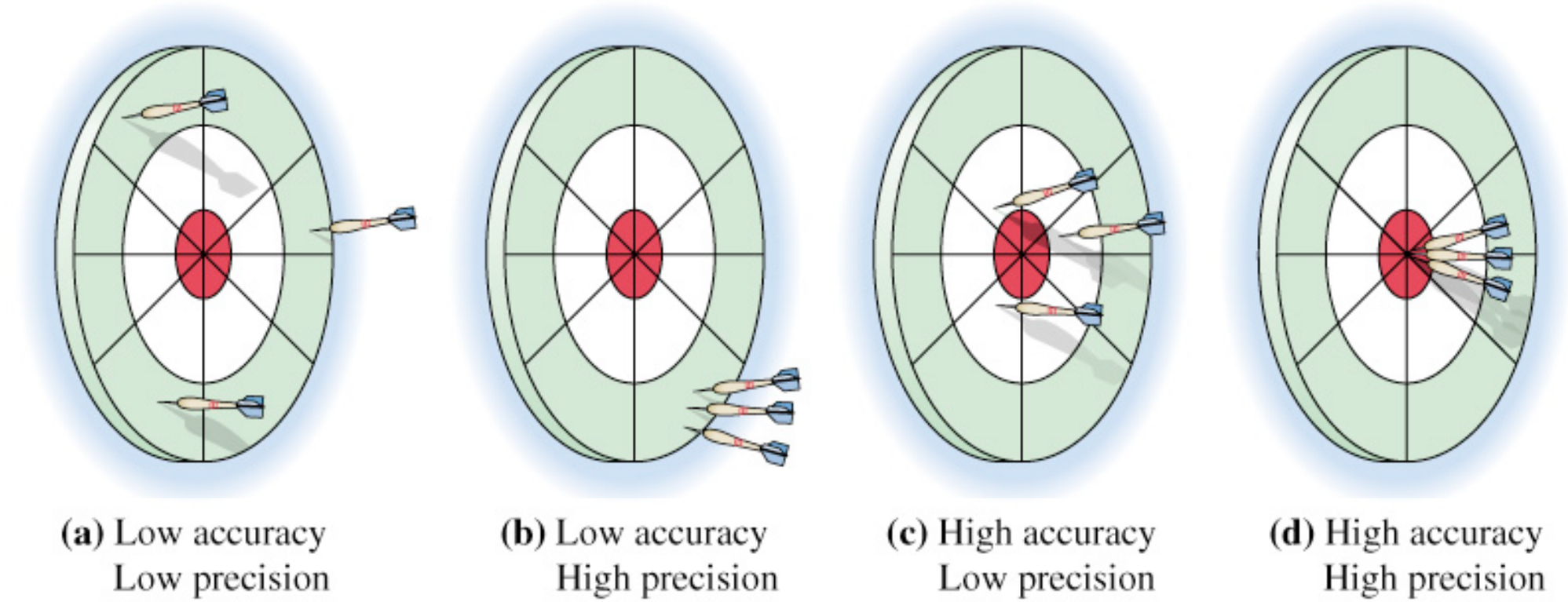
Model Validation at Low Resolution

ECM32 Vienna
August 2019

Paul Adams
Lawrence Berkeley Laboratory and
Department of Bioengineering UC Berkeley

Accuracy & Precision



(a) Low accuracy
Low precision

(b) Low accuracy
High precision

(c) High accuracy
Low precision

(d) High accuracy
High precision

From http://extensionengine.com, by Furqan Nazeeri

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Errors

- Random errors (noise)
  - Typically normally distributed
  - Can be reduced by increasing the number of observations
  - Affect the precision
- Systematic errors (bias)
  - Could arise from a poor experimental design or lack of understanding of the system being studied
  - Are reproducibly biased
  - Affect the accuracy
- Gross errors
  - Incorrect assumptions have been made or serious mistakes undetected
  - May be detectable as outliers compared to prior knowledge

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Mistakes Still Happen

Retraction

WE WISH TO RETRACT OUR RESEARCH ARTICLE "STRUCTURE OF MsbA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters" and both of our Reports "Structure of the ABC transporter MsbA in complex with ADP•vanadate and lipopolysaccharide" and "X-ray structure of the EmrE multidrug transporter in complex with a substrate" (*1*–*3*).

The recently reported structure of Sav1866 (*4*) indicated that our MsbA structures (*1*, *2*, *5*) were incorrect in both the hand of the structure and the topology. Thus, our biological interpretations based on these inverted models for MsbA are invalid.

An in-house data reduction program introduced a change in sign for anomalous differences. This program, which was not part of a conventional data processing package, converted the anomalous pairs (I+ and I–) to (F– and F+), thereby introducing a sign change. As the diffraction data collected for each set of MsbA crystals and for the EmrE crystals were processed with the same program, the structures reported in (*1*–*3*, *5*, *6*) had the wrong hand.

The error in the topology of the original MsbA structure was a consequence of the low resolution of the data as well as breaks in the electron density for the connecting loop regions. Unfortunately, the use of the multicopy refinement procedure still allowed us to obtain reasonable refinement values for the wrong structures.

The Protein Data Bank (PDB) files 1JSQ, 1PF4, and 1Z2R for MsbA and 1S7B and 2F2M for EmrE have been moved to the archive of obsolete PDB entries. The MsbA and EmrE structures will be recalculated from the original data using the proper sign for the anomalous differences, and the new Cα coordinates and structure factors will be deposited.

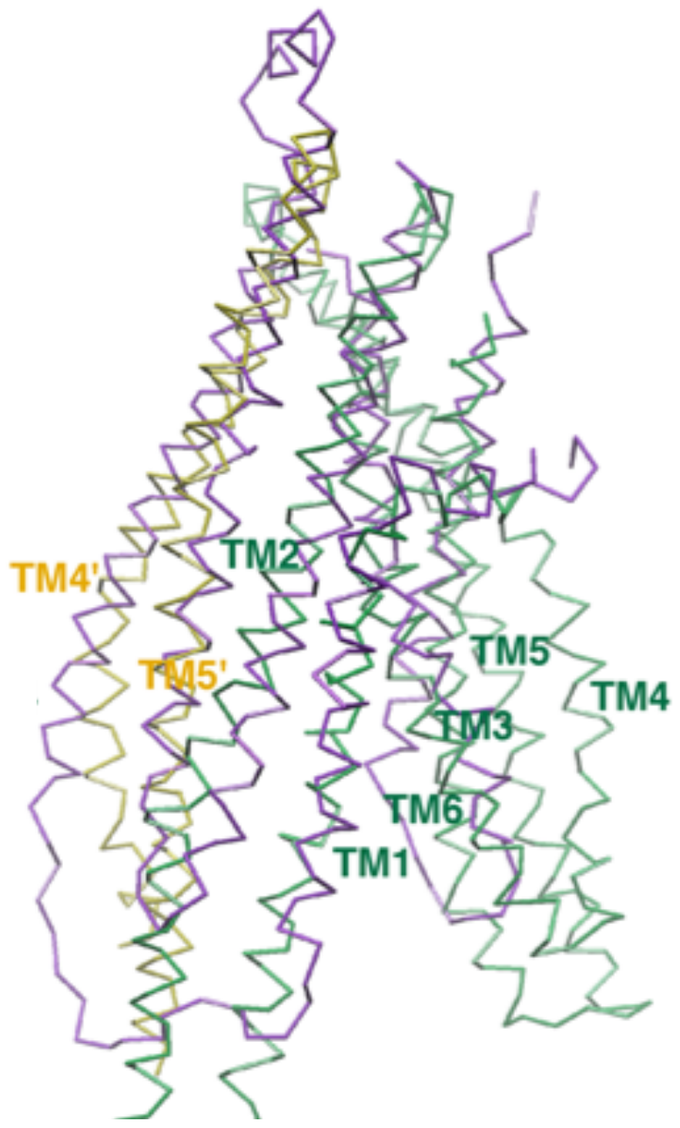
We very sincerely regret the confusion that these papers have caused and, in particular, subsequent research efforts that were unproductive as a result of our original findings.

GEOFFREY CHANG, CHRISTOPHER B. ROTH, CHRISTOPHER L. REYES, OWEN PORNILLOS, YEN-JU CHEN, ANDY P. CHEN

Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.

References

1. G. Chang, C. B. Roth, *Science* **293**, 1793 (2001).
2. C. L. Reyes, G. Chang, *Science* **308**, 1028 (2005).
3. O. Pornillos, Y.-J. Chen, A. P. Chen, G. Chang, *Science* **310**, 1950 (2005).
4. R. J. Dawson, K. P. Locher, *Nature* **443**, 180 (2006).
5. G. Chang, *J. Mol. Biol.* **330**, 419 (2003).
6. C. Ma, G. Chang, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2852 (2004).



Dawson & Locher, Nature
443, 180-185, 2006

Retraction: Cocrystal structure of synaptobrevin-II bound to botulinum neurotoxin type B at 2.0 Å resolution

Michael A Hanson & Raymond C Stevens

Nat. Struct. Biol. 7, 687–692 (2000); retracted 6 July 2009

In this paper, we described both the three-dimensional crystal structure of a botulinum toxin catalytic domain separated from the holotoxin (BoNT/B-LC, PDB 1F82) and a structure of the toxin catalytic domain in complex with a peptide (Sb2–BoNT/B-LC, PDB 1F83). The complex was later refined and deposited in the Protein Data Bank (PDB 3G94). The apo structure (PDB 1F82) remains valid. However, because of the lack of clear and continuous electron density for the peptide in the complex structure, the paper is being retracted. We apologize for any confusion this may have caused.

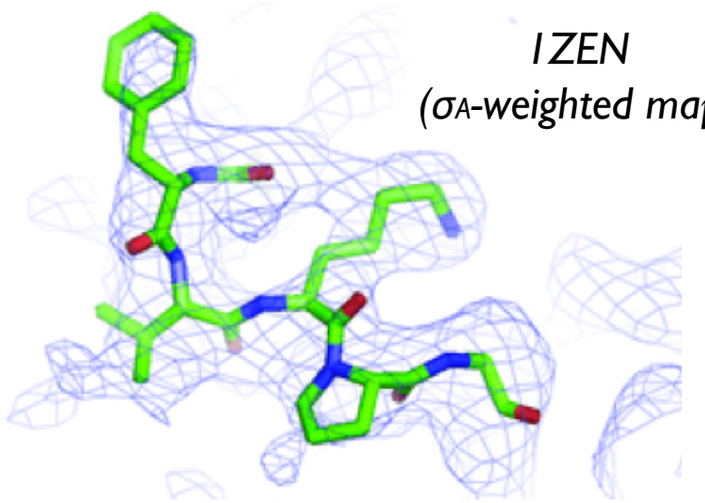
BERKELEY LAB
Lawrence Berkeley National Laboratory

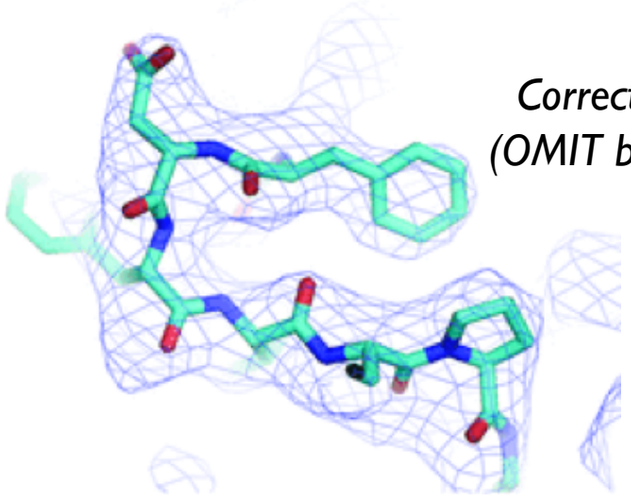
Phenix

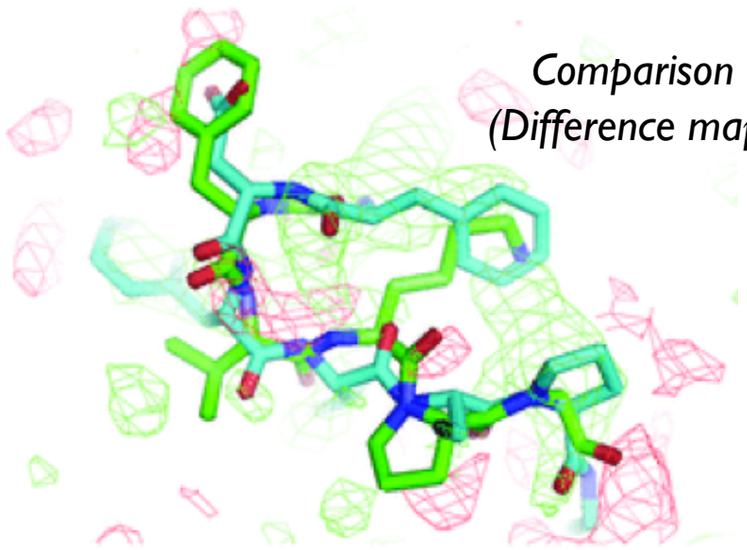
THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

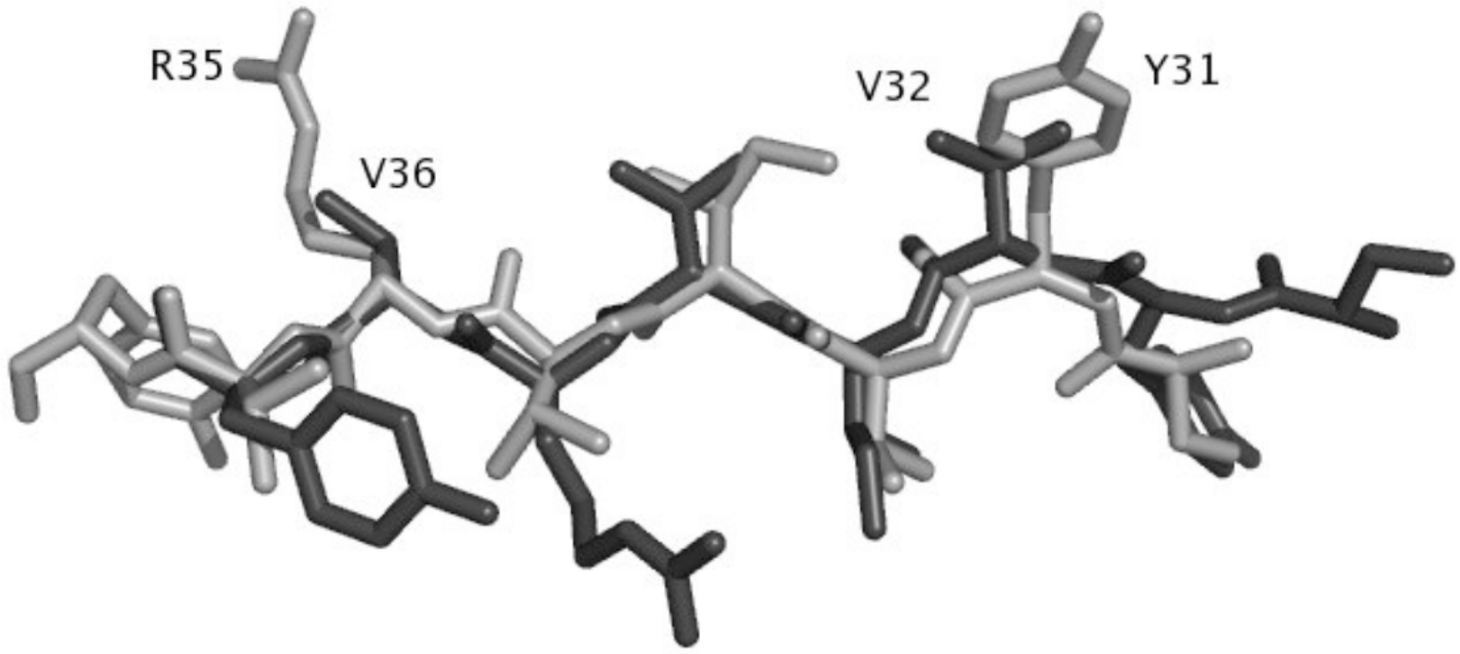
Register Errors

- Register errors typically start in loop regions (over or under building)
- Fixing these errors can be challenging (loop regions often have poor density) - estimated that 1% of structures in PDB have register errors
- Real space analysis can help
- Packing analysis (WHATCHECK, MolProbity)









1CHR, 3.0 (light) versus 2CHR (dark)

Image from Gerard Kleywegt, European Bioinformatics Institute

Terwilliger et al., Acta Cryst D64, 515-524, 2008

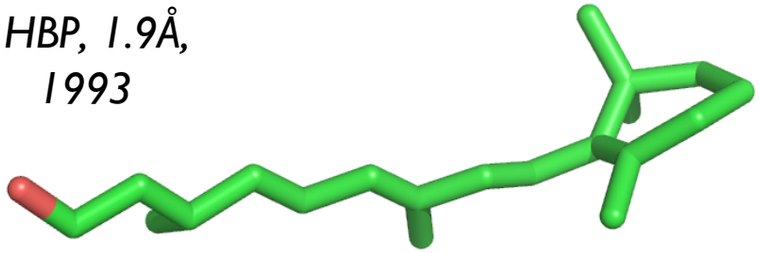
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

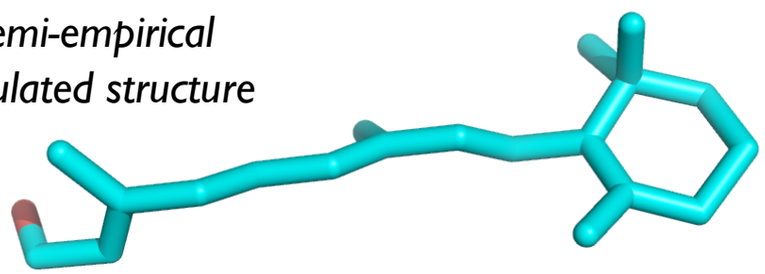
Other Kinds of Errors

- Systematic error in magnification
- Incorrect sequence (less common these days)
- Incorrectly placed waters or too many waters
- Waters fit instead of ions and side chains
- Small molecule geometry (where did you get the restraints from?)

1HBP, 1.9Å, 1993



Semi-empirical calculated structure



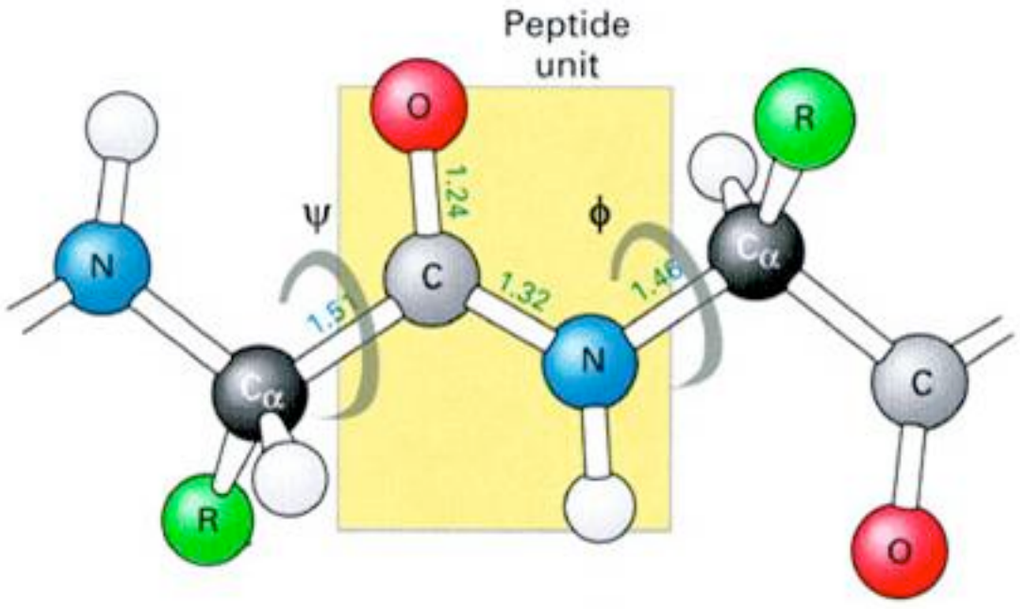
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Geometric Measures

- Some of the best measures for validation are from information not used in the model optimization (e.g. Free R-value)
- For geometry (of proteins) one of the best measures is the Ramachandran distribution - the main chain torsion angles
- The handedness of amino acids, and the steric clashes that occur, given the side chain attachment to the mainchain, results in limits on the distribution of mainchain torsion angles



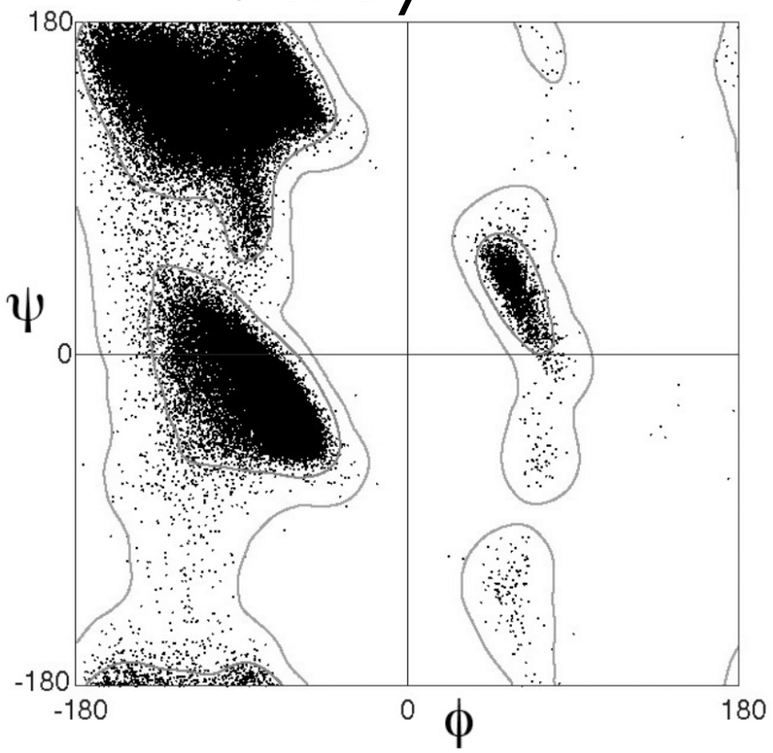
G. N. Ramachandran

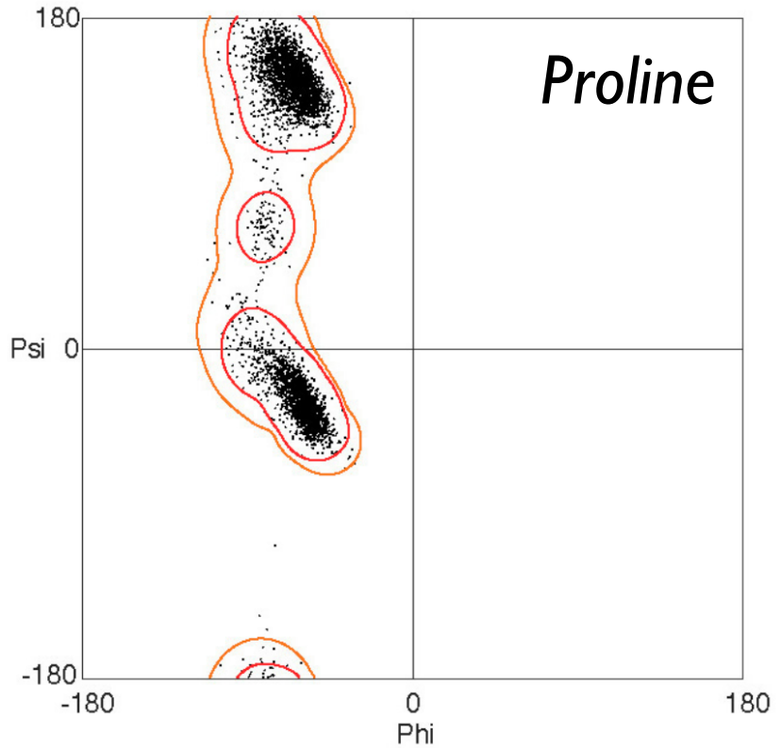
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

The Ramachandran Plot

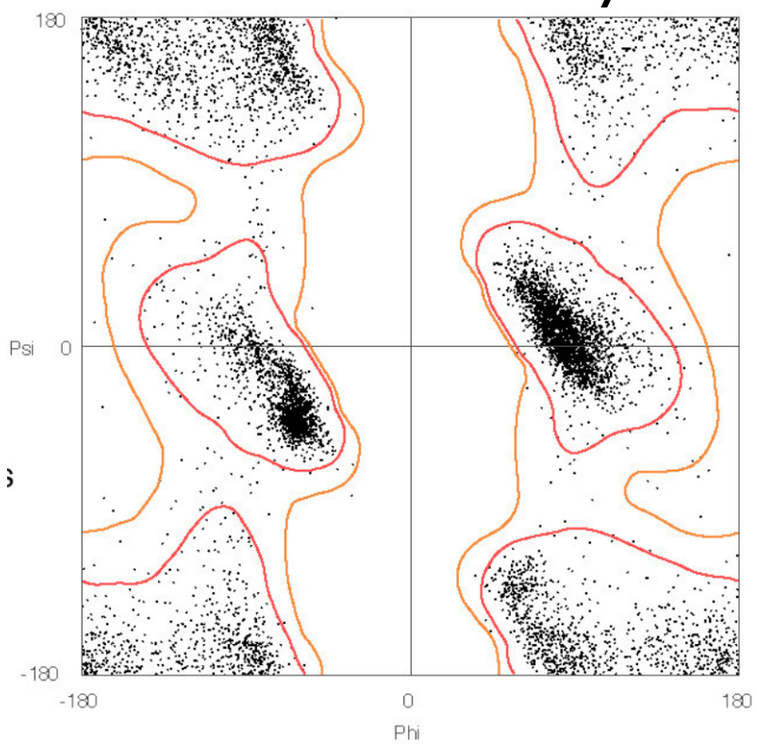
All minus Pro & Gly

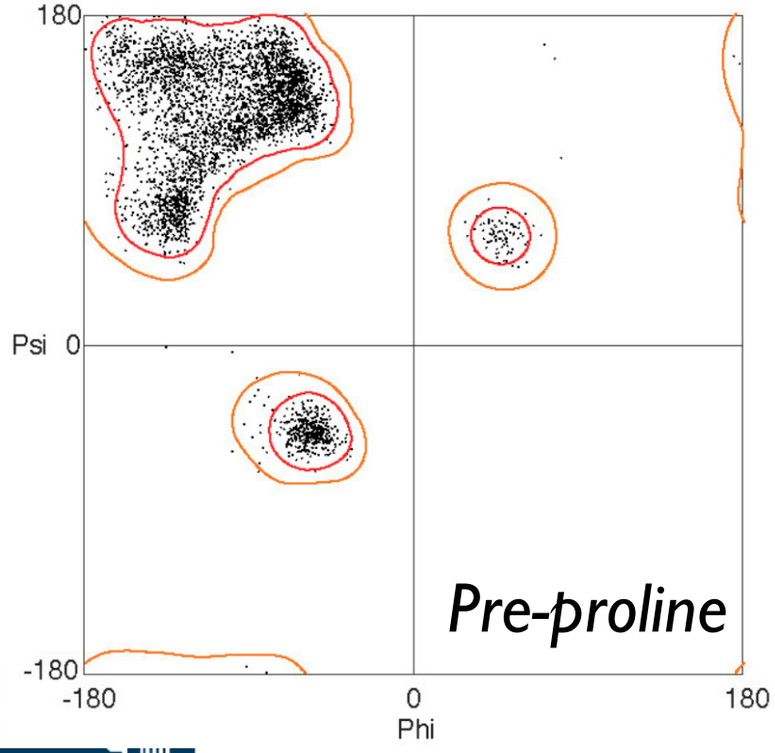




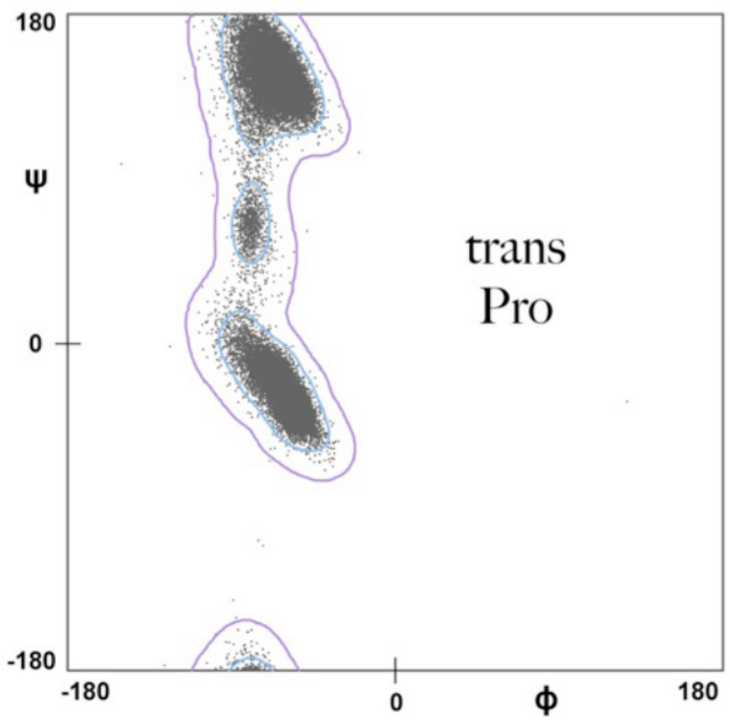
Proline

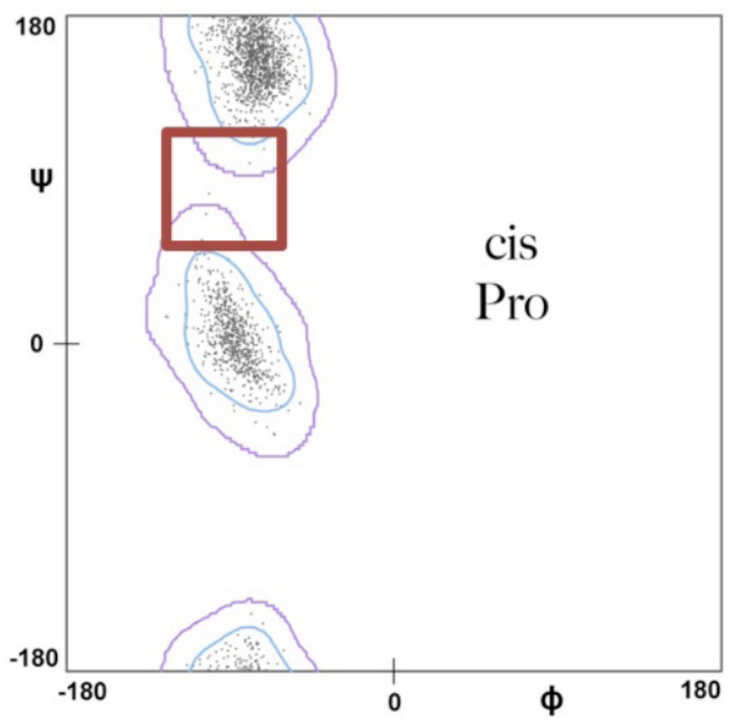
Glycine





Pre-proline





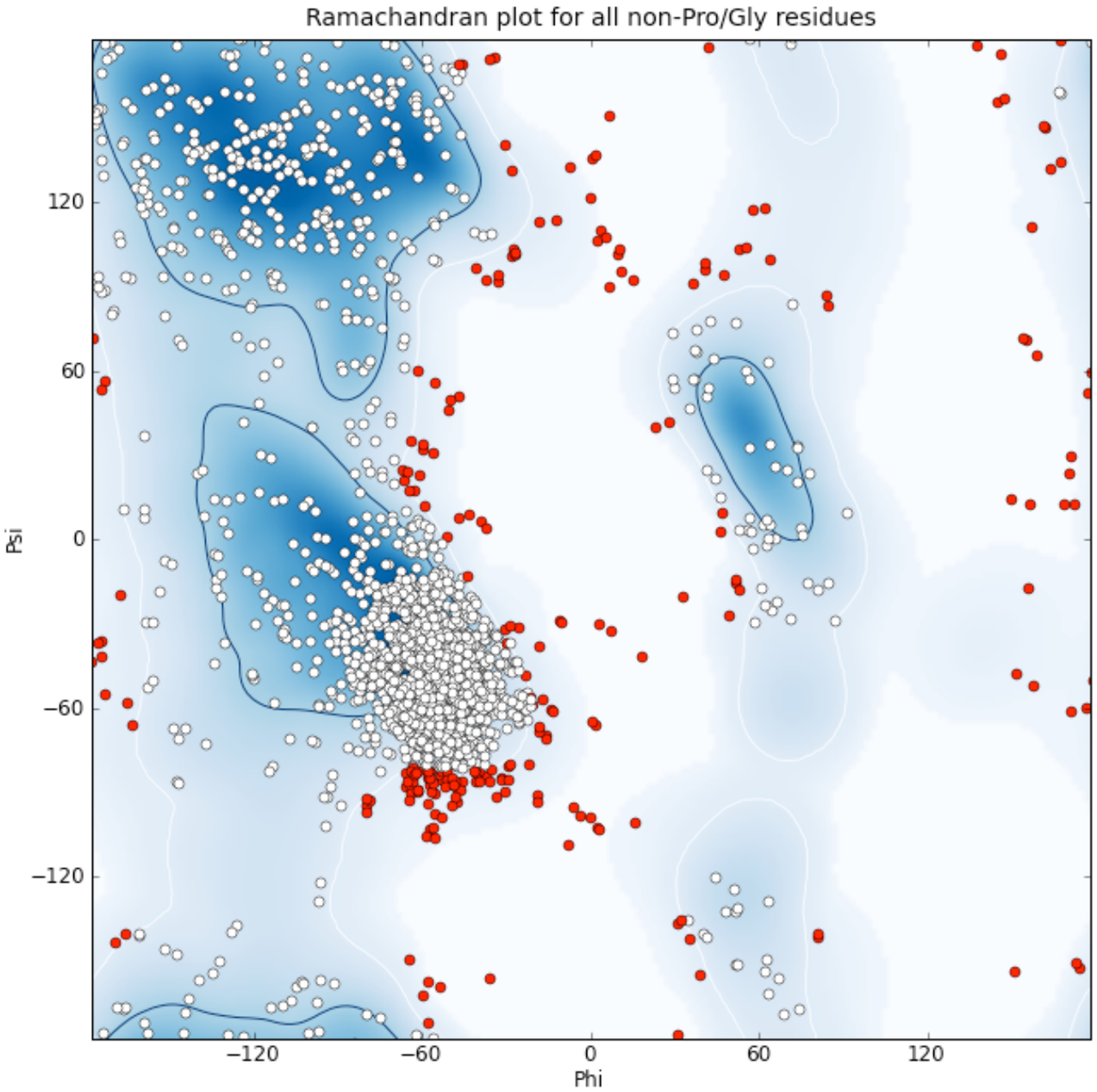
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Images from Jane and David Richardson, Duke University

The Ramachandran Plot

- A protein structure should in general conform to prior expectations (based on theory and prior observation)
- Most (98%+) residues should have a mainchain conformation consistent with the Ramachandran distribution
- A small percentage (0.2%) of residue may show Ramachandran outliers (note they are not necessarily errors)
  - Outliers can be seen in strained regions of the structure (e.g. in the active site)
- Any outliers need to be confirmed by detailed analysis



BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Rotamers

- There are steric clashes between atoms within amino acid side chains
- These clashes lead to preferred conformations, called rotamers
- Different rotamers are generated by rotation of side chain torsion angles (χ_1, χ_2 etc)

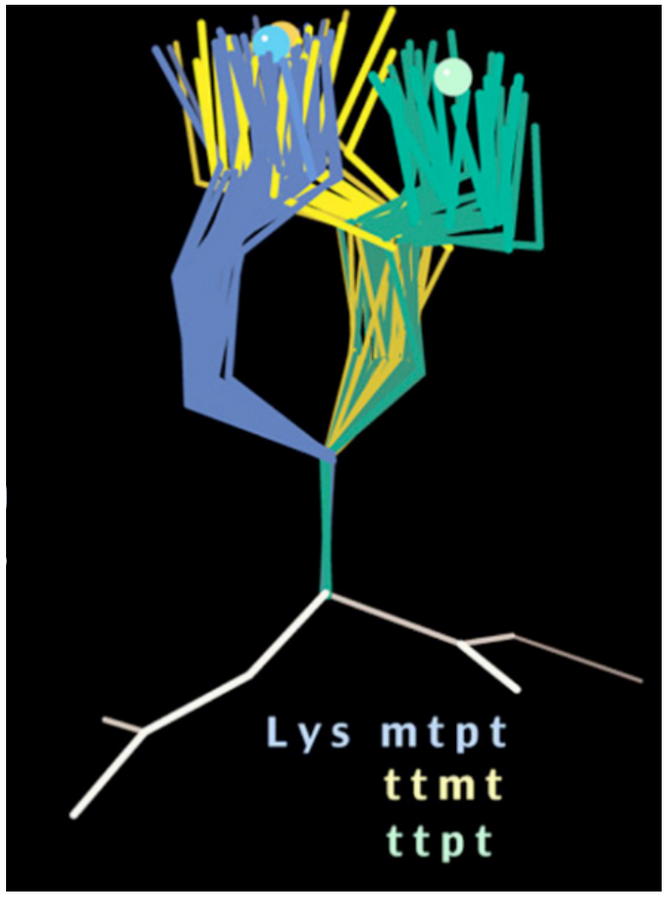
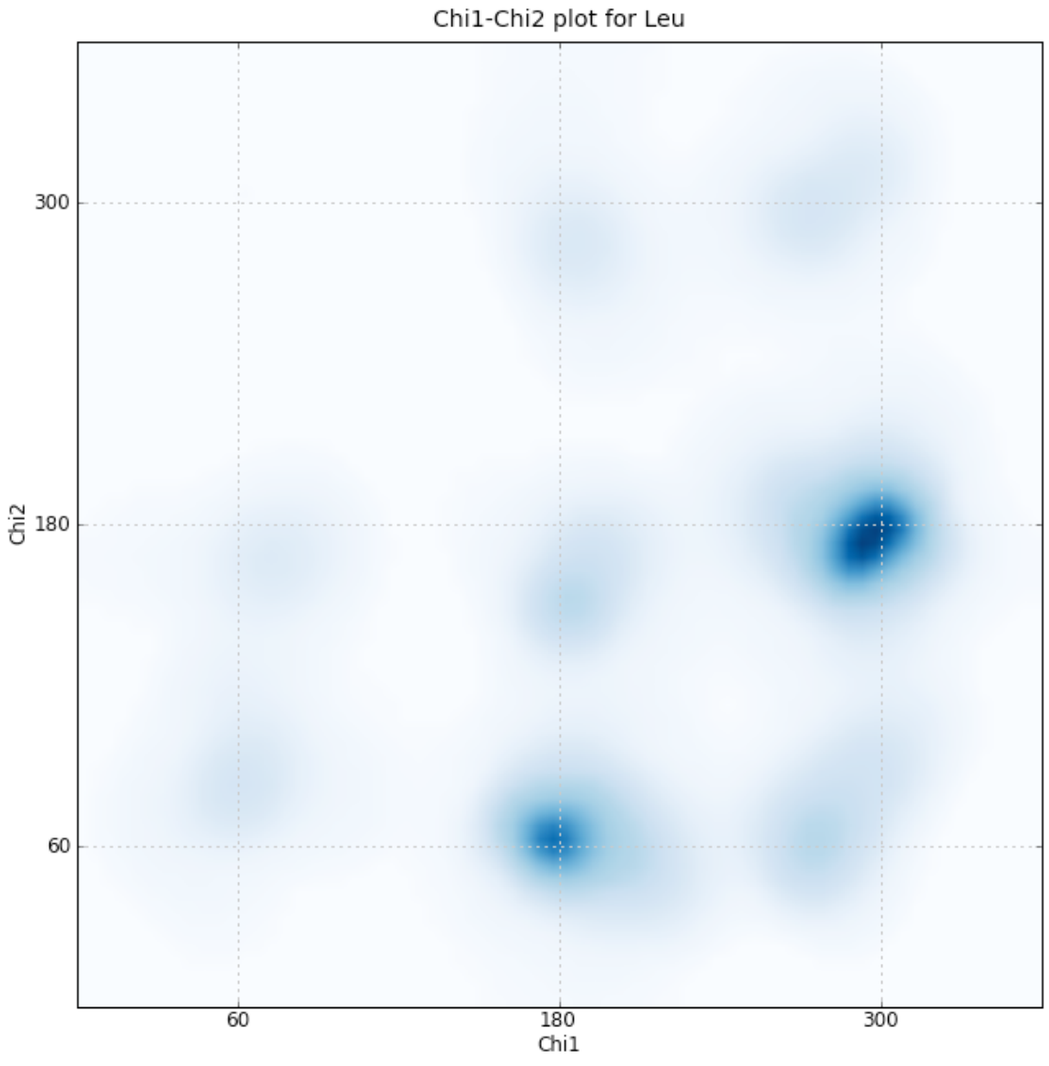


Image from Jane and David Richardson, Duke University



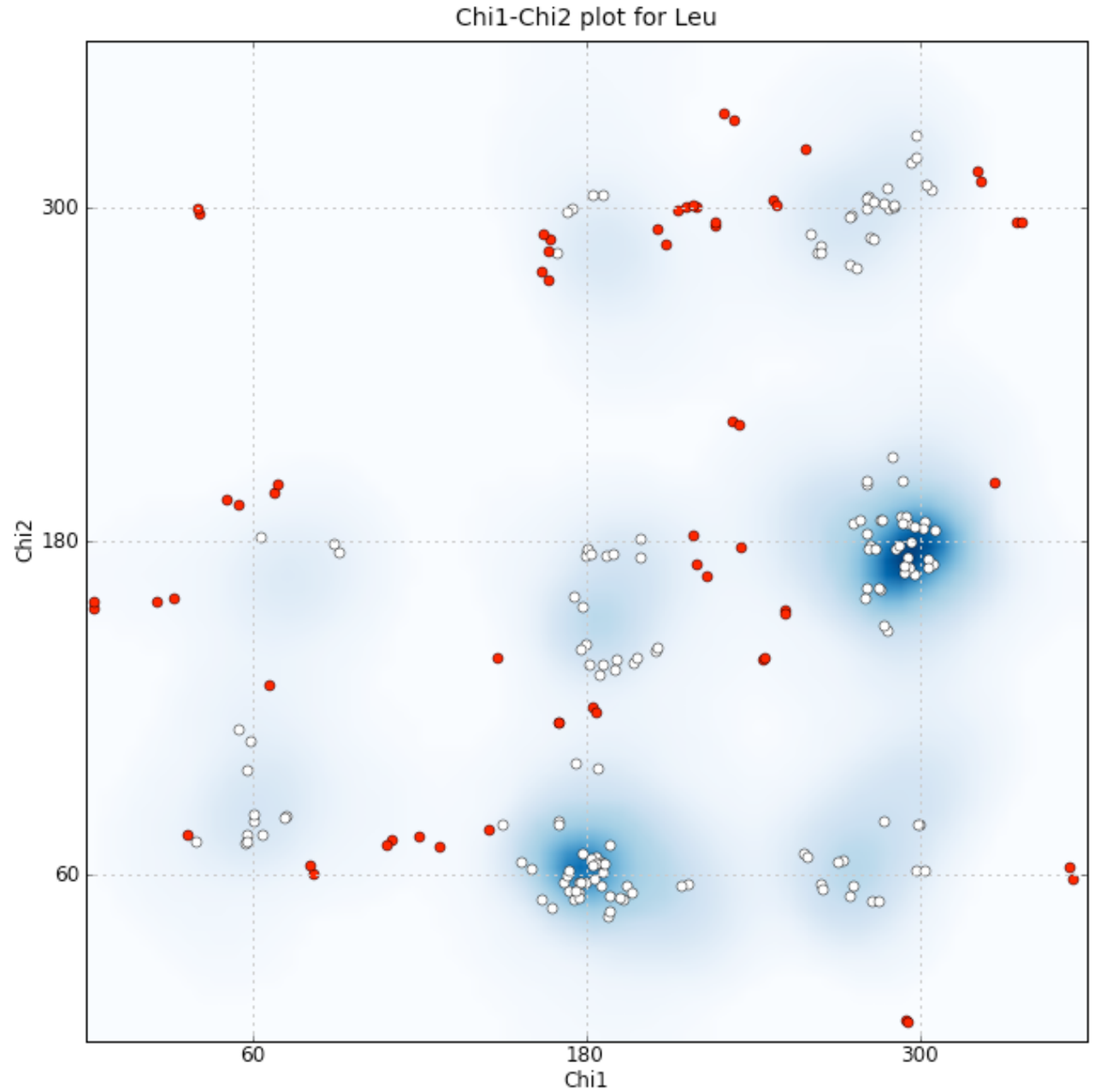
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Rotamers

- As with the Ramachandran distribution, protein side chains are expected to conform to known rotamer distributions
- More variability because of interactions with other sidechains, mainchain or ligands
- Outliers may be meaningful, but need to be verified
- Sidechains on the protein surface will often have little density (disorder)

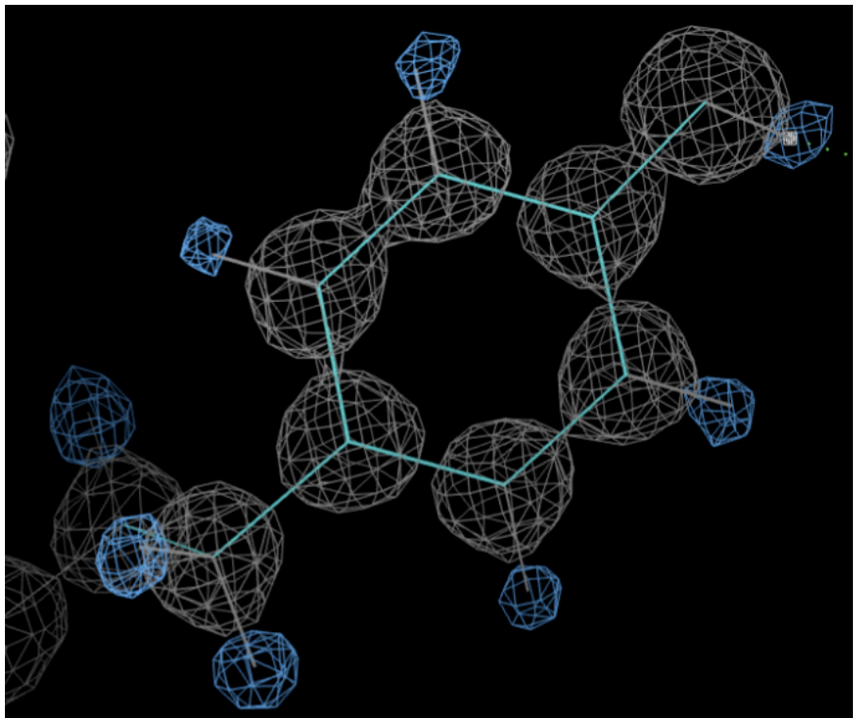


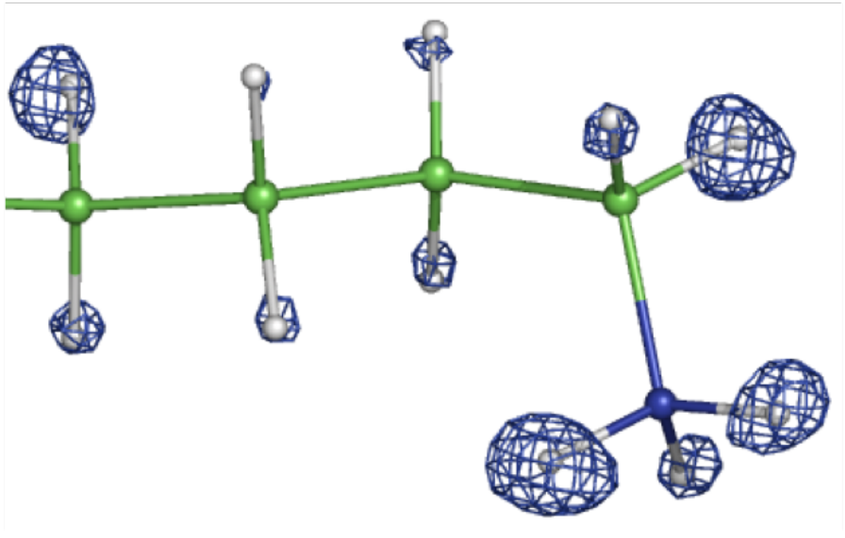
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Hydrogens

- Macromolecules contain hydrogens
  - Approximately half of the atoms in a structure
- Hydrogens make the majority of contacts in a structure
- Typically ignored because they aren't typically seen experimentally
- But, the hydrogens are there!
- The Richardson group (Duke University) have pioneered the use of hydrogens in calculating packing (and clashes) inside macromolecules
- The quality of packing and the nature of clashes can be used to validate and correct structures





Images from Jane and David Richardson, Duke University

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

All Atom Contacts

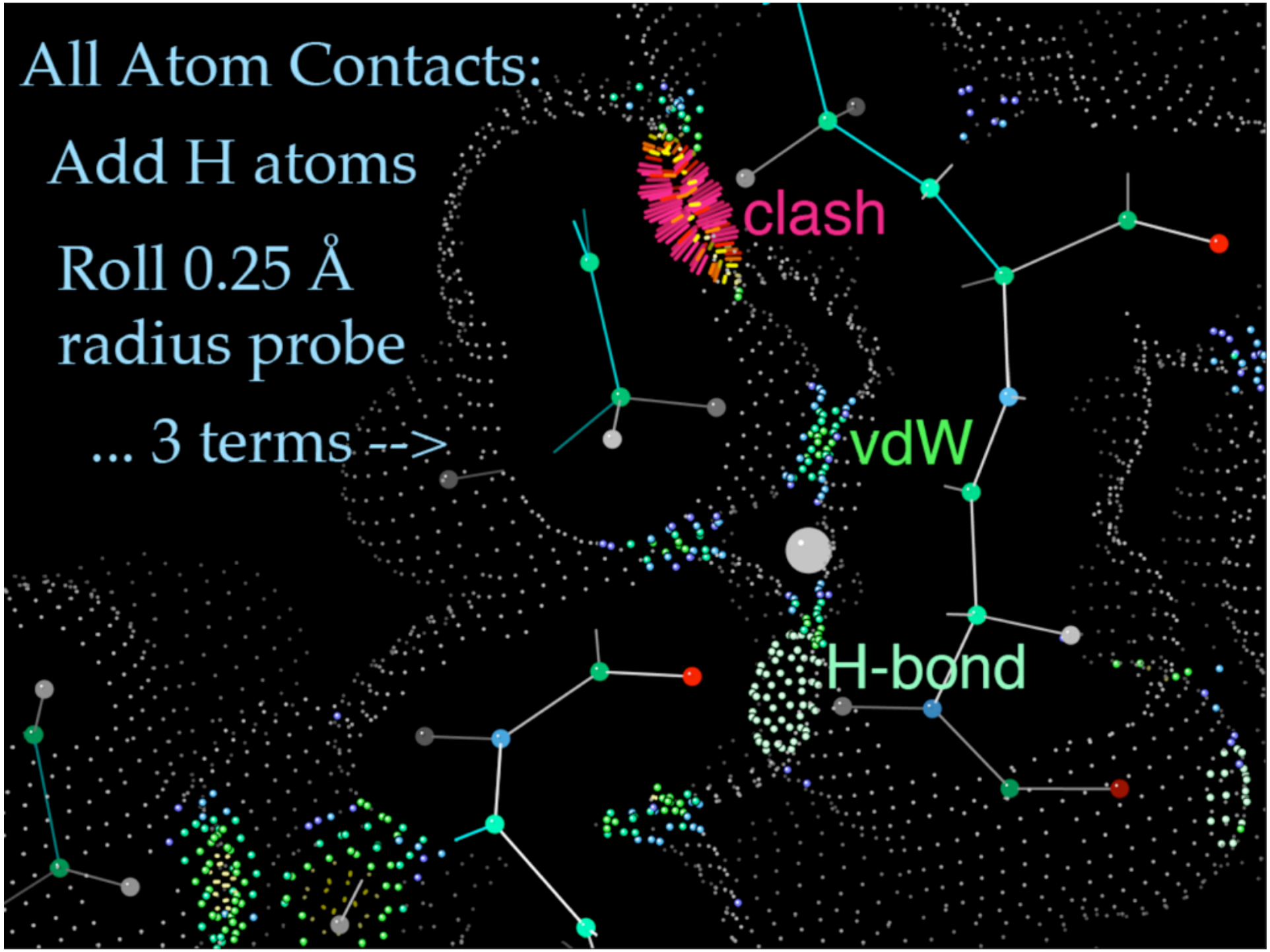


Image from Jane and David Richardson, Duke University

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

MolProbity

- MolProbity has been developed to validate structures (purely on coordinates)
- Performs all atom contacts, Ramachandran, rotamer and other geometry analyses

Analysis output: all-atom contacts and geometry for 3g5uH.pdb

Summary statistics

All-Atom Contacts	Clashscore, all atoms:	159.56	2nd percentile* (N=37, 3Å - 9999Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	20.10%	Goal: <1%
	Ramachandran outliers	11.33%	Goal: <0.2%
	Ramachandran favored	62.35%	Goal: >98%
	Cβ deviations >0.25Å	6	Goal: 0
	MolProbity score^	4.55	4th percentile* (N=342, 3.25Å - 4.05Å)
	Residues with bad bonds:	0.04%	Goal: 0%
	Residues with bad angles:	3.85%	Goal: <0.1%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

^ MolProbity score is defined as the following: 0.42574*log(1+clashscore) + 0.32996*log(1+max(0,pctRotOut-1)) + 0.24979*log(1+max(0,100-pctRamaFavored-2)) + 0.5

By adding H to this model and allowing Asn/Gln/His flips, we could *automatically* improve your clashscore by 2.05 points.

http://molprobity.biochem.duke.edu/

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA · BERKELEY · 1868

MolProbity

- Generates detailed problem list
- Problems can be fixed more easily by using validation lists viewed visually (e.g. Coot from Phenix)

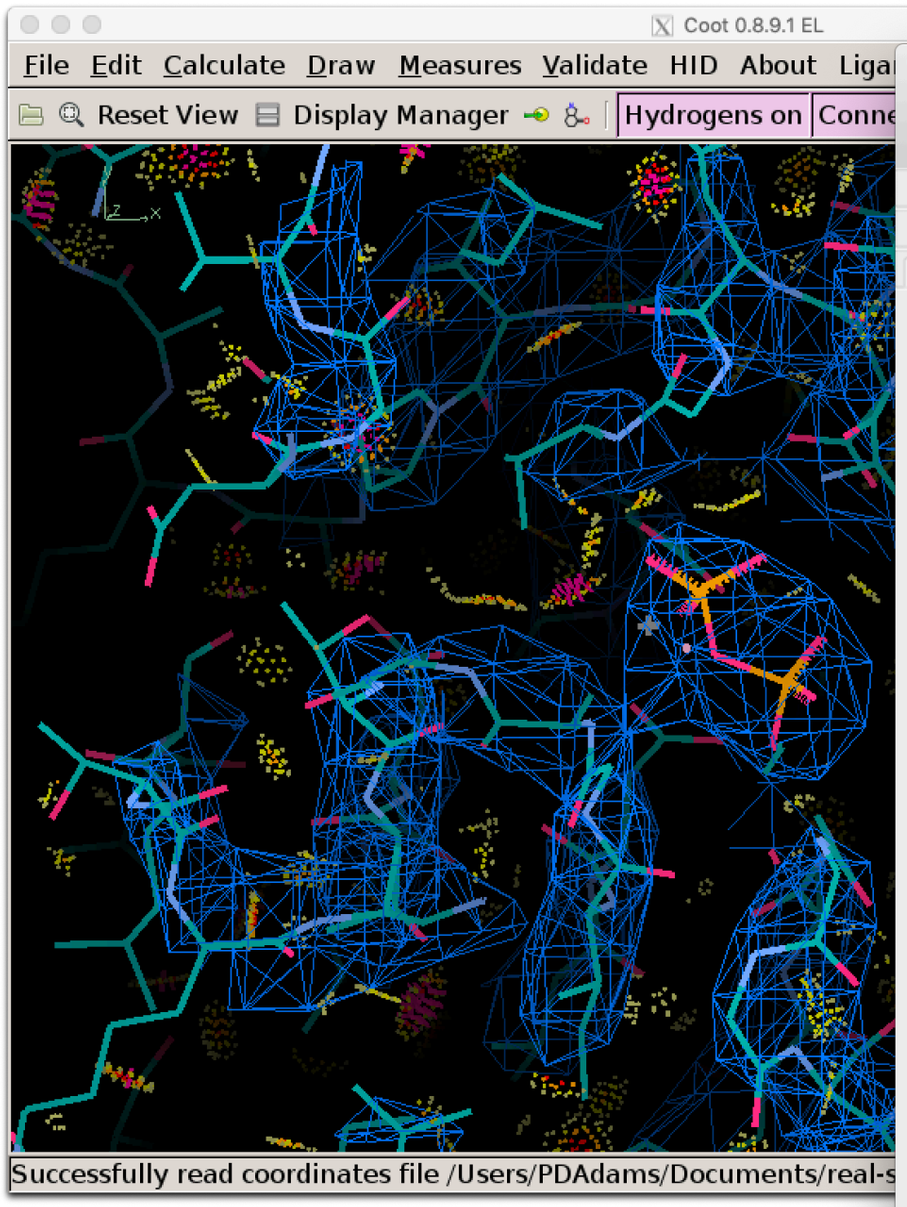
#	Res	High B	Clash > 0.4Å	Ramachandran	Rotamer	Cβ deviation	Bond lengths.	Bond angles.
		Avg: 154.80	Clashscore: 159.56	Outliers: 267 of 2356	Poor rotamers: 392 of 1950	Outliers: 6 of 2174	Outliers: 1 of 2364	Outliers: 91 of 2364
A 33	VAL	207.38	0.761Å N with A 36 LEU HD11	-	5.4% (*m*) chi angles: 285.2	0.04Å	-	-
A 34	SER	186.49	1.084Å HA with A 38 MET HB2	OUTLIER (0%) General case / 1.9,-66.3	82.4% (*p*) chi angles: 69.5	0.015Å	-	-
A 35	VAL	204.94	1.221Å HG12 with A 359 TYR CE2	Allowed (0.14%) General case / -50.4,-77.0	19.1% (*m*) chi angles: 304.1	0.049Å	-	-
A 36	LEU	142.23	1.095Å H with A 35 VAL HG23	Favored (42.64%) General case / -73.2,-49.2	1% chi angles: 55.4,111.8	0.07Å	-	-
A 37	THR	170.59	0.723Å H with A 36 LEU HG	Favored (3.45%) General case / -69.8,-60.5	64.9% (*m*) chi angles: 296.4	0.052Å	-	-
A 38	MET	155.79	1.084Å HB2 with A 34 SER HA	Favored (36.13%) General case / -49.2,-42.4	0% chi angles: 229.3,294.4,131.7	0.054Å	-	-
A 39	PHE	122.57	1.127Å HB2 with A 35 VAL O	Favored (49.78%) General case / -51.0,-39.2	31.2% (*t80*) chi angles: 196.6,84	0.027Å	-	-
A 40	ARG	102.85	0.683Å N with A 37 THR O	Allowed (0.94%) General case / -116.3,69.9	9.1% (*mtp180*) chi angles: 248.9,176.6,56.5,191.6	0.078Å	-	-
			0.484Å	Allowed (0.6%)	0.5%	°		

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Results - Rebuilding and Validation



Real-space refinement (Project: real-space-refine-5ljv_0)

Preferences Help Run Abort Save Ask for help

Input/Output | Refinement Settings | RealSpaceRefine_1

Results | Validation

Summary | MolProbity | Model vs. Data | Data

Open in Coot

Export Table 1

Model

Composition (#)	
Chains	2
Atoms	2500 (Hydrogens: 0)
Residues	Protein: 325 Nucleotide: 0
Water	0
Ligands	MG: 1 ADP: 1
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.005 (0)
Angles (°) (# > 4σ)	0.754 (0)
MolProbity score	1.81
Clash score	10.33
Ramachandran plot (%)	
Outliers	0.00
Allowed	4.02
Favored	95.98
Rotamer outliers (%)	0.00
Cβ outliers (%)	0.00
Peptide plane (%)	
Cis proline/general	5.6/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	2.18
ADP (B-factors)	
Iso/Aniso (#)	2500/0
min/max/mean	
Protein	38.70/81.46/52.85
Nucleotide	---
Ligand	48.04/49.31/48.09
Water	---
Occupancy	
Mean	1.00

Data

Box		
Lengths (Å)	49.58, 68.34, 80.40	
Angles (°)	90.00, 90.00, 90.00	
Supplied Resolution (Å)	3.7	
Resolution Estimates (Å)	Masked	Unmasked
d FSC (half maps; 0.143)	---	---
d 99 (full/half1/half2)	3.7/---/---	3.2/---/---
d model	3.7	3.7
d FSC model (0/0.143/0.5)	3.2/3.4/3.6	3.4/3.6/3.9
Map min/max/mean	-0.42/0.80/0.03	

Model vs. Data

CC (mask)	0.81
CC (box)	0.57
CC (peaks)	0.40
CC (volume)	0.78
Mean CC for ligands	0.73

Idle

Project: real-space-refine-5ljv_0

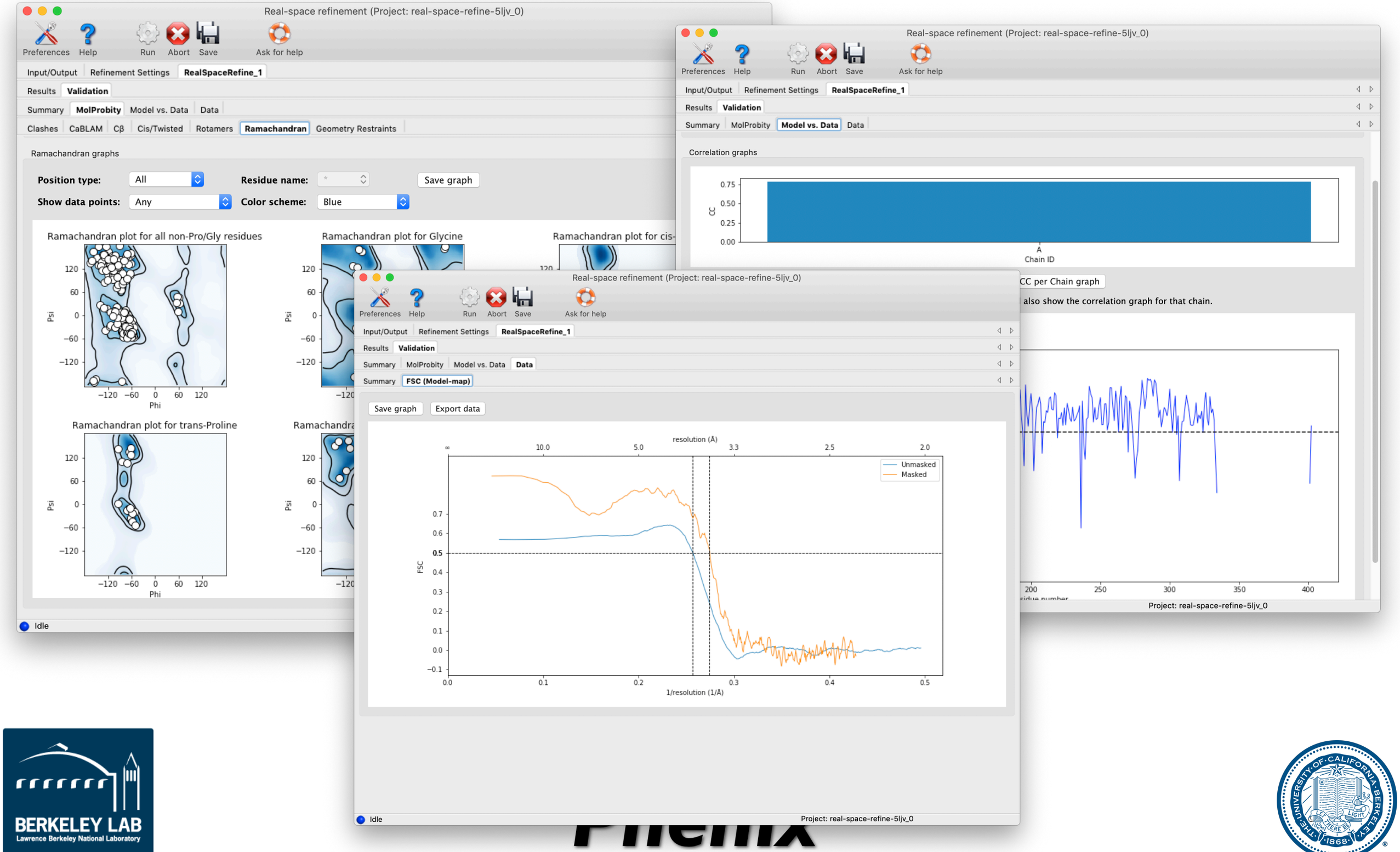
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

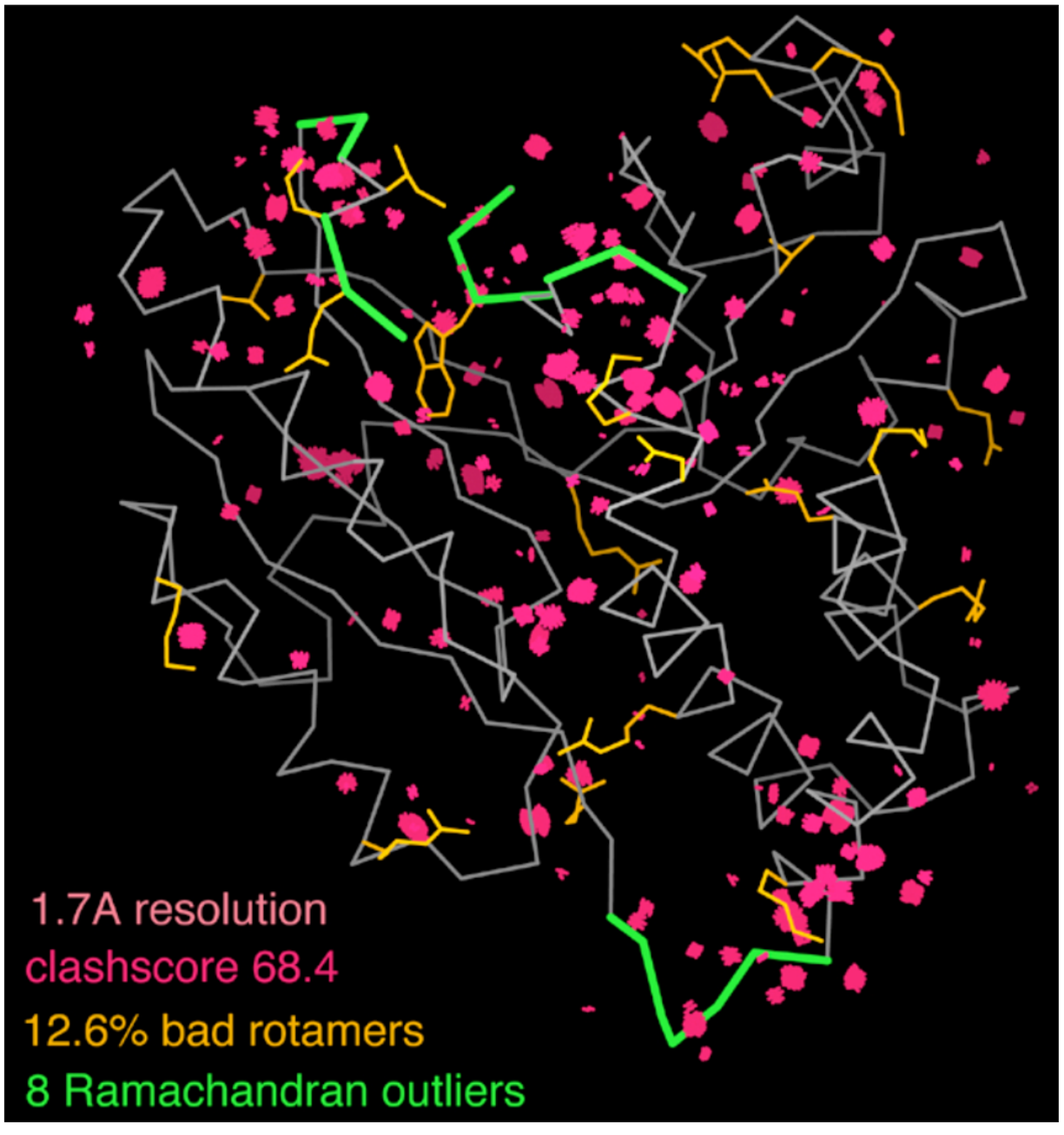
THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

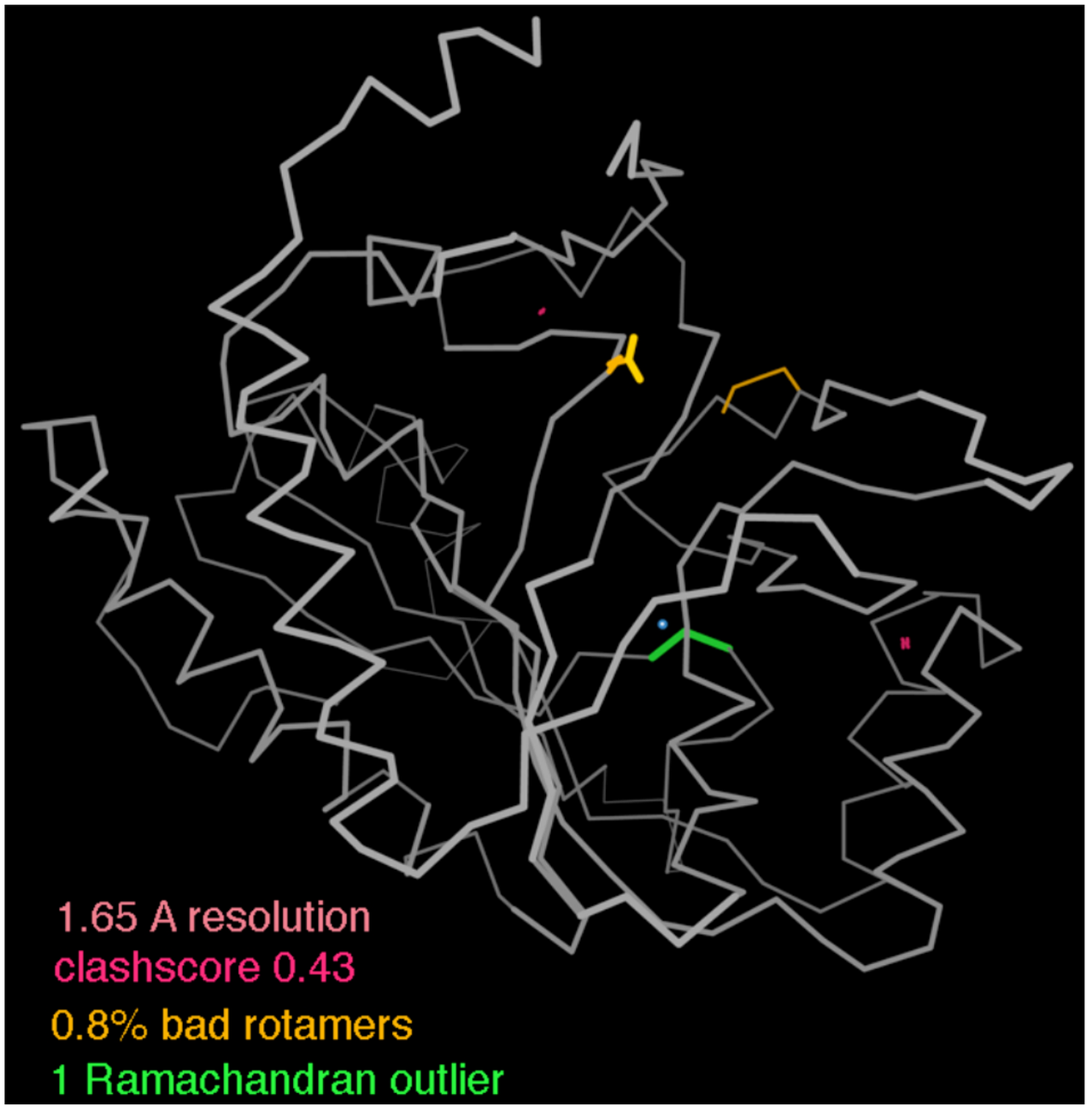
Validation

- Outlier lists recenter Coot view; Probe dots automatically loaded



Using Validation Tools Improves Models





Images from Jane and David Richardson, Duke University

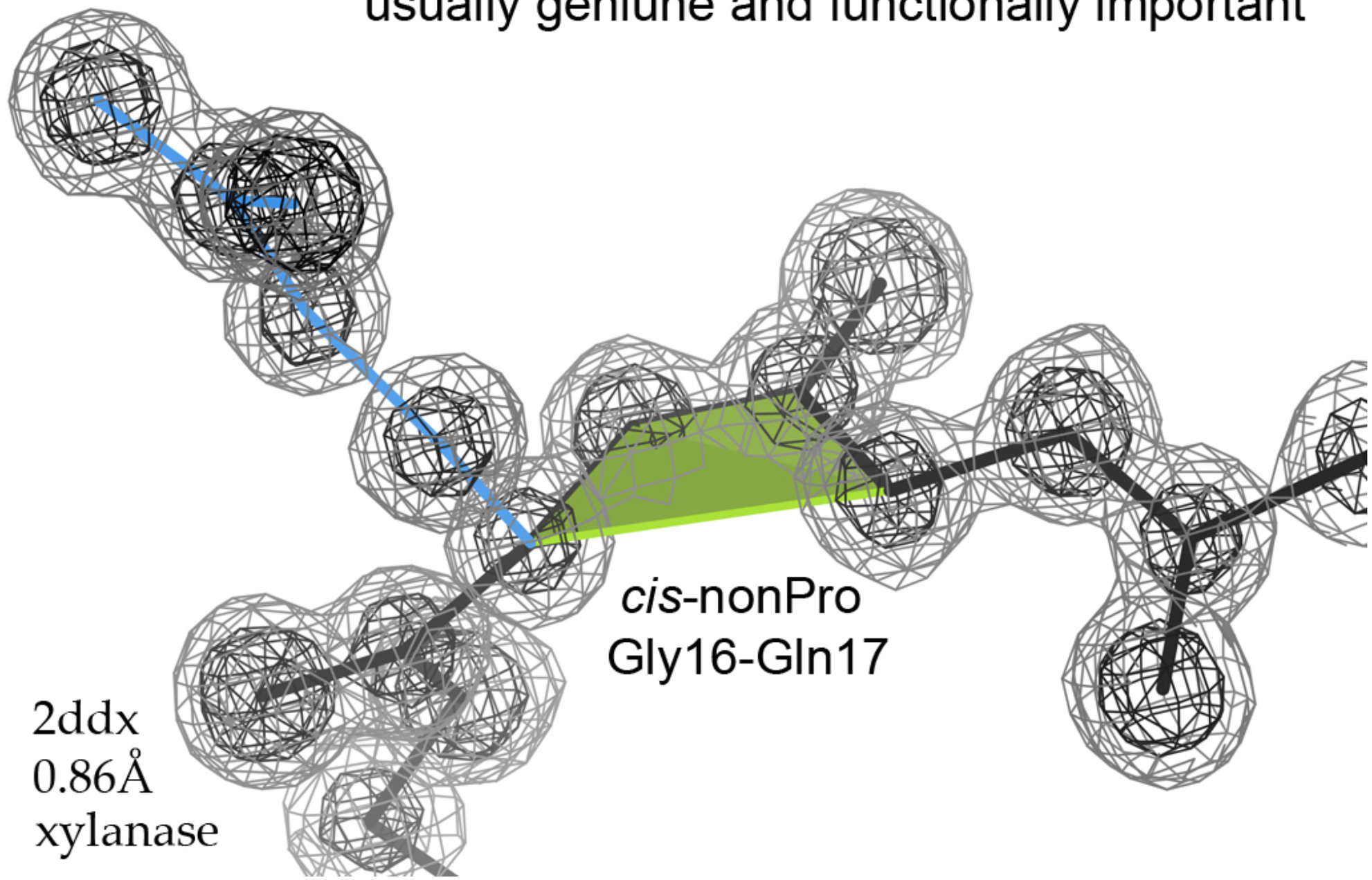
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Cis-Peptides

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

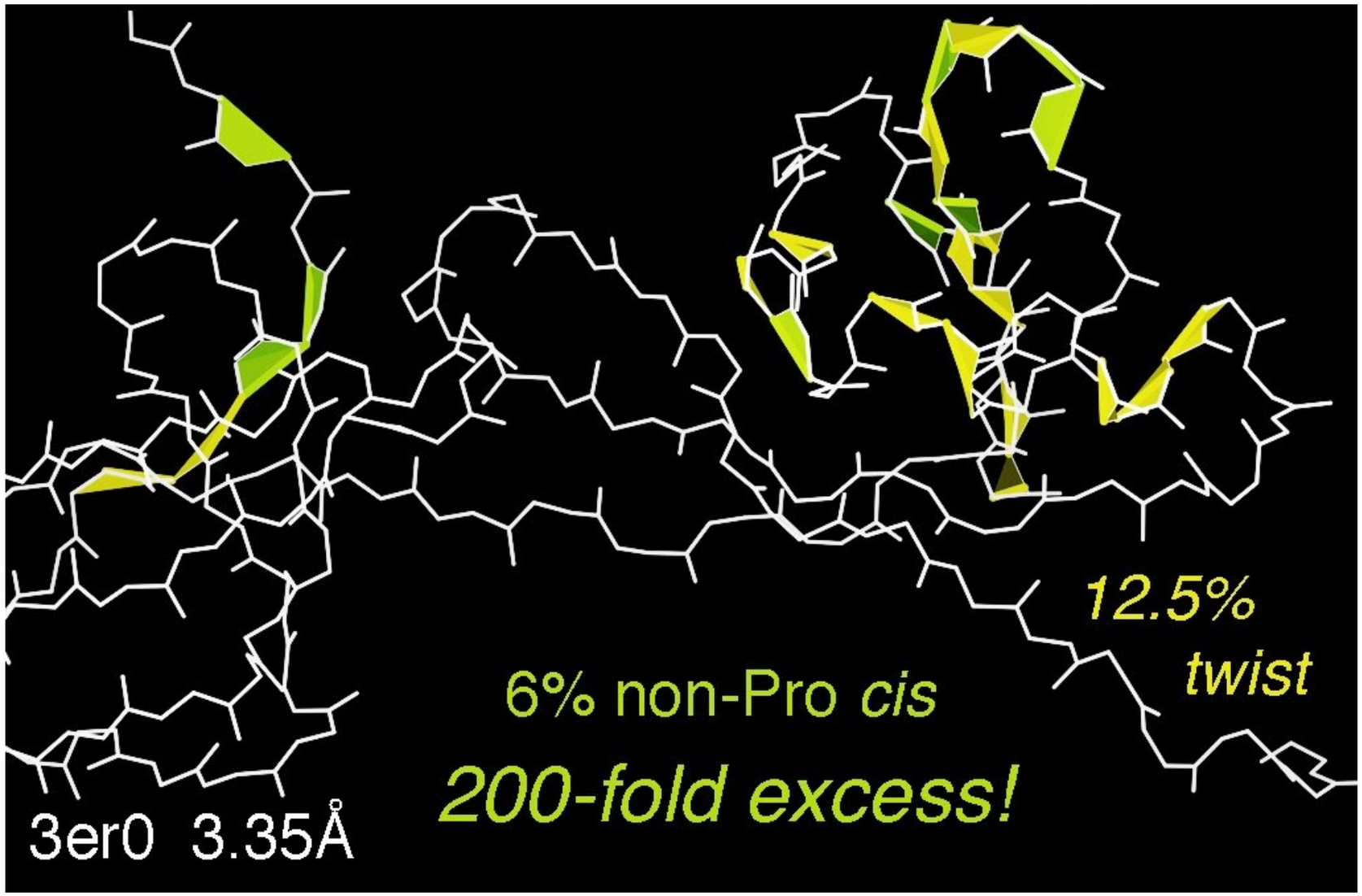


BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Too Many Cis-Peptides



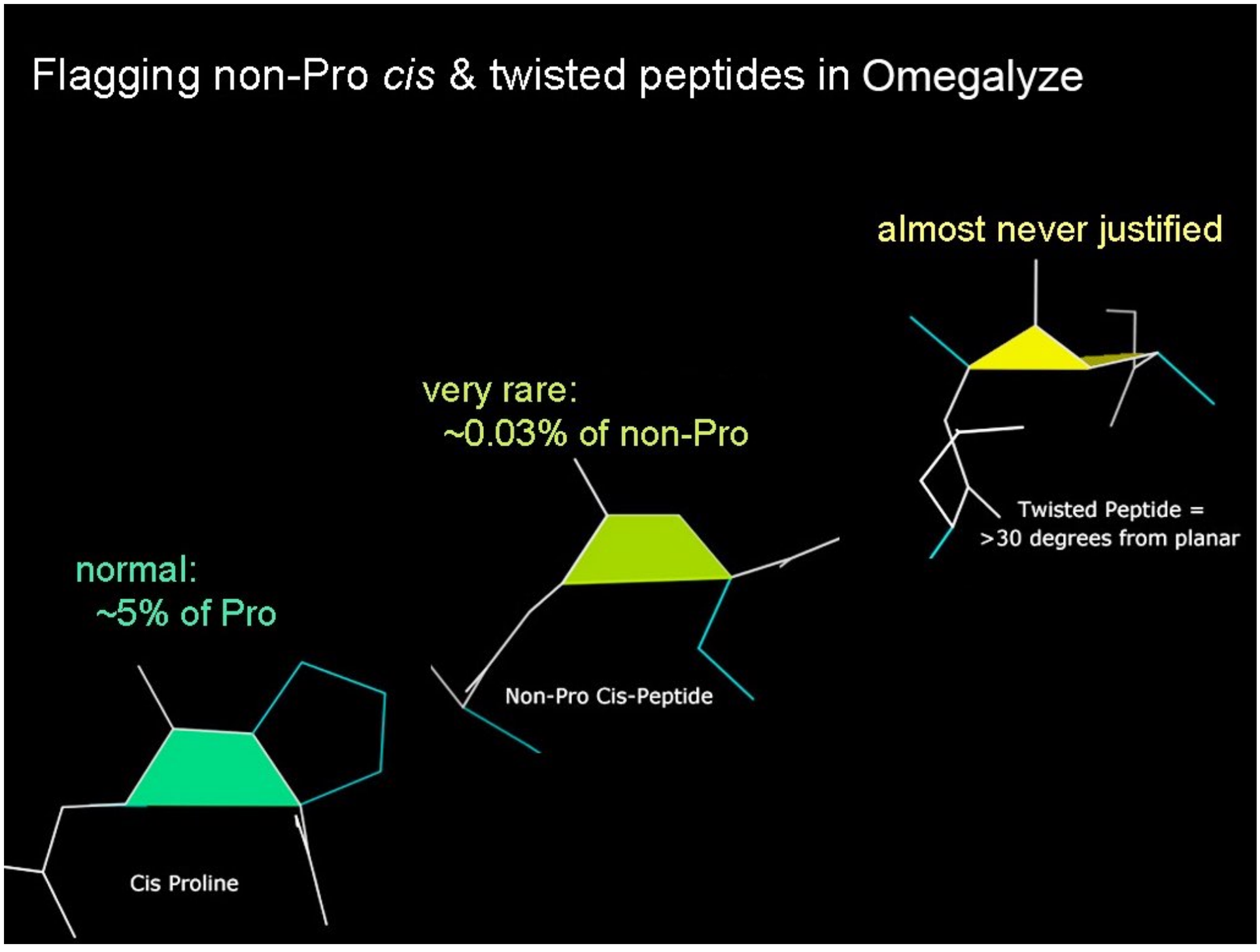
- Cis non-Prolines are chosen much more often than chance, because they are more compact than trans and fit better into the shrunken & rather featureless low-resolution density (esp. for loops).
- Automated building with a no-cis fragment library (Tom Terwilliger)

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Omegalyze

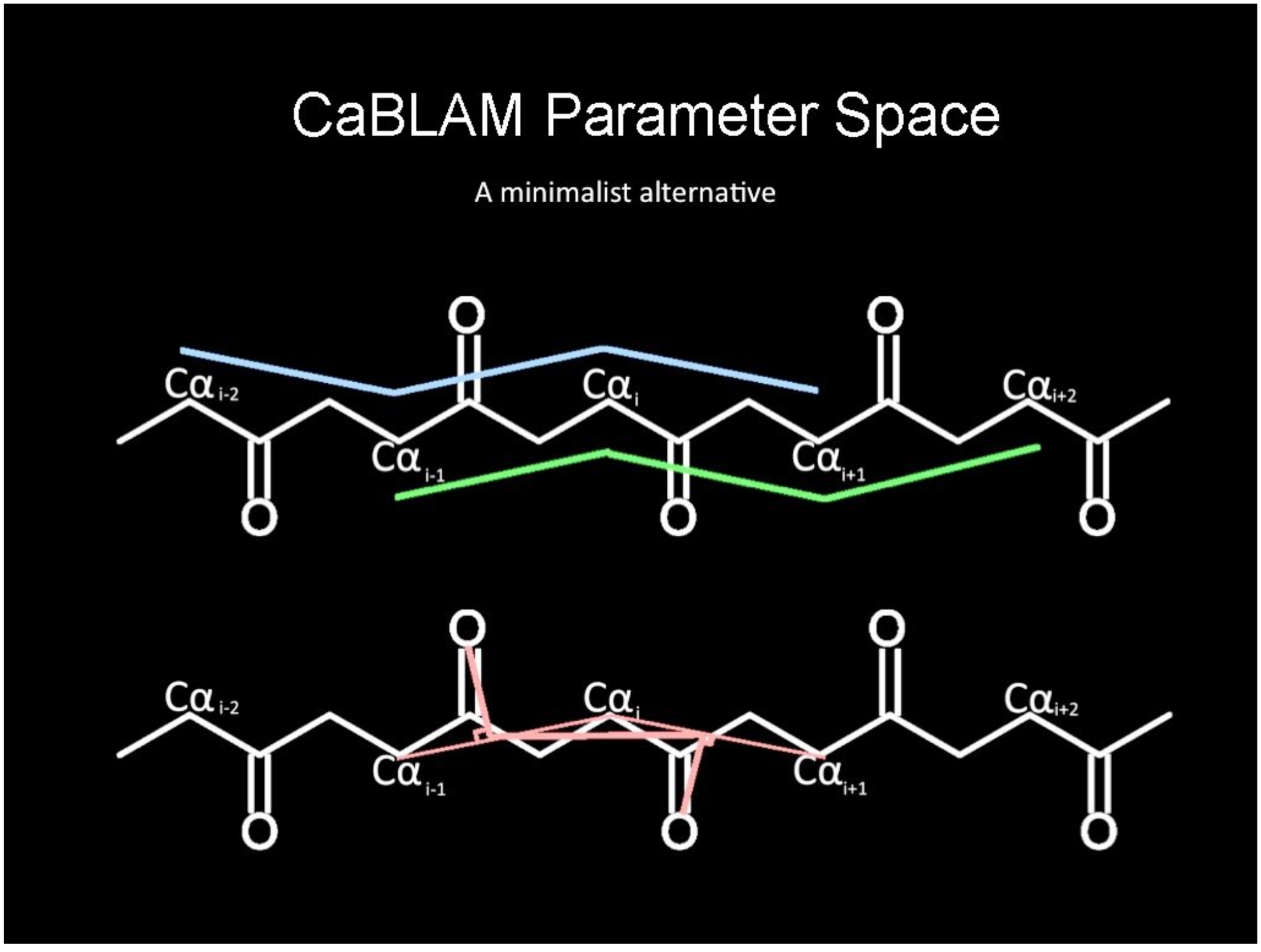


BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Christopher Williams,
Duke University

Validation Using Cα Atoms

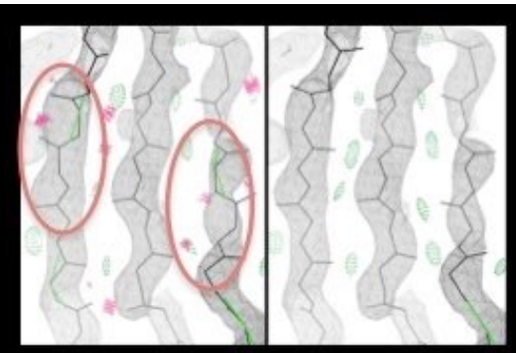


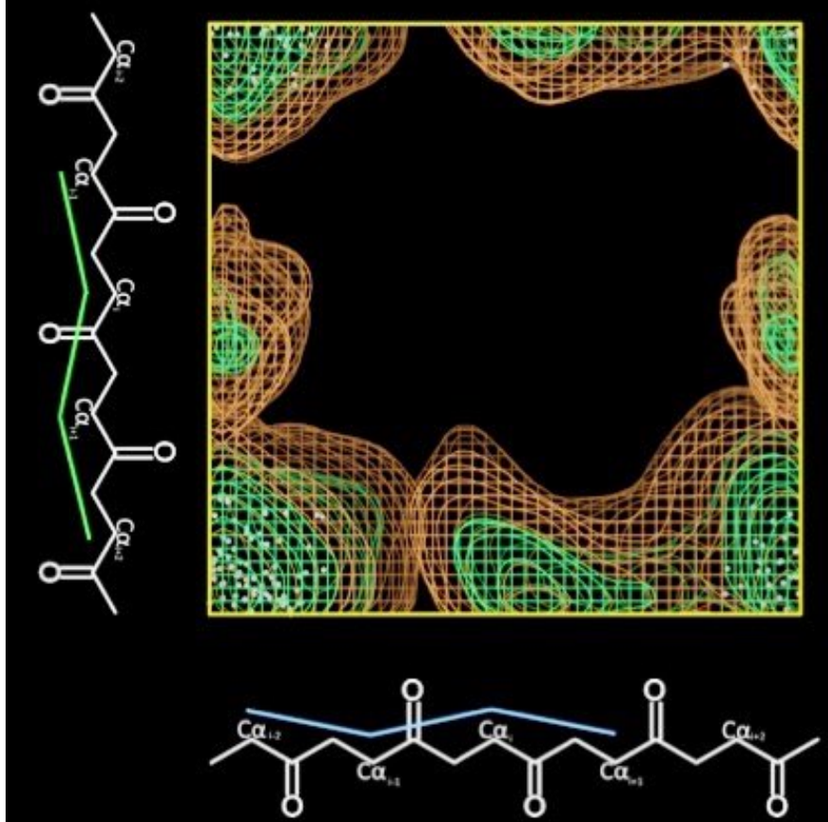
Christopher Williams,
Duke University

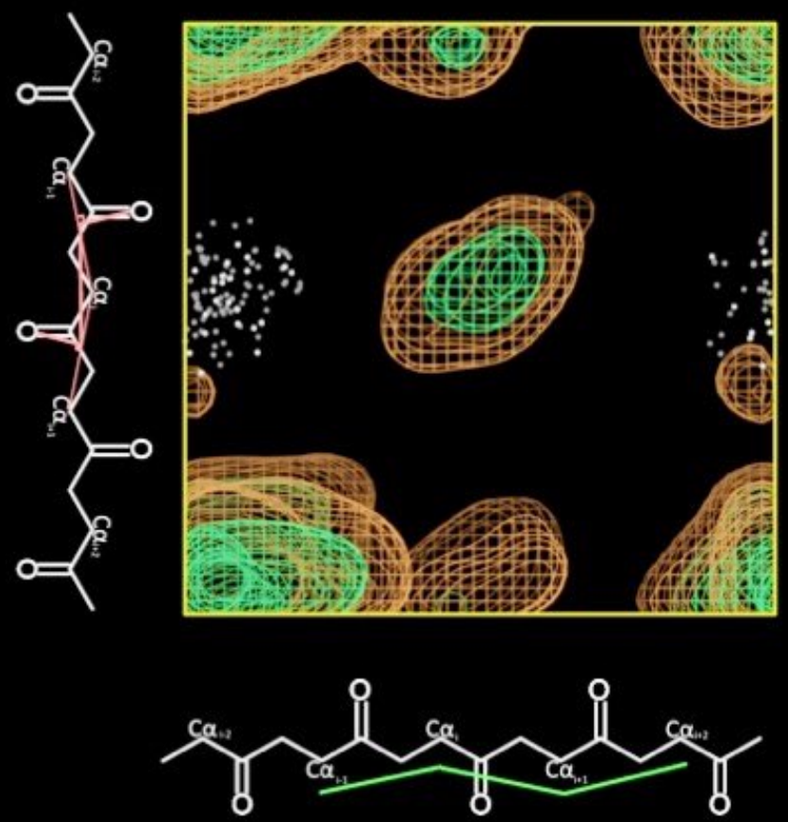
Identifying Distorted Secondary Structure

Diagnosing Strands

Pathological strands from 70S Ribosome



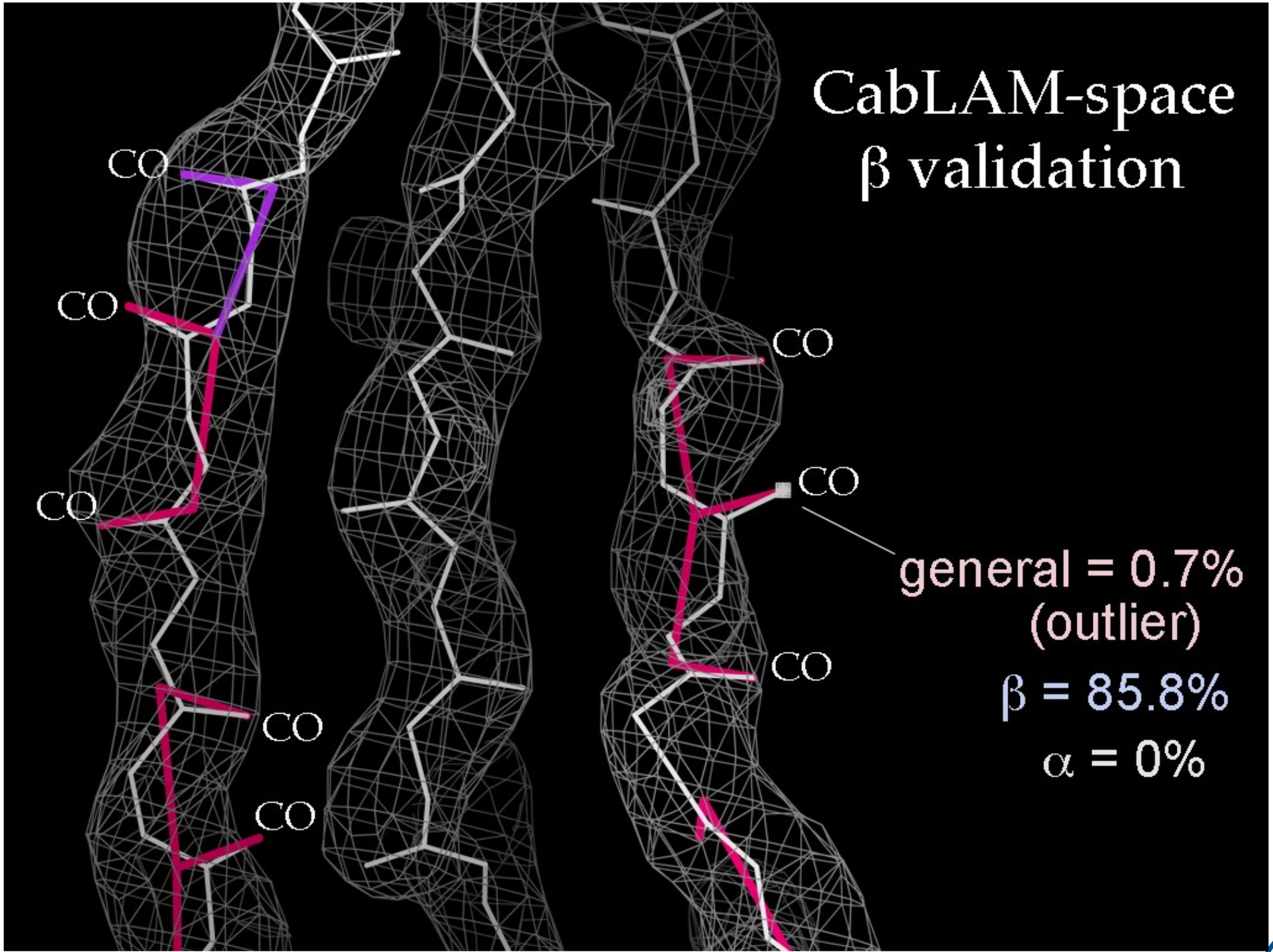




Phenix

Christopher Williams,
Duke University

Assessing Secondary Structure Probability



BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Christopher Williams,
Duke University

Comprehensive Validation

Comprehensive validation (CryoEM) (Project: rea-space-refine-6crz)

Preferences Help Run Abort Ask for help

Input/Output | ValidationCryoEM_7

Summary | Model | Model vs. Data | Data

Files

Model: /Users/PDAdams/Documents/rea-space-refine-6crz/mo

Map: /Users/PDAdams/Documents/rea-space-refine-6crz/ma

Open in Coot

White cells are mostly informational.
Green cells imply that the values are in an acceptable range.
Yellow cells imply that the values need to be checked carefully.
Red cells imply that the values are cocerning and that the model sho
Clicking on a row will bring up a panel with more detailed informatio

Model

MolProbity

MolProbity score	1.72	
Clash score	5.44	
Rotamer outliers (%)	0.00	(Goal: < 1%)
Cβ outliers	0	(Goal: 0)

CaBLAM

Outliers (%)	3.88	(Goal: <= 1%)
Disfavored (%)	8.96	(Goal: <= 5%)
Cα outliers (%)	1.19	(Goal: <= 0.5%)

Geometry Restraints

Bond **Angle**

Idle

Comprehensive validation (CryoEM) (Project: rea-space-refine-6crz)

Preferences Help Run Abort Ask for help

Input/Output | ValidationCryoEM_7

Summary | Model | Model vs. Data | Data

MolProbity | Rotamers | Ramachandran | Clashes | Geometry Restraints

These statistics are computed using the same underlying distributions as the MolProbity web server. The overall score represents the experimental resolution expected for a model of this quality; ideally the score should be lower than the actual resolution.

Overall scores

MolProbity score: 1.72 **Clash score:** 5.44

CaBLAM

Outliers (%): 3.88 **Disfavored (%):** 8.96 **Cα outliers (%):** 1.19

Chain	Residue	Evaluation	CaBLAM Score	CA Geometry Score	Secondary Struct
A	ILE 955	CaBLAM Disfav...	0.03762	0.01447	
A	PRO 969	CaBLAM Disfav...	0.02931	0.46424	try alpha helix
A	SER 1012	CaBLAM Outlier	0.00273	0.67504	try alpha helix
A	LEU 1016	CaBLAM Outlier	0.00086	0.07553	

Cβ deviation analysis

No Cβ position outliers detected.

Cis and twisted peptides

Cis conformations are observed in about 5% of Prolines.
Cis conformations are observed in about 0.03% of general residues.
Twisted peptides are almost certainly modeling errors.

No non-trans peptides detected.

Idle Project: rea-space-refine-6crz

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

Map Resolution and Map/Model Fit

Summary of map resolution estimates.

Metric	Objects used	Purpose	Values	Meaning, possible actions
d_{FSC}	Half-maps	Highest resolution at which the experimental data are confident	The higher the better	Resolution determined using half-maps method
d_{99}	Map	Resolution cutoff beyond which Fourier coefficients are negligibly small	$d_{99} \geq d_{FSC}$	Expected values
			$d_{99} < d_{FSC}$	Verify d_{FSC}; omit coefficients with $d_{99} \leq d < d_{FSC}$
			$d_{99} >> d_{FSC}$	Sharpen the map
d_{model}	Map and model	Resolution cutoff at which the model map is the most similar to the target map	$d_{model} \geq d_{FSC}$	Expected values
			$d_{model} < d_{FSC}$	Verify d_{FSC}; check ADP (too large?); validate map details
			$d_{model} >> d_{FSC}$	Sharpen the map
			$d_{model} << d_{99}$	Check ADP (too large?)
			$d_{model} >> d_{99}$	Check ADP (too small?); check the model
d_{FSC_model}	Map and model	Resolution cutoff up to which the model and map Fourier coefficients are similar	$d_{FSC\_model} \geq d_{FSC}$	Expected values
			$d_{FSC\_model} < d_{FSC}$	Verify d_{FSC}; omit coefficients with $d_{FSC\_model} \leq d < d_{FSC}$
			$d_{FSC\_model} \geq d_{FSC}$	Sharpen the map
			$d_{FSC\_model} >> d_{model}$	Omit coefficients with $d_{model} \leq d < d_{FSC\_model}$
			$d_{FSC\_model} << d_{model}$	Sharpen the map

Summary of map correlation coefficients used in this work.

Metric	Region of the map used in calculation	Purpose
CC_{box}	Whole map	Similarity of maps
CC_{mask}	Jiang & Brünger (1994) mask with a fixed radius	Fit of the atomic centers
CC_{volume}	Mask of points with the highest values in the model map	Fit of the molecular envelope defined by the model map
CC_{peaks}	Mask of points with the highest values in the model and in the target maps	Fit of the strongest peaks in the model and target maps
CC_{vr_mask}	Same as CC_{mask} but atomic radii are variable and function of resolution, atom type and ADP	Fit of the atomic images in the given map

Afonine et al: New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Cryst.* 2018, **D74**:814-840.

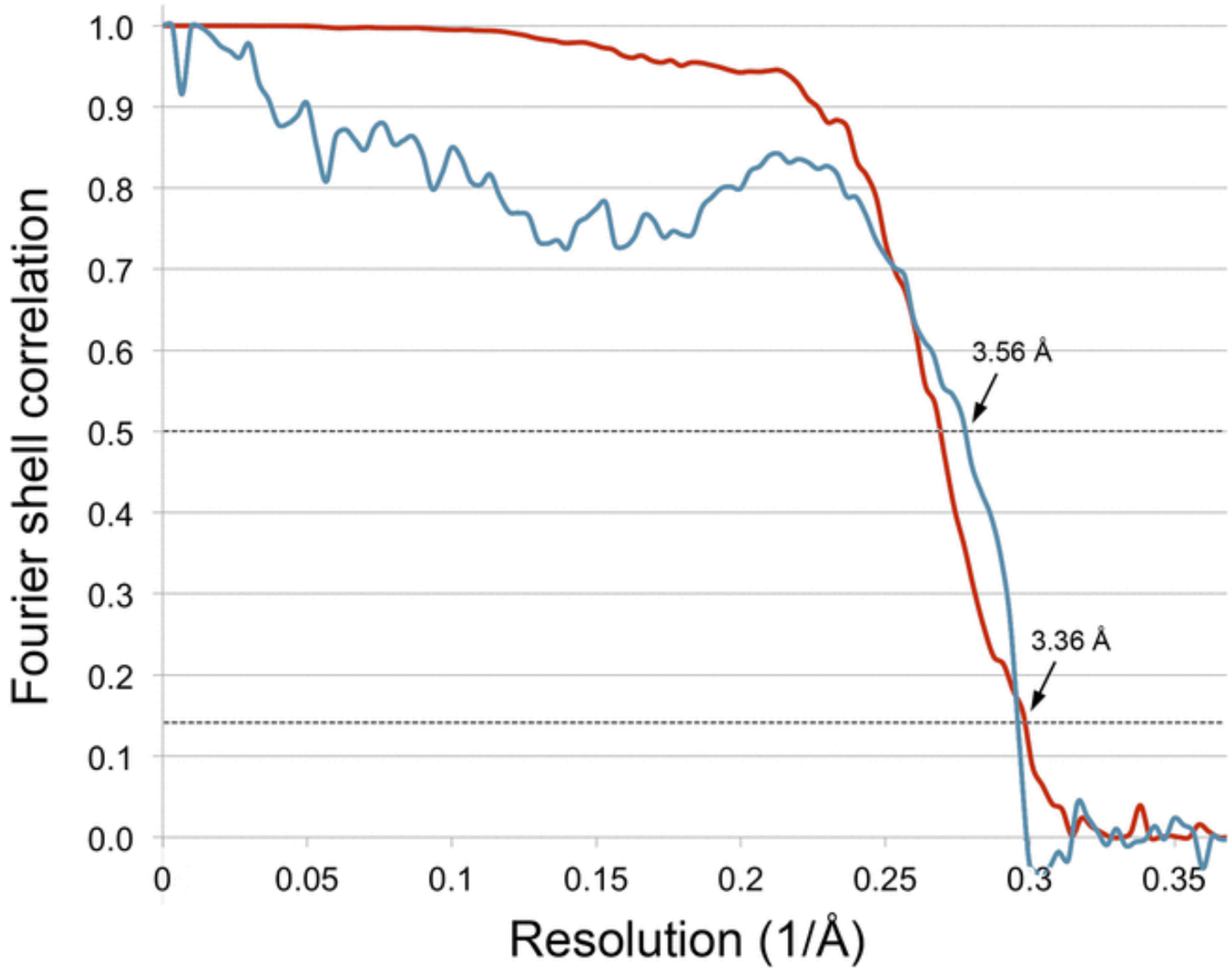
BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA · BERKELEY · 1868

Resolution Determination

$$FSC(r) = \frac{\sum_{r_i \in r} F_1(r_i) \cdot F_2(r_i)^*}{\sqrt[2]{\sum_{r_i \in r} |F_1(r_i)|^2 \cdot \sum_{r_i \in r} |F_2(r_i)|^2}}$$



BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Cross Validation with Half Maps

- Perturb model (random shift of coordinates)
- Re-refine against 1 half map
- Calculate FSC of model against 2nd half map
- FSC curve shouldn't show signal beyond the half map resolution

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

Model/Map Validation

Benjamin Barad, Yifan Cheng, Jaime Fraser
University of California San Francisco
Ray Yu-Ruei Wang, Frank DiMaio
University of Washington
Nat Echols
Lawrence Berkeley National Laboratory

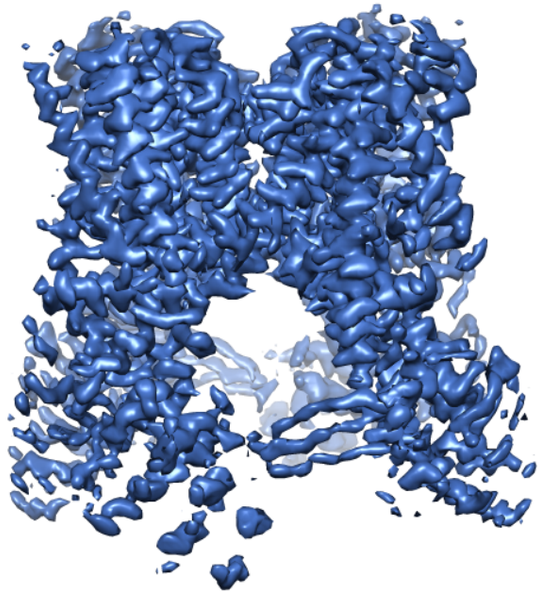
UCSF
University of California
San Francisco

Phenix

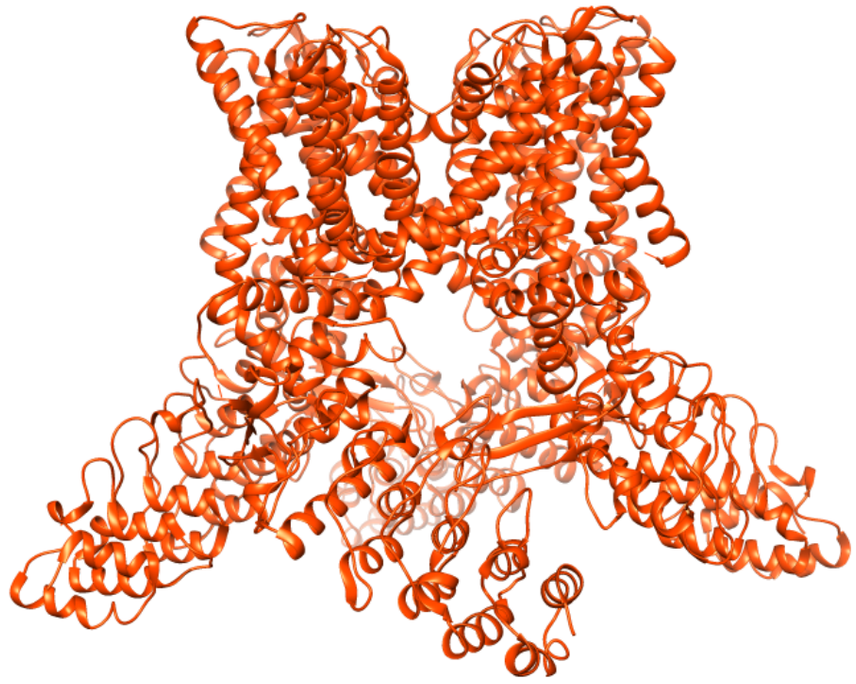
BERKELEY LAB
Lawrence Berkeley National Laboratory

Validation and Cryo-EM

- Do the map make sense?
  - Gold Standard FSC of half maps
- Does the model make sense?
  - MolProbity



- Does the model fit the map?
  - Overall and local density correlation
  - What about the detailed local fit?



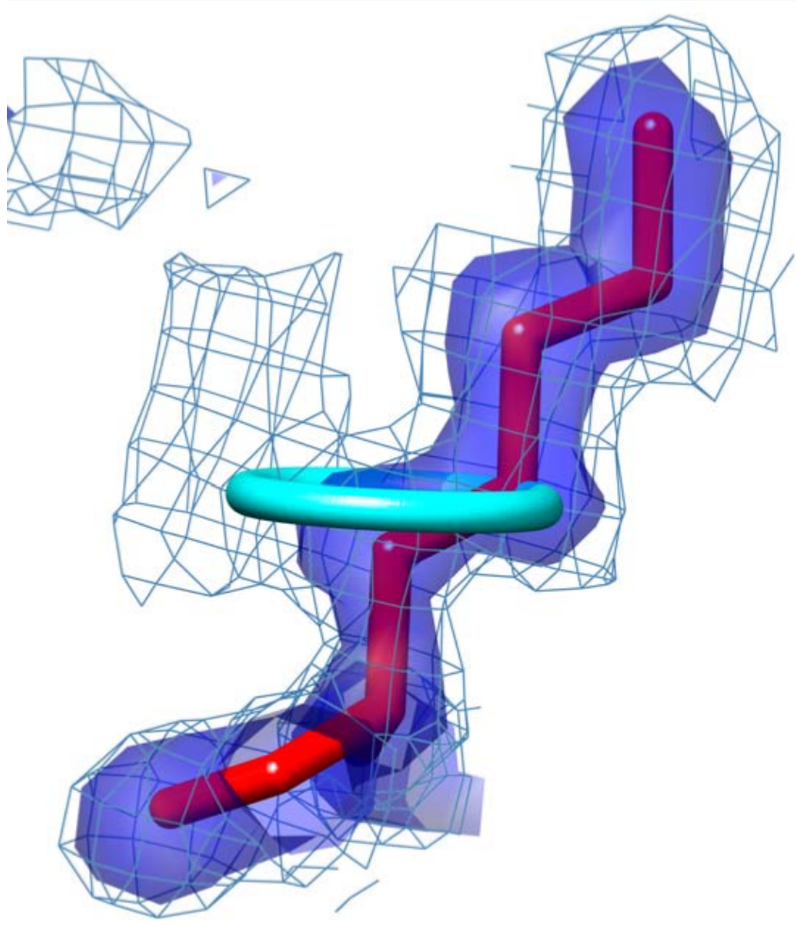
UCSF
University of California
San Francisco

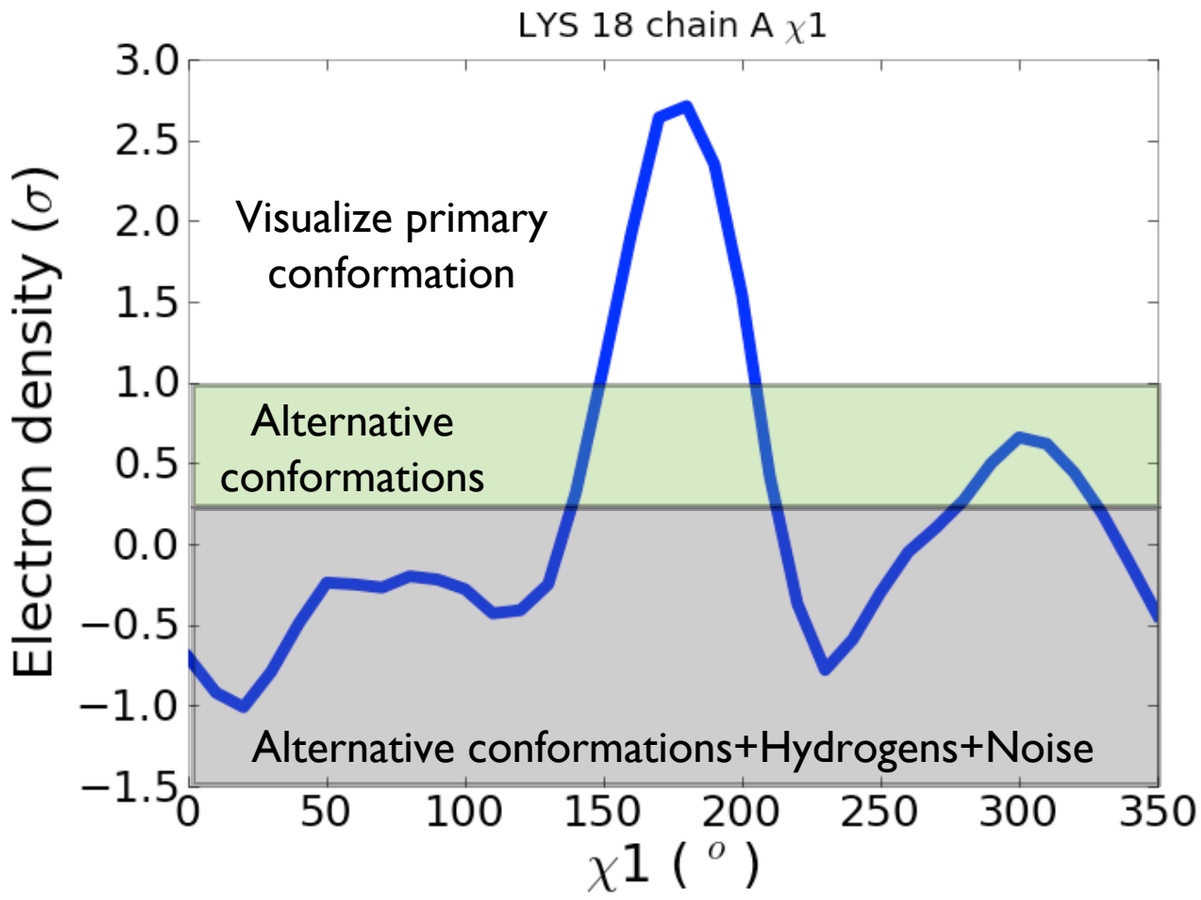
Phenix

BERKELEY LAB
Lawrence Berkeley National Laboratory

Look at the Density Around Sidechains

Ringer





Lang PT, et al. Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Science*. 2010.

UCSF University of California San Francisco

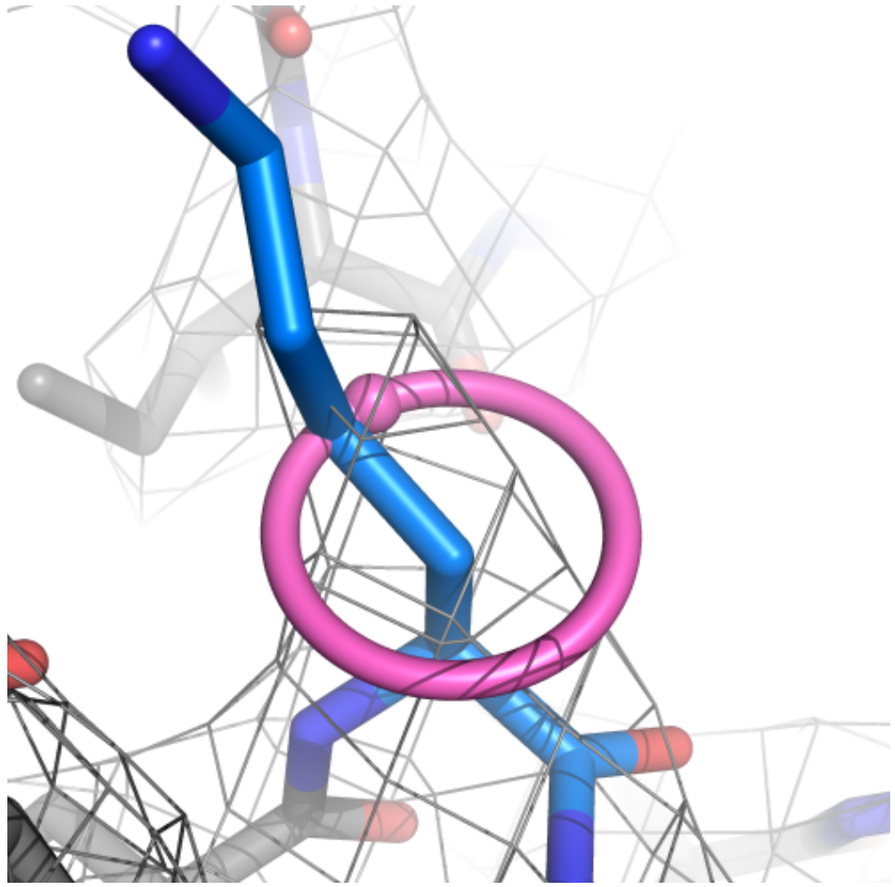
Phenix

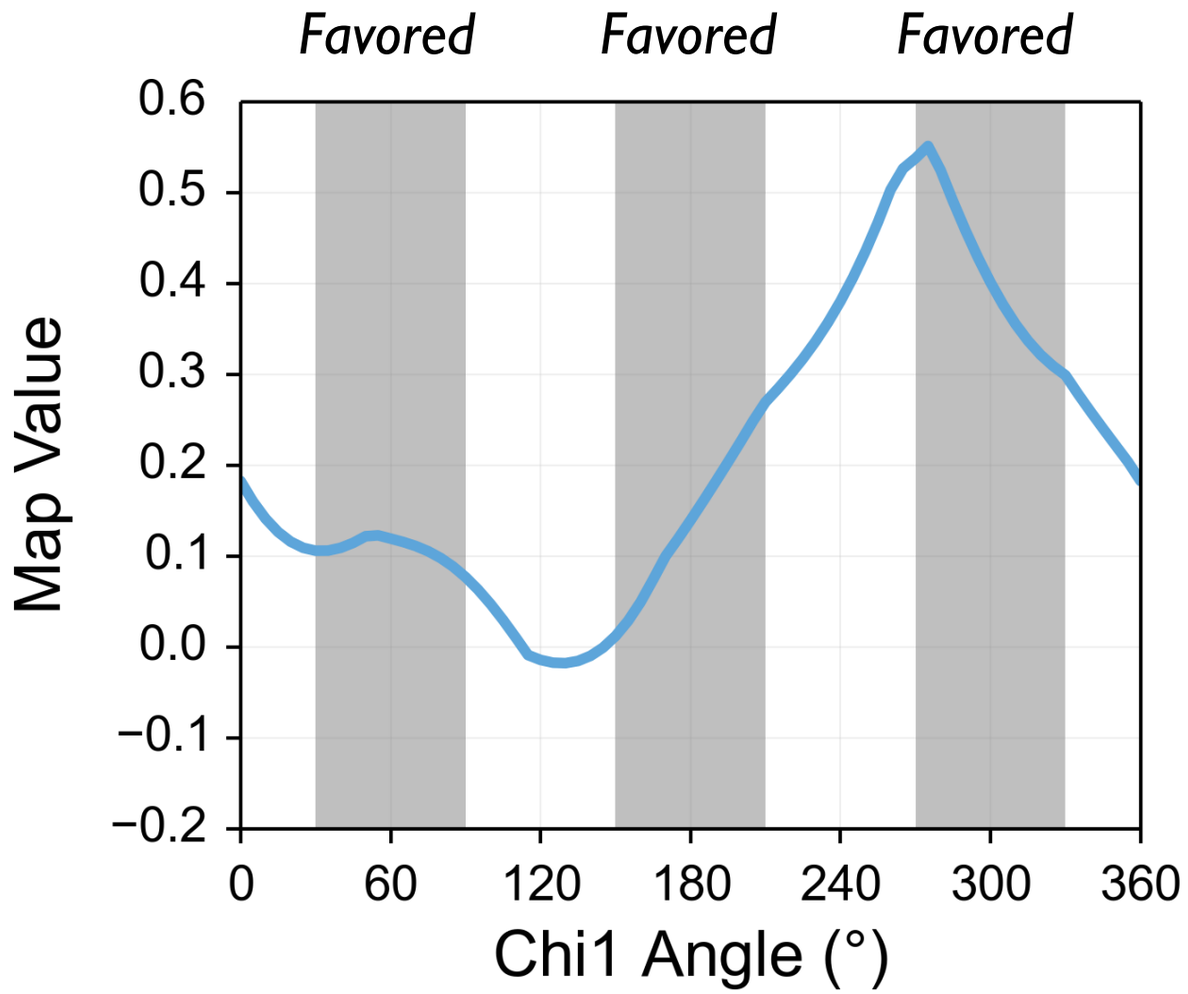
Ben Barad, Jaime Fraser, UCSF

BERKELEY LAB Lawrence Berkeley National Laboratory

Look at the Density Around Sidechains

EMRinger





Barad BA, et al. EMRinger: Side-chain-directed model and map validation for 3D Electron Cryomicroscopy. *Nature Methods*. 2015

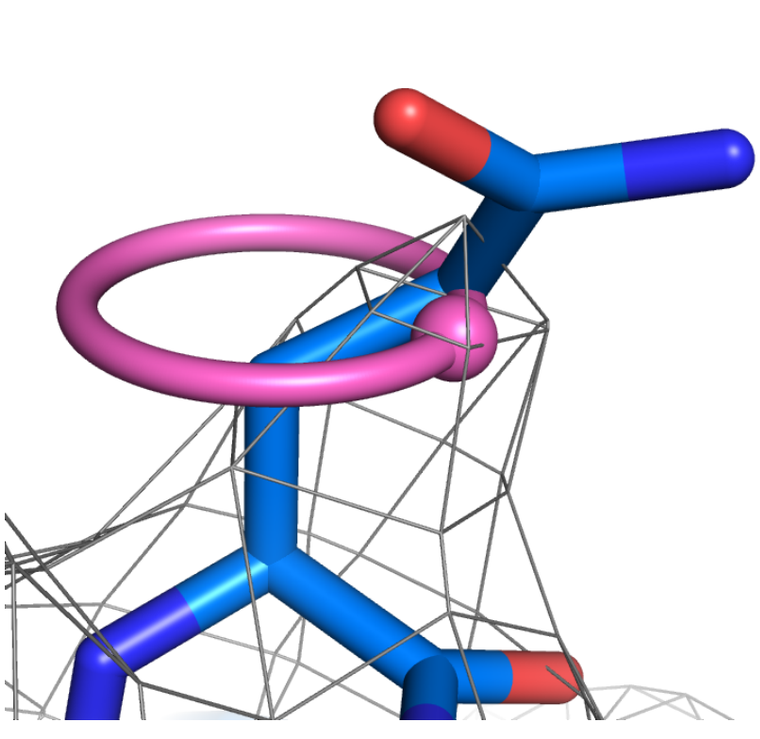
UCSF University of California San Francisco

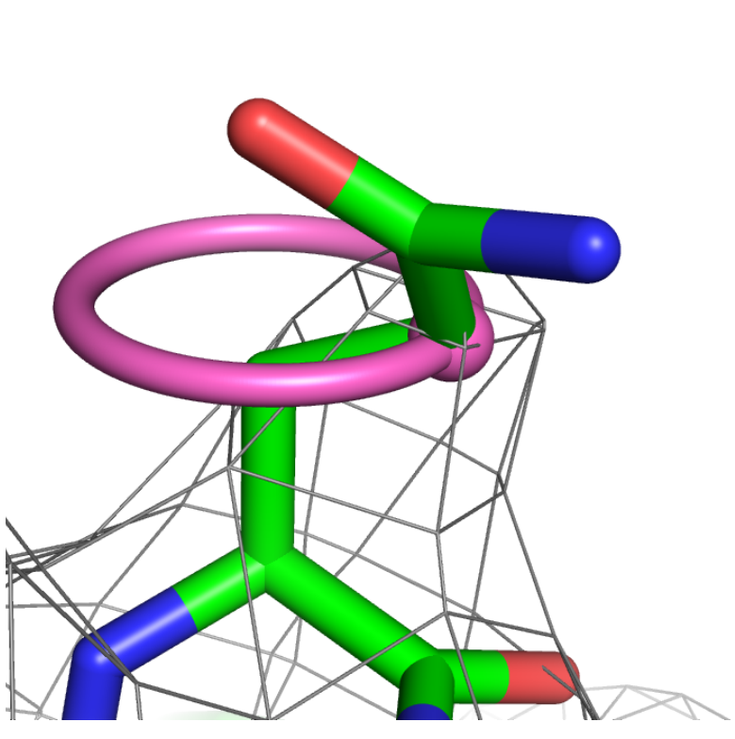
Phenix

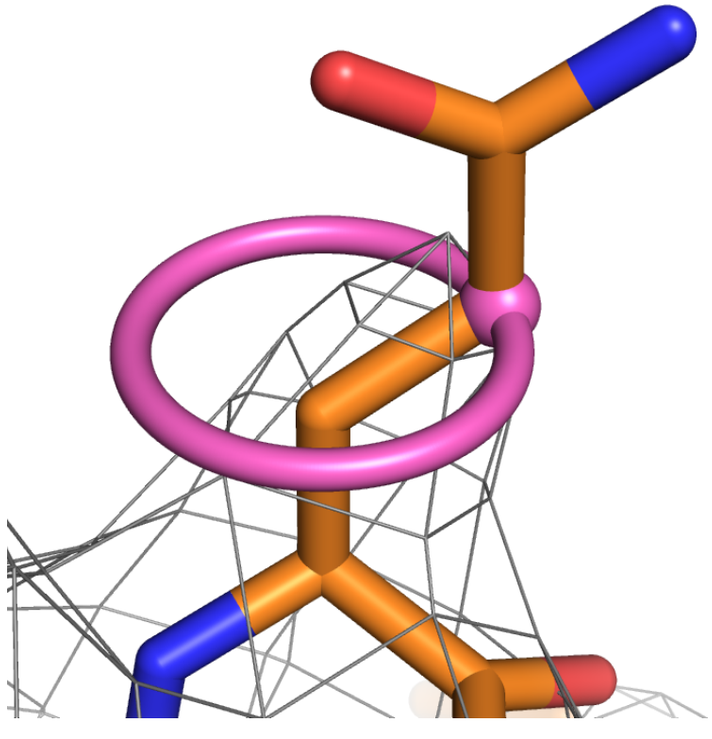
Ben Barad, Jaime Fraser, UCSF

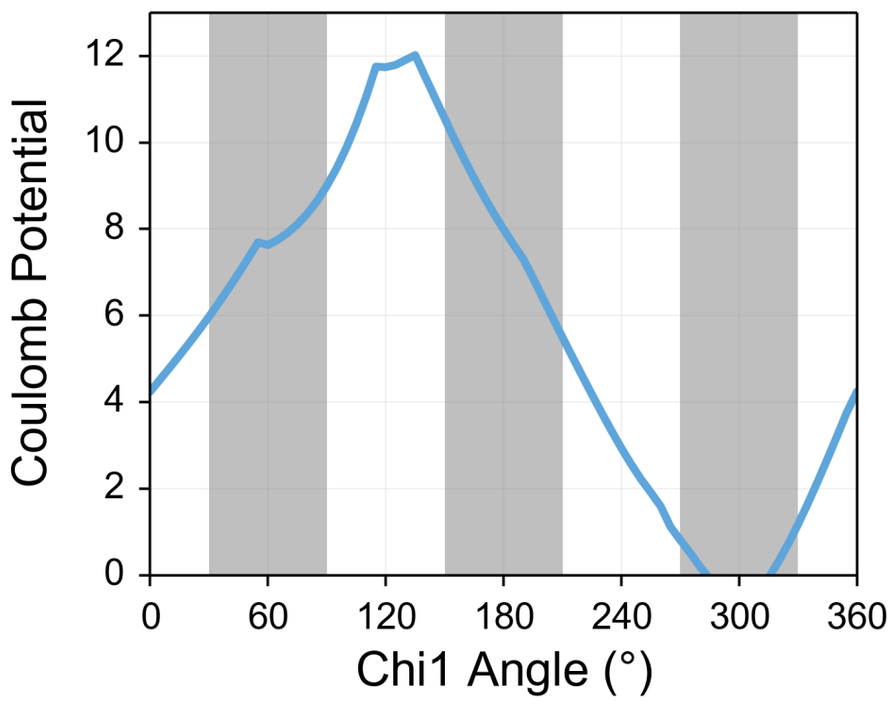
BERKELEY LAB Lawrence Berkeley National Laboratory

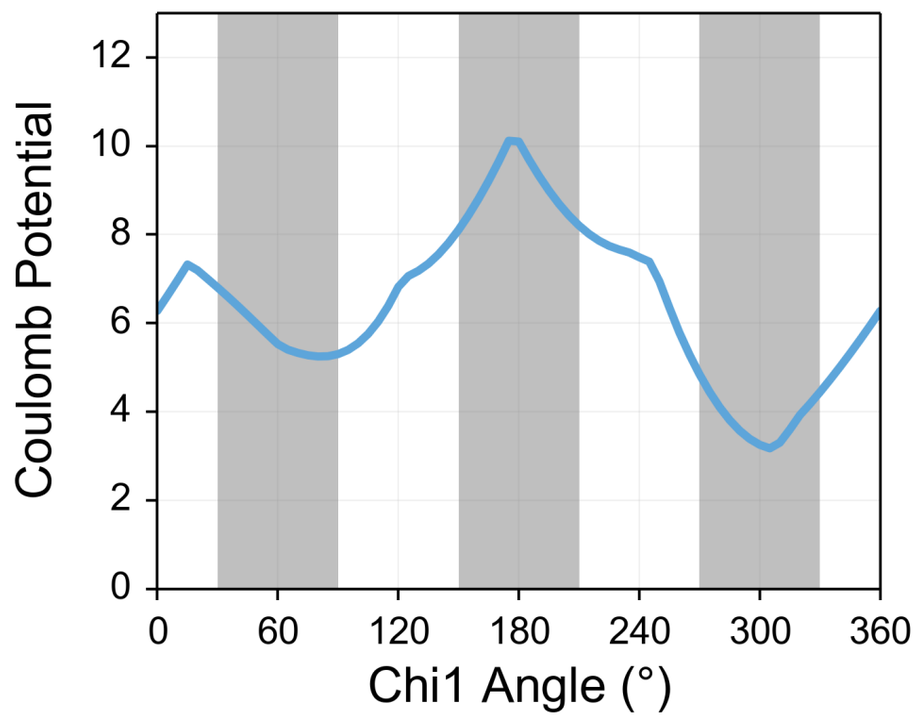
EMRinger reports on backbone placement











UCSF
University of California
San Francisco

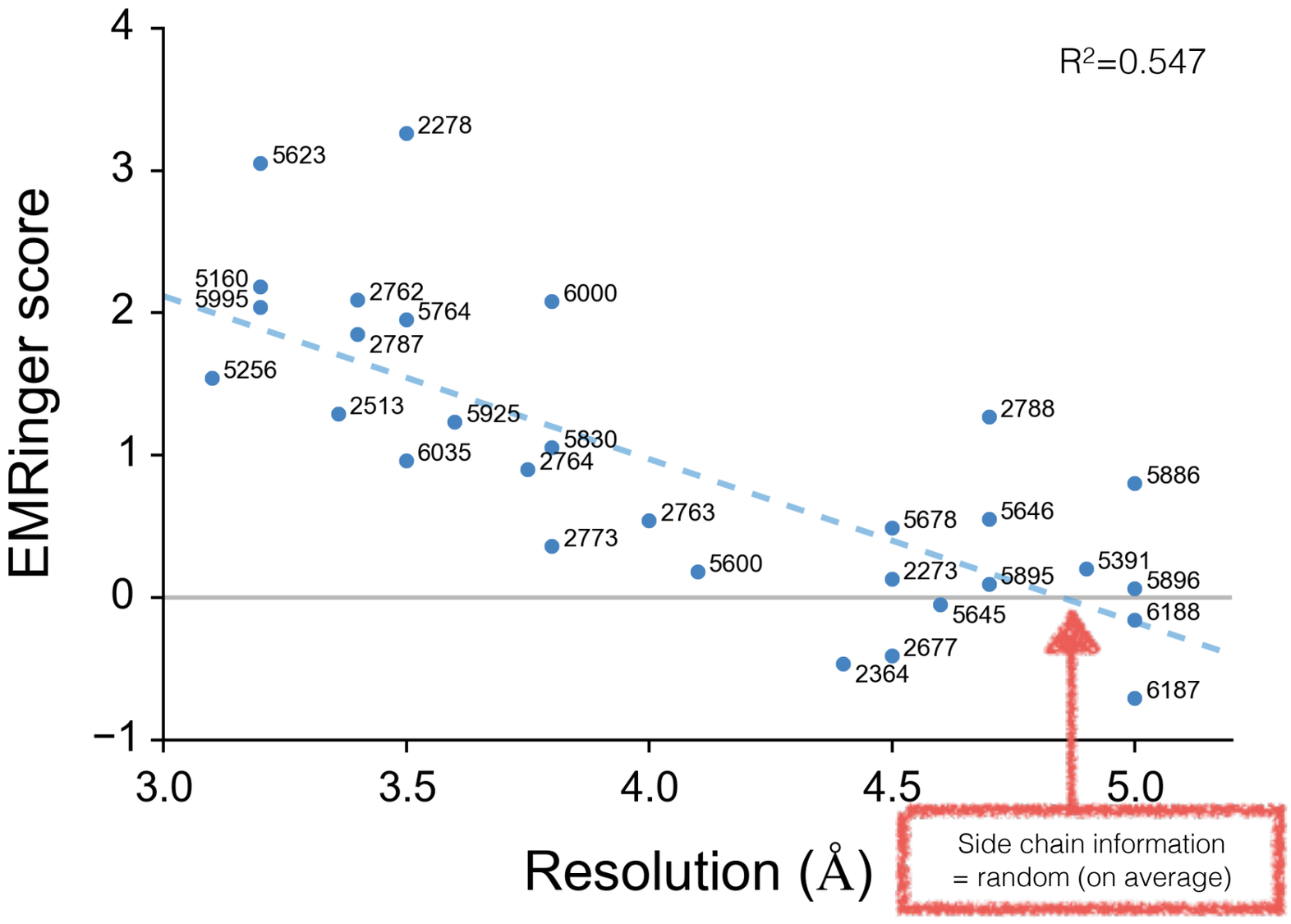
Phenix

Ben Barad, Jaime Fraser, UCSF

BERKELEY LAB
Lawrence Berkeley National Laboratory

EMRinger Score to Validate Model vs Data

- Quantify how well the model backbone puts side chains in places where there are density peaks consistent with rotameric conformations



http://emringer.com

Phenix

- Available in GUI and command line
- `phenix.emringer model.pdb map.ccp4`

University of California San Francisco

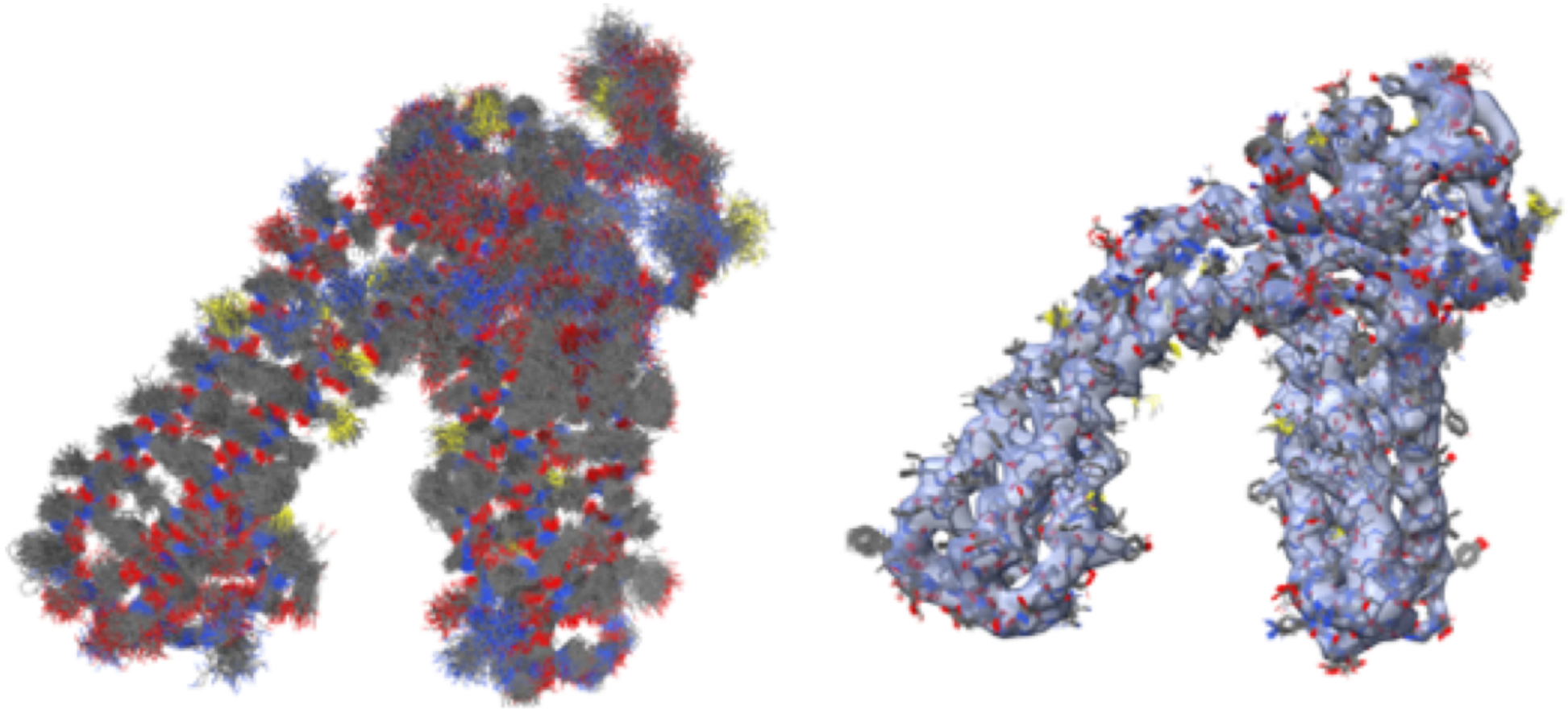
Phenix

Ben Barad, Jaime Fraser, UCSF

BERKELEY LAB
Lawrence Berkeley National Laboratory

Ensembles

- At lower resolution ensemble models are probably more appropriate
- Can be used to help assess map variability (Herzik, Fraser, Lander. Structure. 2019)



BERKELEY LAB
Lawrence Berkeley National Laboratory

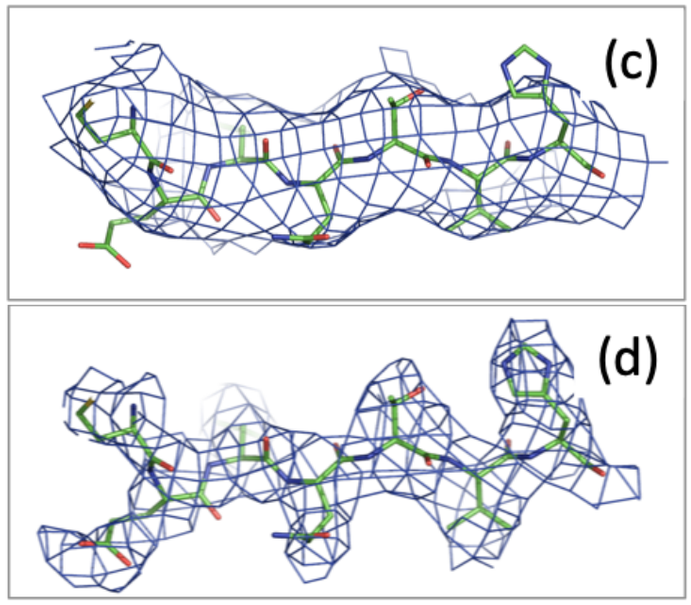
Phenix

THE UNIVERSITY OF CALIFORNIA · BERKELEY · 1868

Deposition Issues

- Successful re-analysis of cryo-EM data relies on accurate data/model deposition
- Current practice has led to significant issues:
  - Models misplaced wrt maps
  - Inconsistent map deposition (sharpened, masked, filtered, wrong map)
  - Absence of half-maps
  - Very variable assessments of resolution
  - Optimistic ligand placement (probably unintentional)

Afonine et al: New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Cryst.* 2018, **D74**:814-840.



BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA · BERKELEY · 1868

Conclusions

- Many of the validation metrics developed to assess models can be readily applied to cryo-EM structures
- Many of the pitfalls of low resolution from other fields apply to cryo-EM
- Care needs to be taken to ensure that validation metrics can be used when restraints are applied in refinement
- Additional validation metrics for the model w.r.t. the data are needed
- We do not have cross-validation metrics for the model/data

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA · BERKELEY · 1868

Acknowledgements

Berkeley Laboratory
Pavel Afonine, Youval Dar, Nat Echols, Jeff Headd, Richard Gildea, Ralf Grosse-Kunstleve, Dorothee Liebschner, Nigel Moriarty, Nader Morshed, Billy Poon, Ian Rees, Nicholas Sauter, Oleg Sobolev, Peter Zwart

Los Alamos Laboratory/New Mexico Consortium
Tom Terwilliger, Li-Wei Hung

Baylor College of Medicine
Matt Baker

Cambridge University
Randy Read, Airlie McCoy, Gabor Bunckozi, Tristan Croll, Rob Oeffner, Kaushik Hatti, Massimo Sammito, Duncan Stockwell, Laurent Storoni

Duke University
Jane Richardson & David Richardson, Ian Davis, Vincent Chen, Jeff Headd, Chris Williams, Bryan Arendall, Bradley Hintze, Laura Murray

UC San Francisco
Ben Barad, Yifan Cheng, Jaime Fraser

University of Washington
Frank DiMaio, Ray Wang, David Baker

Oak Ridge National Laboratory
Marat Mustyakimov, Paul Langan

Other Collaborators
Corey Hryc, Zhao Wang, Wah Chiu
Pawel Janowski, David Case
Dale Tronrud, Donnie Berholz, Andy Karplus
Alexandre Urzhumtsev & Vladimir Lunin
Garib Murshudov & Alexi Vagin
Paul Emsley, Bernhard Lohkamp, Kevin Cowtan
David Abrahams
PHENIX Testers & Users

Funding
- NIH/NIGMS: P01GM063210, P50GM062412, P01GM064692, R01GM071939
- PHENIX Industrial Consortium
- Lawrence Berkeley Laboratory

BERKELEY LAB
Lawrence Berkeley National Laboratory

Phenix

THE UNIVERSITY OF CALIFORNIA · BERKELEY · 1868 · LET THERE BE LIGHT