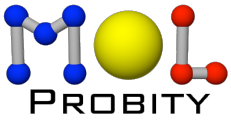
Model Validation in Phenix

Jeffrey J. Headd
Lawrence Berkeley National Lab

Phenix



1

the *PHENIX* project

Lawrence Berkeley Laboratory

Paul Adams, Ralf Grosse-Kunstleve, Pavel Afonine, Nat Echols, Jeff Headd, Nigel Moriarty, Nicholas Sauter, Peter Zwart

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

Duke University

Randy Read, Airlie McCoy, Gabor Bunkoczi, Rob Oeffner

Cambridge University

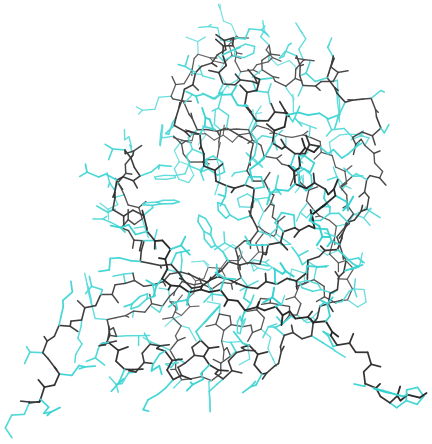
Jane & Dave Richardson, Vincent Chen, Chris Williams, Swati Jain, Bradley Hintze

An NIH/NIGMS funded Program Project

Phenix

2

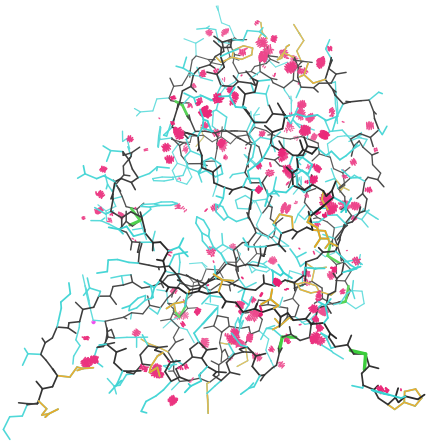
What is model validation?



Cyclic Nucleotide Phosphodiesterase (2.4 Å)

3

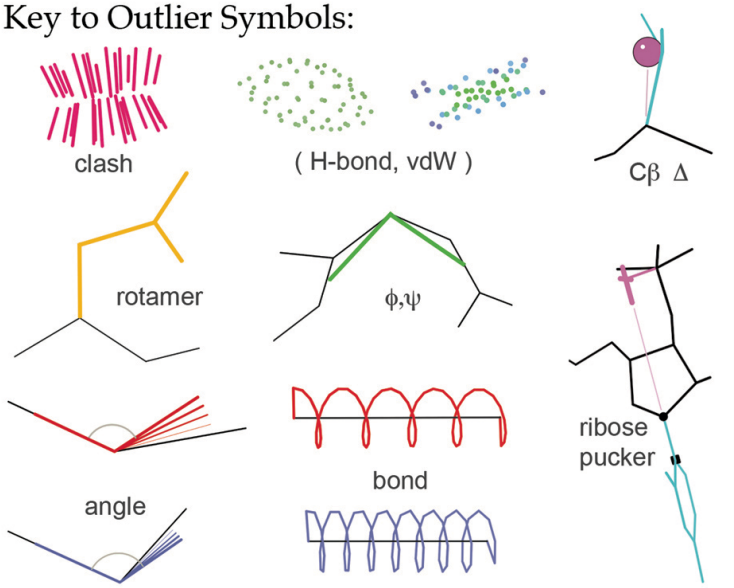
What is model validation?



Cyclic Nucleotide Phosphodiesterase (2.4 Å)

4

Key to Outlier Symbols:



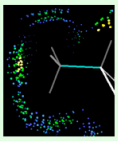
5

I. MolProbity Validation Criteria

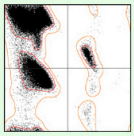
6

MolProbity Structure Improvement Criteria

All-atom contacts, clashscore

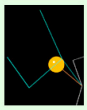


Ramachandran criteria

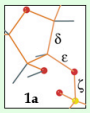


Sidechain rotamers

Geometry



RNA bb

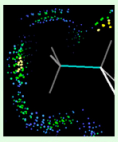


Crystallographic: R_{free}, electron density fit

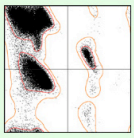
7

MolProbity Structure Improvement Criteria

All-atom contacts, clashscore

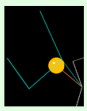


Ramachandran criteria

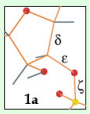


Sidechain rotamers

Geometry

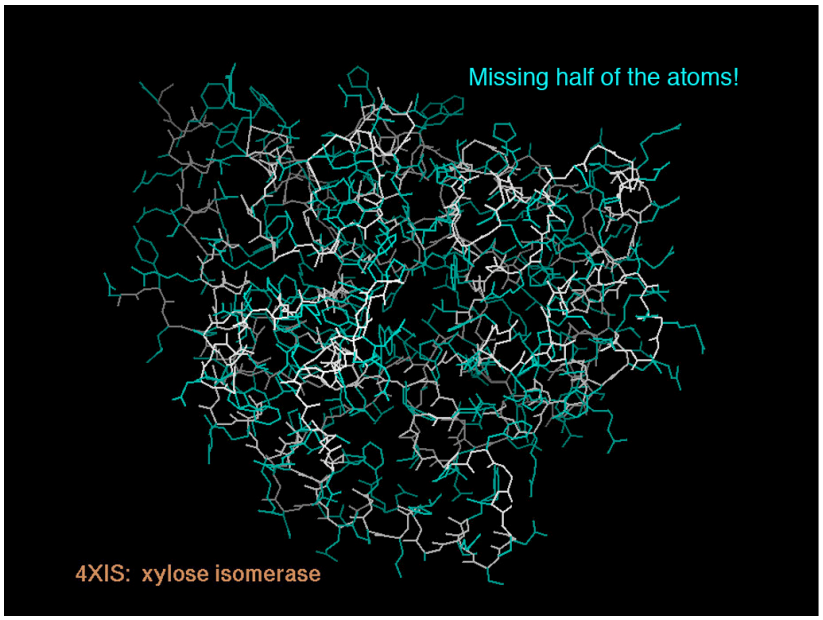


RNA bb

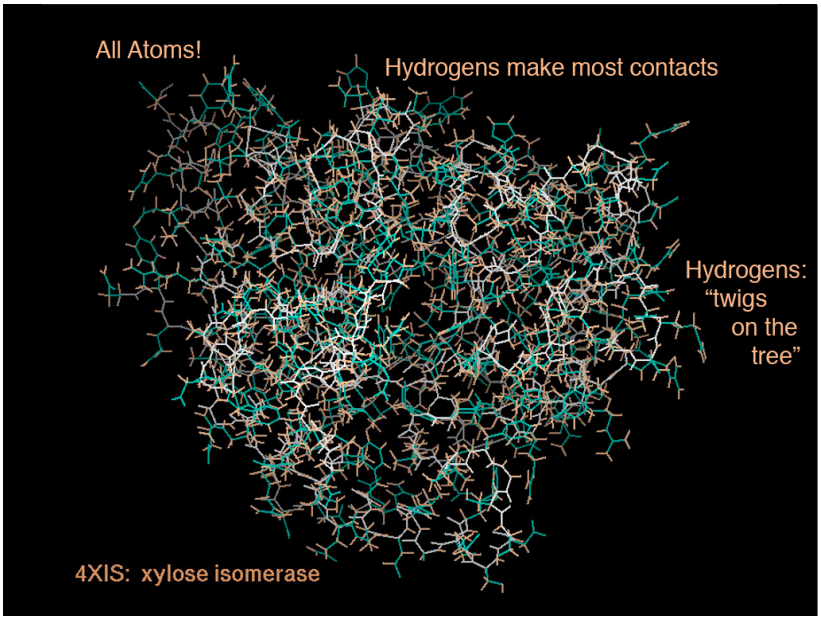


Crystallographic: R_{free}, electron density fit

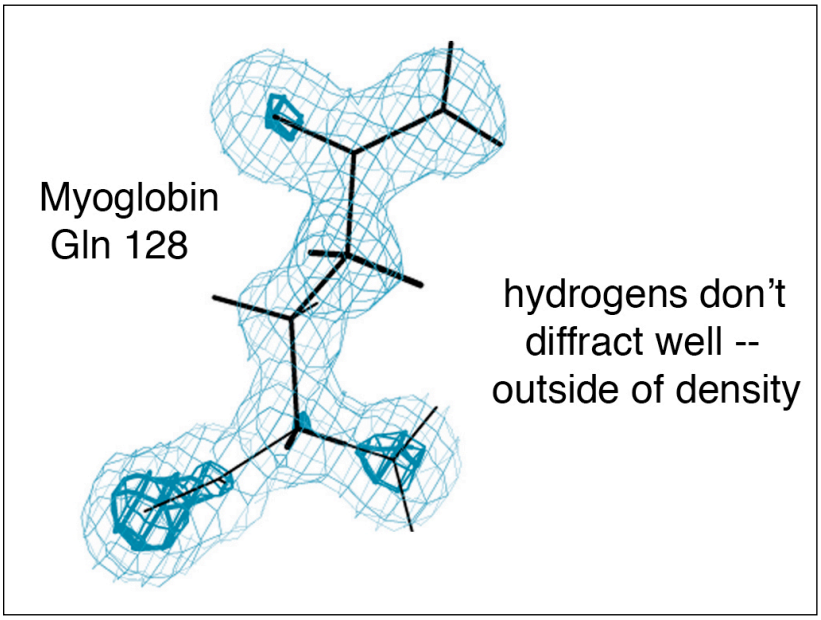
8



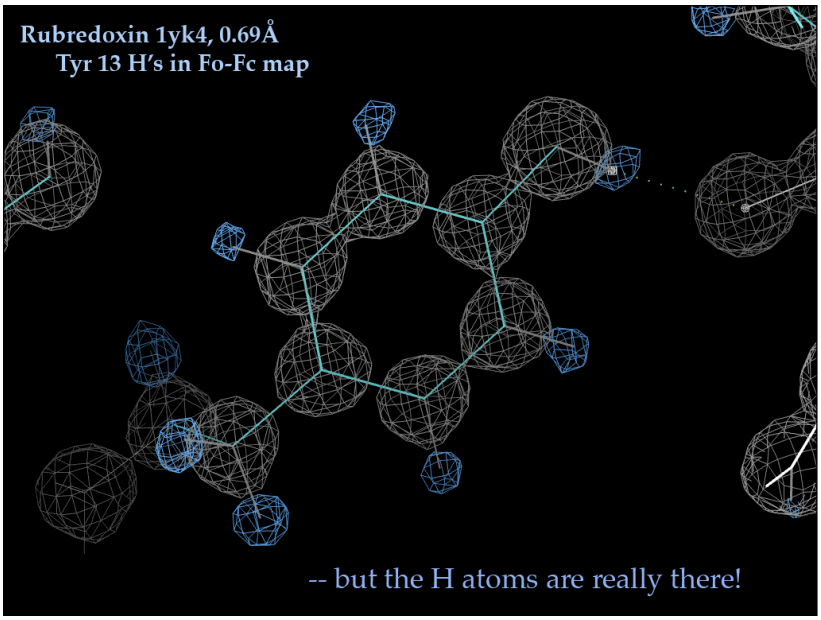
9


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

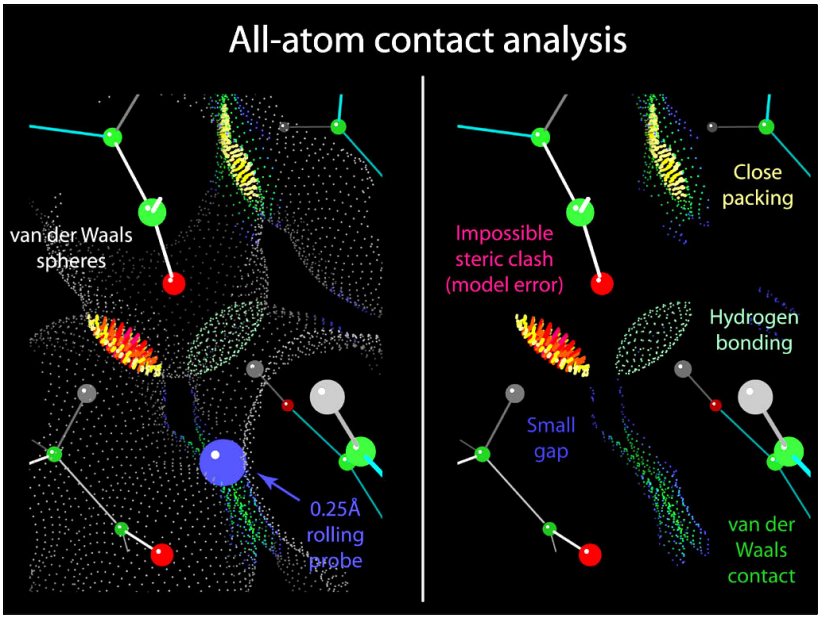
10


Myoglobin
Gln 128
hydrogens don't
diffract well --
outside of density

11


Rubredoxin 1yk4, 0.69Å
Tyr 13 H's in Fo-Fc map
-- but the H atoms are really there!

12

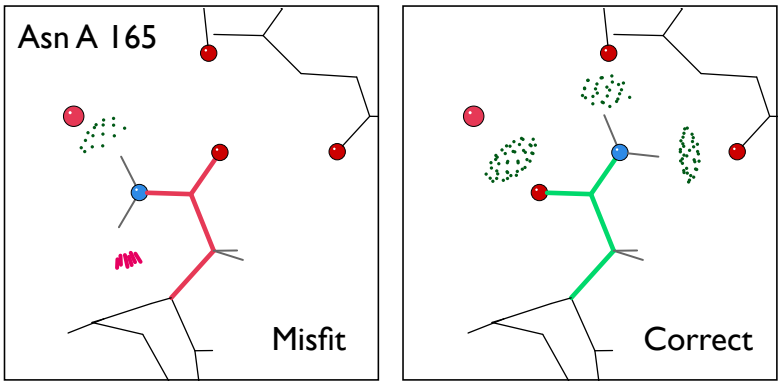


13

Asn / Gln / His Correction

* Automatically detect and correct flipped N/Q/H residues at each macrocycle

* Uses MolProbity/Reduce methodology (H-bonds, clashes) to determine correct orientation

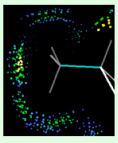


Sulfate Binding Protein (1SBP)

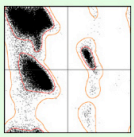
14

MolProbity Structure Improvement Criteria

All-atom contacts, clashscore

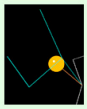


Ramachandran criteria

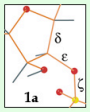


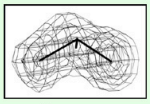
Sidechain rotamers

Geometry



RNA bb





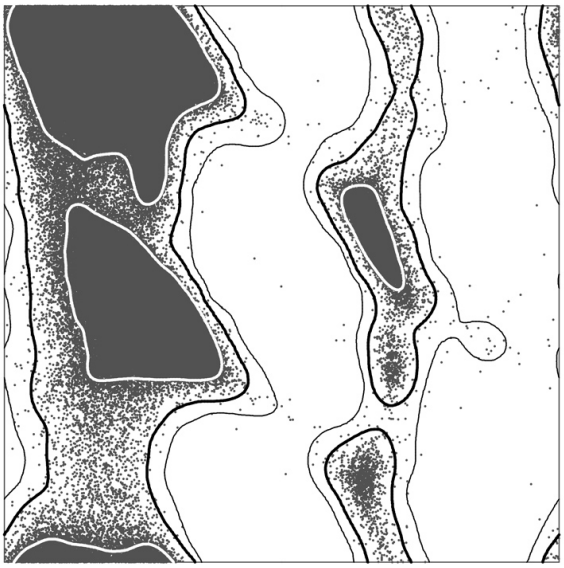
Crystallographic: R_{free}, electron density fit

15

general-case ϕ,ψ

(non-Gly, non-Pro, non-prePro)

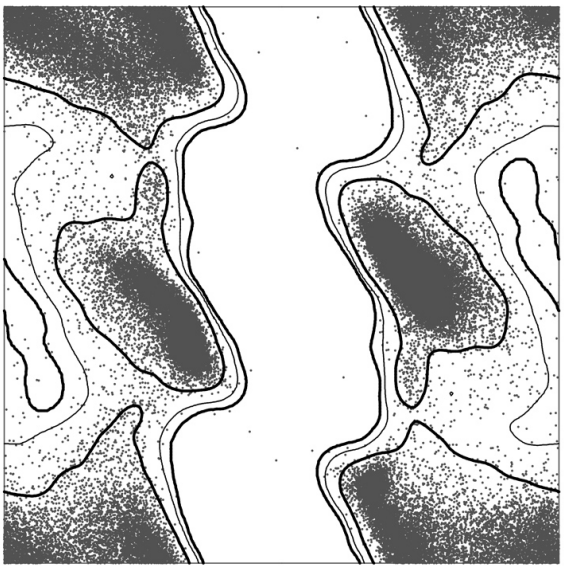
>750,000 aa
bb B <30



16

Gly ϕ,ψ

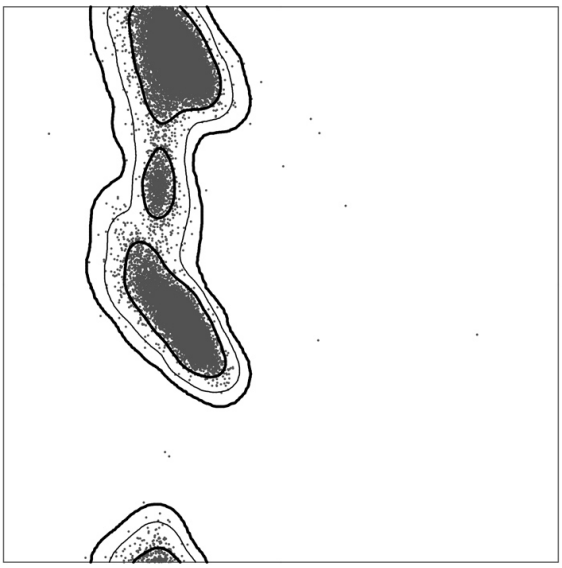
Top5200



17

Pro ϕ,ψ

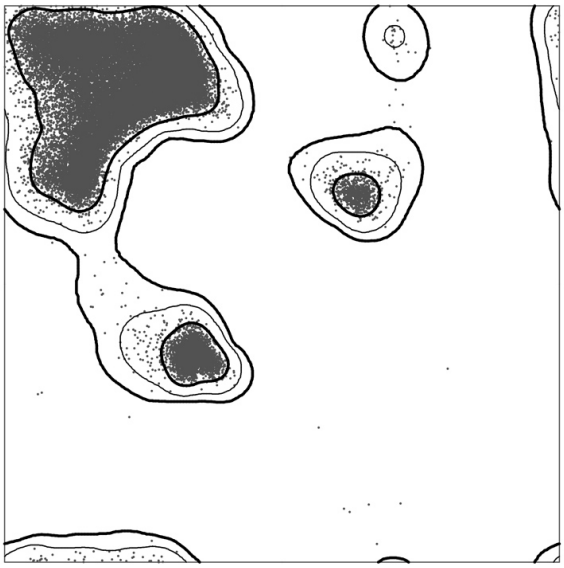
Top5200



18

prePro ϕ,ψ

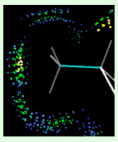
Top5200



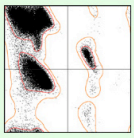
19

MolProbity Structure Improvement Criteria

All-atom contacts, clashscore

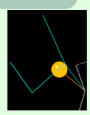


Ramachandran criteria

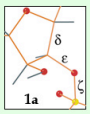


Sidechain rotamers

Geometry



RNA bb



Crystallographic : R_{free}, electron density fit

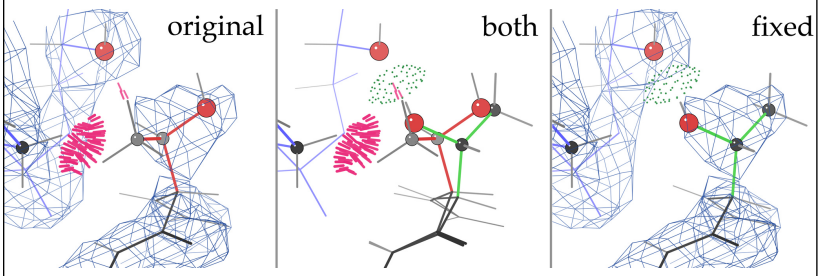
20

Rotamers are tight and distinct

21

Example Rotamer Correction

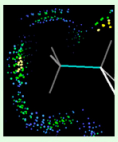
Thr O 3 from 1YHQ



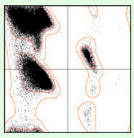
22

MolProbity Structure Improvement Criteria

All-atom contacts, clashscore

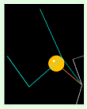


Ramachandran criteria

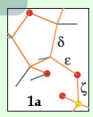


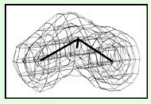
Sidechain rotamers

Geometry



RNA bb

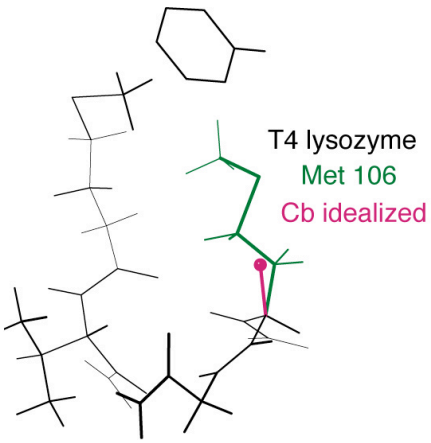




Crystallographic: R_{free}, electron density fit

23

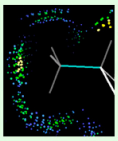
Cβ Deviations



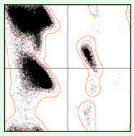
24

MolProbity Structure Improvement Criteria

All-atom contacts, clashscore

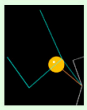


Ramachandran criteria

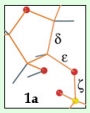


Sidechain rotamers

Geometry

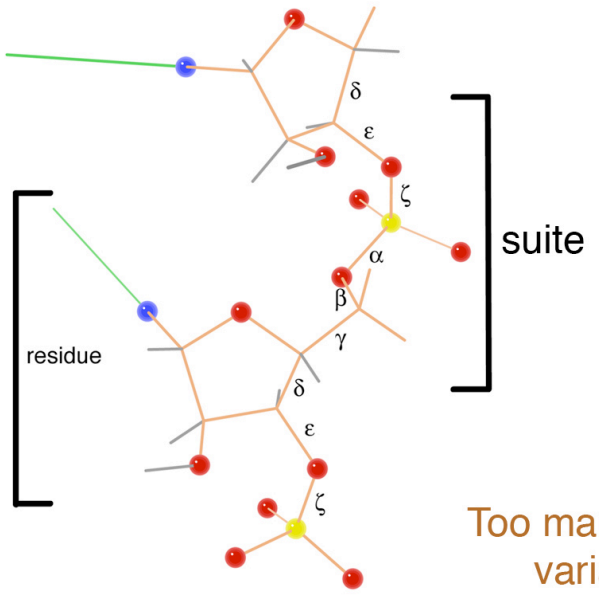


RNA bb



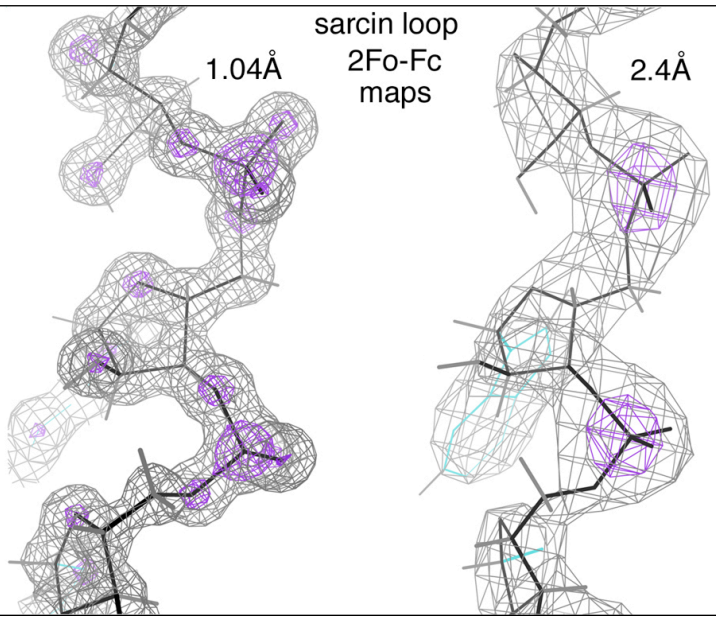
Crystallographic : R_{free}, electron density fit

25

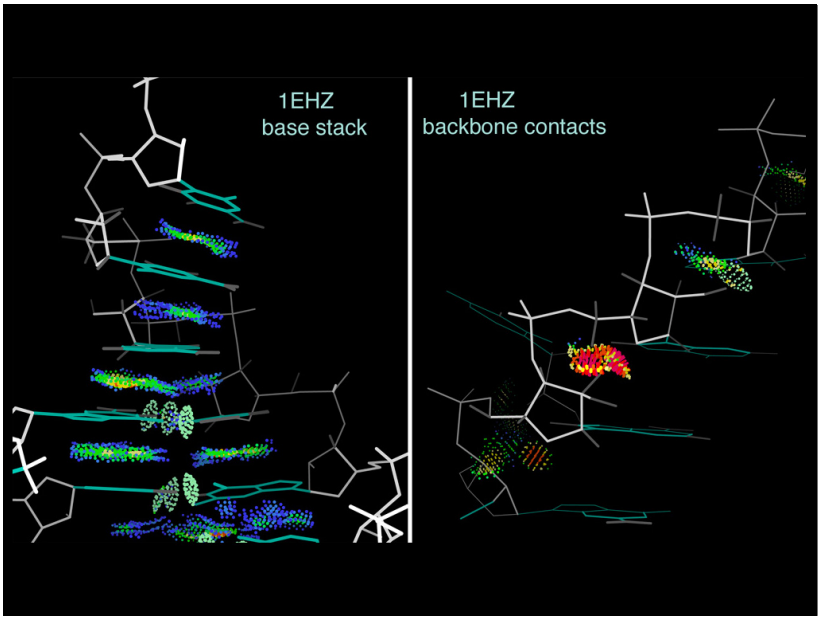


Too many variables!

26

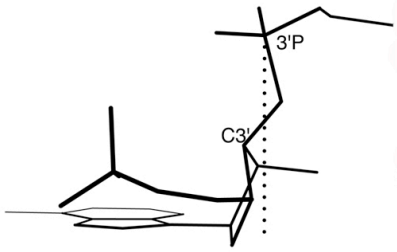


27

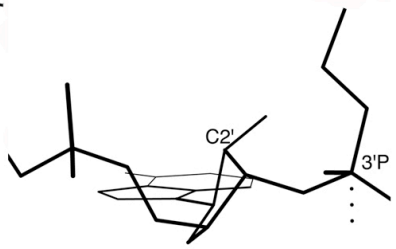


28

C3’ vs. C2’ Pucker

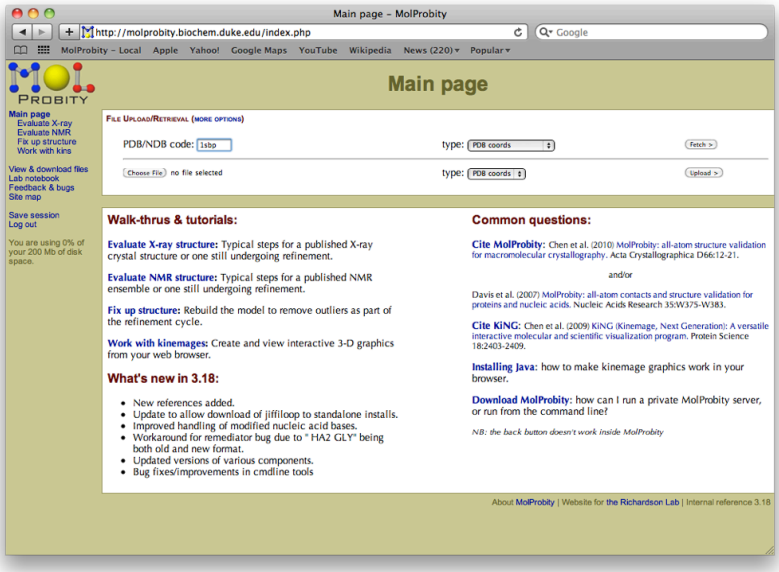


C3' pucker, long 3'P ⊥

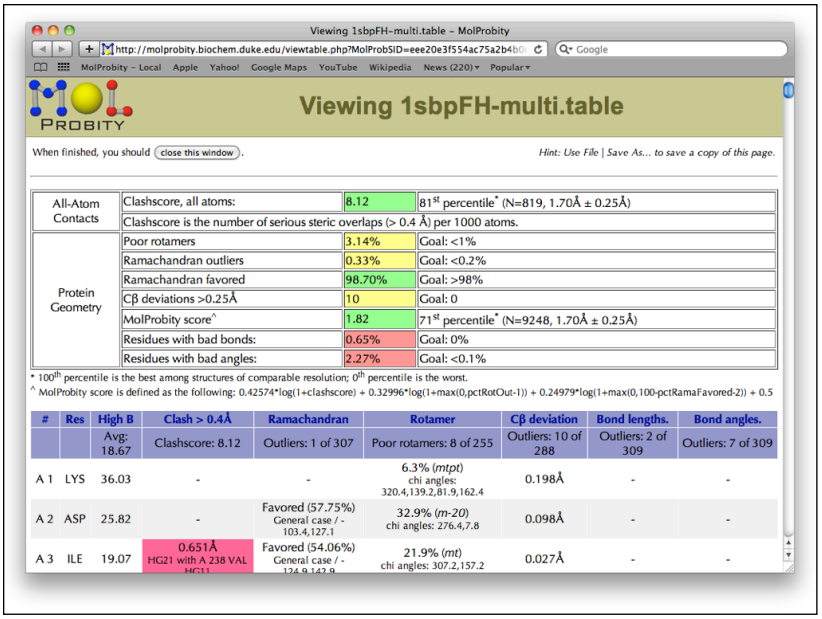


C2' pucker, short 3'P ⊥

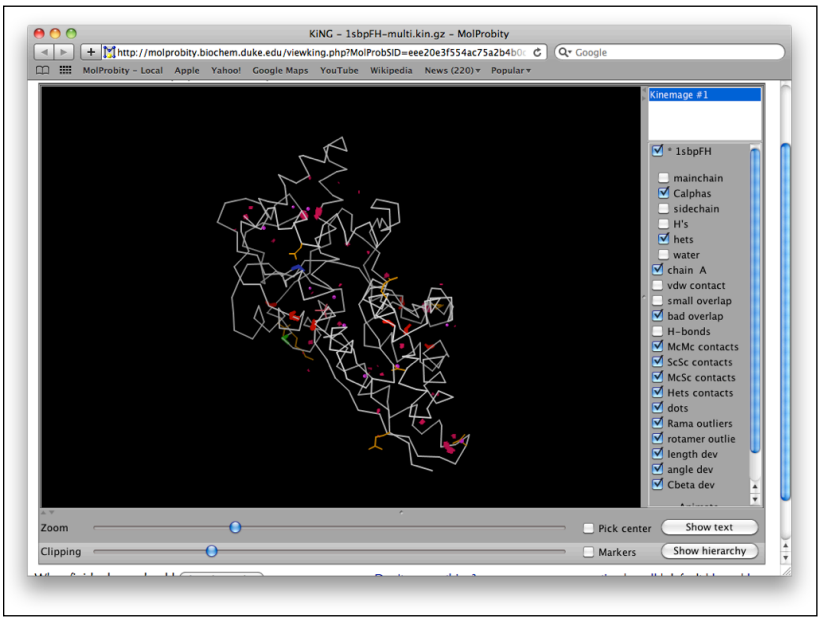
29



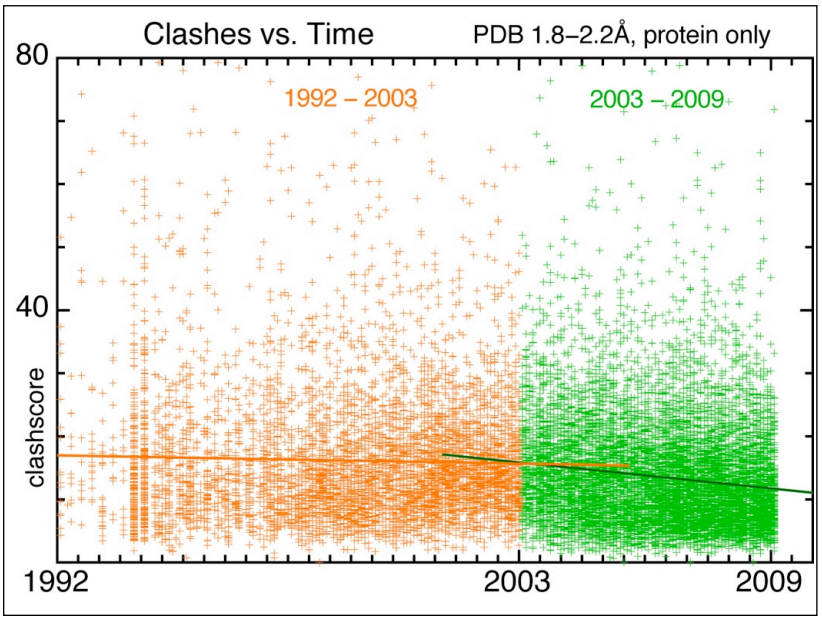
30


Viewing 1sbpFH-multi.table - MolProbity
http://molprobity.biochem.duke.edu/viewtable.php?MolProbSID=eee20e3f554ac75a2b4b0
MolProbity - Local Apple Yahoo! Google Maps YouTube Wikipedia News (220) Popular
Google
PROBITY
Viewing 1sbpFH-multi.table
When finished, you should close this window.
Hint: Use File | Save As... to save a copy of this page.
All-Atom Contacts
Clashscore, all atoms: 8.12 81st percentile* (N=819, 1.70Å ± 0.25Å)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.
Protein Geometry
Poor rotamers 3.14% Goal: <1%
Ramachandran outliers 0.33% Goal: <0.2%
Ramachandran favored 98.70% Goal: >98%
Cβ deviations >0.25Å 10 Goal: 0
MolProbity score^ 1.82 71st percentile* (N=9248, 1.70Å ± 0.25Å)
Residues with bad bonds: 0.65% Goal: 0%
Residues with bad angles: 2.27% 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
# Res High B Clash > 0.4Å Ramachandran Rotamer Cβ deviation Bond lengths. Bond angles.
Avg: 18.67 Clashscore: 8.12 Outliers: 1 of 307 Poor rotamers: 8 of 255 Outliers: 10 of 288 Outliers: 2 of 309 Outliers: 7 of 309
A 1 LYS 36.03 - - 6.3% (mtpt) chi angles: 320.4,139.2,81.9,162.4 0.198Å - -
A 2 ASP 25.82 - Favored (57.75%) General case / - 103.4,127.1 32.9% (m-20) chi angles: 276.4,7.8 0.098Å - -
A 3 ILE 19.07 0.651Å HG21 with A 238 VAL Favored (54.06%) General case / - 21.9% (mt) chi angles: 307.2,157.2 0.027Å - -

31


KiNG - 1sbpFH-multi.kin.gz - MolProbity
http://molprobity.biochem.duke.edu/viewking.php?MolProbSID=eee20e3f554ac75a2b4b0
MolProbity - Local Apple Yahoo! Google Maps YouTube Wikipedia News (220) Popular
Google
Kinemage #1
* 1sbpFH
mainchain
Calphas
sidechain
H's
hets
water
chain A
vdw contact
small overlap
bad overlap
H-bonds
McMc contacts
ScSc contacts
McSc contacts
Hets contacts
dots
Rama outliers
rotamer outlie
length dev
angle dev
Cbeta dev
Zoom
Clipping
Pick center
Markers
Show text
Show hierarchy

32


Clashes vs. Time
PDB 1.8–2.2Å, protein only
80
40
clashscore
1992 – 2003
2003 – 2009
1992
2003
2009

33

II. Validation in PHENIX

command line validation tools in PHENIX

$ phenix.rotalyze 1sbp.pdb

```
residue:score%:chi1:chi2:chi3:chi4:rotamer
A   1 LYS:6.3:320.4:139.2:81.9:162.4:mtpt
A   2 ASP:32.9:276.4:7.8:::m-20
A   3 ILE:21.9:307.2:157.2:::mt
A   4 GLN:0.0:222.6:108.3:159.9::OUTLIER
...
```

$ phenix.ramalyze 1sbp.pdb

```
residue:score%:phi:psi:evaluation:type
A   2 ASP:57.75:-103.37:127.06:Favored:General
A   3 ILE:54.06:-124.92:142.95:Favored:General
A   4 GLN:57.01:-121.44:122.25:Favored:General
A   5 LEU:31.42:-116.62:153.66:Favored:General
...
```

$ phenix.cbetadev 1sbp.pdb

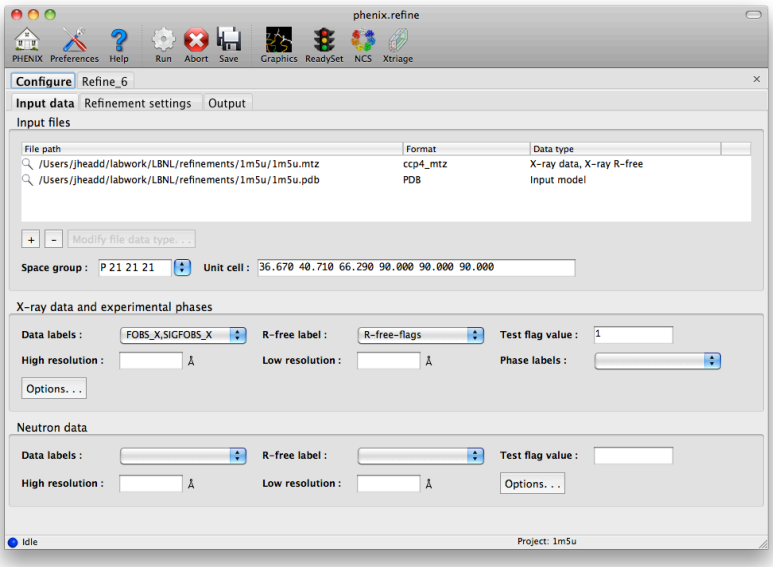
```
pdb:alt:res:chainID:resnum:dev:dihedralNABB:Occ:ALT:
  1sbp : :lys: A:   1  :  0.198: 112.84:   1.00: :
  1sbp : :asp: A:   2  :  0.098: -84.34:   1.00: :
  1sbp : :ile: A:   3  :  0.027:   7.64:   1.00: :
  1sbp : :gln: A:   4  :  0.204:  94.49:   1.00: :
  ...
```

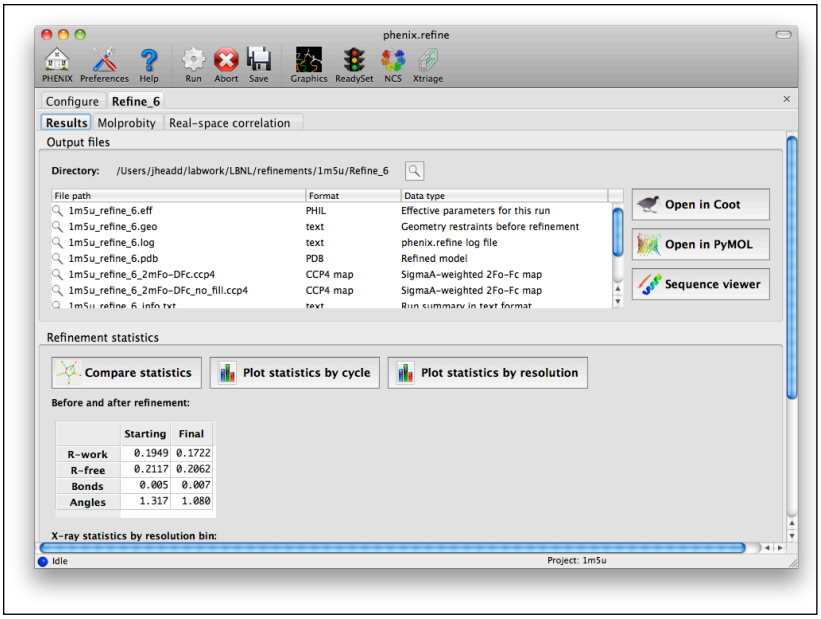
$ phenix.clashscore 1sbp.pdb

```
Bad Clashes >= 0.4 Angstrom:
  A  29 LYS  HE2  A  35 ASN HD21 :-0.822
  A 297 HIS  ND1  A 303 THR HG21 :-0.918
  A  86 LEU  HA   A  87 PRO  HD3 :-0.408
  A 160 LYS  O    A 164 LYS  HG2 :-0.509
  ...
```

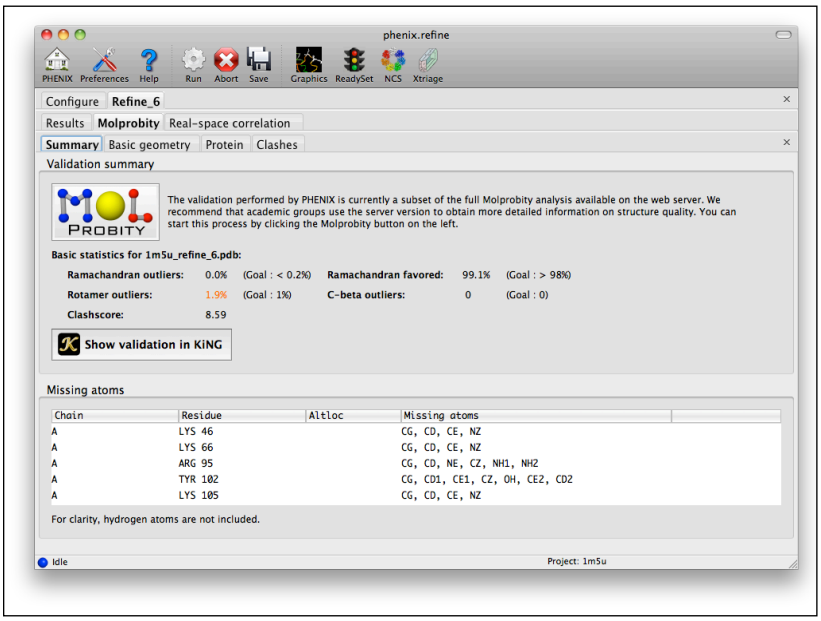
$ phenix.rna_validate 1u8d.pdb

```
RNA Validation
-----------------------------------------------
Pucker Outliers:
#residue:delta_angle:is_d_outlier:epsilon_angle:is_e_outlier
   G A  15 :74.054:no:103.458:yes
   U A  48 :119.081:yes:283.965:no
   U A  63 :90.366:yes:74.220:yes

-----------------------------------------------
Bond Length Outliers:
#residue:atom_1:atom_2:num_sigmas
   G A  15 : P  : O5':-4.003

-----------------------------------------------
Angle Outliers:
#residue:atom_1:atom_2:atom_3:num_sigmas
   U A  49 : C2': C1': N1 :-4.062
   C A  39 : C5': C4': C3':-7.941
   C A  39 : OP1: P  : O5':-4.214
   C A  39 : OP2: P  : O5':6.661
   C A  54 : C5': C4': C3':-5.912
   C A  54 : OP2: P  : O5':4.545
   U A  63 : O3': C3': C2':-6.602
   A A  40 : OP2: P  : O5':4.240

-----------------------------------------------
Suite Outliers:
#suiteID:triaged_angle
   G A  16 :epsilon-1
   U A  48 :delta
   U A  49 :delta-1
   A A  64 :epsilon-1
```

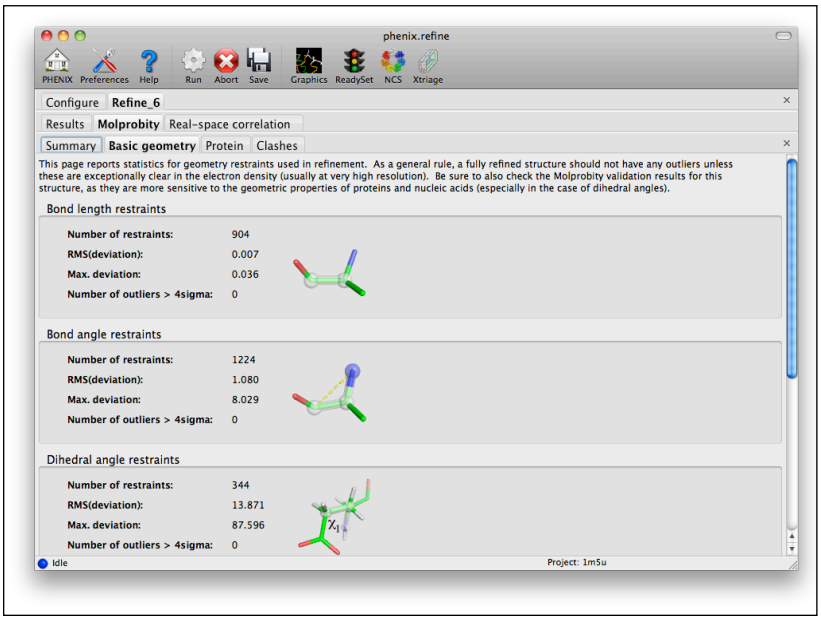




phenix.refine
PHENIX Preferences Help Run Abort Save Graphics ReadySet NCS Xtriage
Configure Refine_6
Results Molprobity Real-space correlation
Output files
Directory: /Users/jheadd/labwork/LBNL/refinements/1m5u/Refine_6
File path Format Data type
1m5u_refine_6.eff PHIL Effective parameters for this run
1m5u_refine_6.geo text Geometry restraints before refinement
1m5u_refine_6.log text phenix.refine log file
1m5u_refine_6.pdb PDB Refined model
1m5u_refine_6_2mFo-DFc.ccp4 CCP4 map SigmaA-weighted 2Fo-Fc map
1m5u_refine_6_2mFo-DFc_no_fill.ccp4 CCP4 map SigmaA-weighted 2Fo-Fc map
Open in Coot
Open in PyMOL
Sequence viewer
Refinement statistics
Compare statistics
Plot statistics by cycle
Plot statistics by resolution
Before and after refinement:
Starting Final
R-work 0.1949 0.1722
R-free 0.2117 0.2062
Bonds 0.005 0.007
Angles 1.317 1.080
X-ray statistics by resolution bin:
Idle
Project: 1m5u

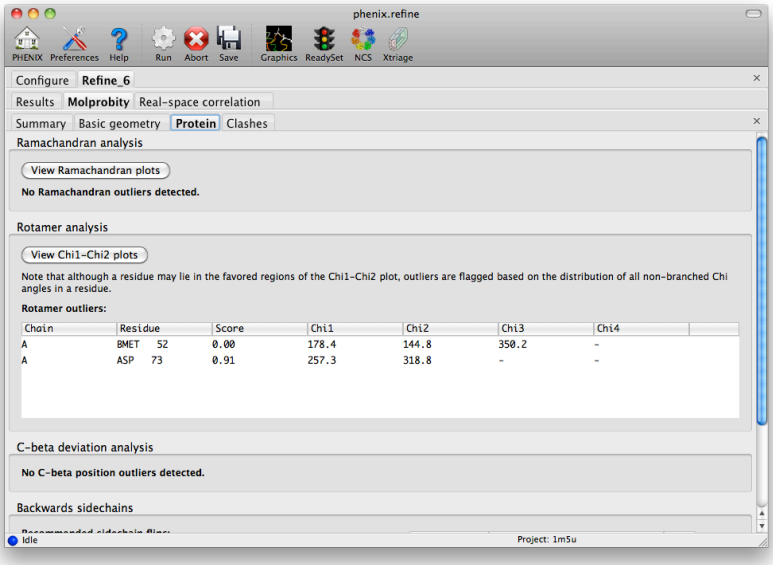
37


phenix.refine
PHENIX Preferences Help Run Abort Save Graphics ReadySet NCS Xtriage
Configure Refine_6
Results Molprobity Real-space correlation
Summary Basic geometry Protein Clashes
Validation summary
MOLPROBITY
The validation performed by PHENIX is currently a subset of the full Molprobity analysis available on the web server. We recommend that academic groups use the server version to obtain more detailed information on structure quality. You can start this process by clicking the Molprobity button on the left.
Basic statistics for 1m5u_refine_6.pdb:
Ramachandran outliers: 0.0% (Goal : < 0.2%) Ramachandran favored: 99.1% (Goal : > 98%)
Rotamer outliers: 1.9% (Goal : 1%) C-beta outliers: 0 (Goal : 0)
Clashscore: 8.59
Show validation in KiNG
Missing atoms
Chain Residue Altloc Missing atoms
A LYS 46 CG, CD, CE, NZ
A LYS 66 CG, CD, CE, NZ
A ARG 95 CG, CD, NE, CZ, NH1, NH2
A TYR 102 CG, CD1, CE1, CZ, OH, CE2, CD2
A LYS 105 CG, CD, CE, NZ
For clarity, hydrogen atoms are not included.
Idle
Project: 1m5u

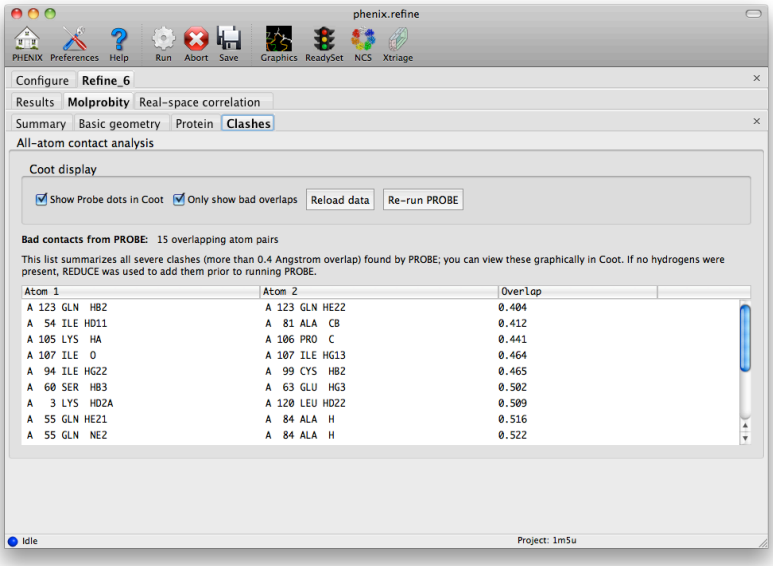
38


phenix.refine
PHENIX Preferences Help Run Abort Save Graphics ReadySet NCS Xtriage
Configure Refine_6
Results Molprobity Real-space correlation
Summary Basic geometry Protein Clashes
This page reports statistics for geometry restraints used in refinement. As a general rule, a fully refined structure should not have any outliers unless these are exceptionally clear in the electron density (usually at very high resolution). Be sure to also check the Molprobity validation results for this structure, as they are more sensitive to the geometric properties of proteins and nucleic acids (especially in the case of dihedral angles).
Bond length restraints
Number of restraints: 904
RMS(deviation): 0.007
Max. deviation: 0.036
Number of outliers > 4sigma: 0
Bond angle restraints
Number of restraints: 1224
RMS(deviation): 1.080
Max. deviation: 8.029
Number of outliers > 4sigma: 0
Dihedral angle restraints
Number of restraints: 344
RMS(deviation): 13.871
Max. deviation: 87.596
Number of outliers > 4sigma: 0
Idle
Project: 1m5u

39



40



41

Live example: 1SBP

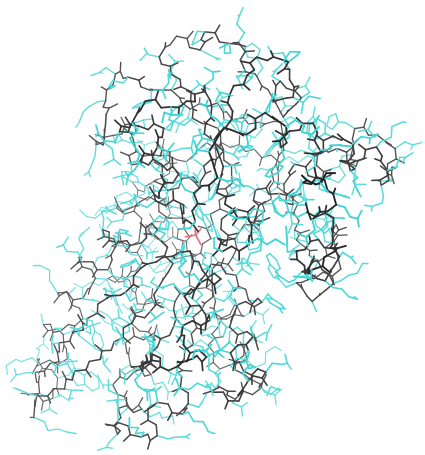
Sulfate-Binding Protein

1.7 Å

Deposited: 1993

R: 0.178
R-Free: N/A

310 Residues



42

Acknowledgements

LBNL:
Paul Adams
Ralf Grosse-Kunstleve
Pavel Afonine
Nigel Moriarty
Nat Echols
Nick Sauter

LANL:
Tom Terwilliger
Li-Wei Hung

Cambridge:
Randy Read
Airlie McCoy
Gabor Bunkoczi
Rob Oeffner

Duke:
Dave Richardson
Jane Richardson
Bryan Arendall
Vincent Chen
Jeremy Block
Gary Kapral
Dan Keedy
Christopher Williams
Swati Jain
Bradley Hintze
Lizbeth Videau

Collaborators:
Jack Snoeyink
Paul Emsley

Funding: NIH grant #GM063210 (and its ARRA supplement) and the Phenix Industrial Consortium