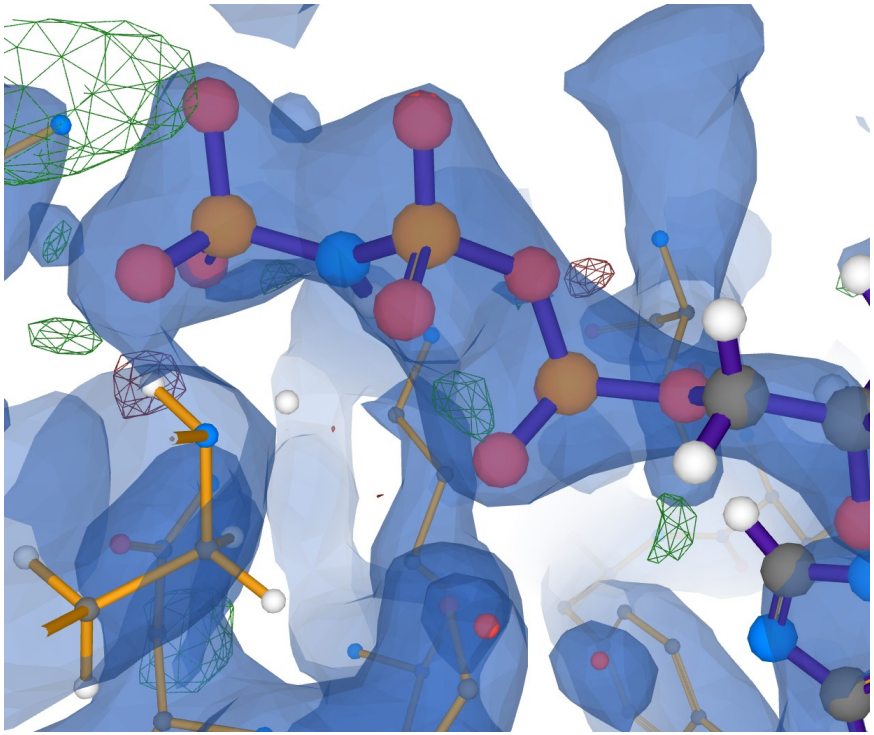
Validation - proteins

Electron density maps

- When looking at PDB structures Electrons Density (ED) maps are more/as important as the 3D atomic model!
- ED is a 3D map of where the scattering electron cloud is according to the measured X-ray data.
- 2Fo-Fc map indicates where electrons are (according to SF and model). Normally colour blue or grey.
- Fo-Fc difference map:

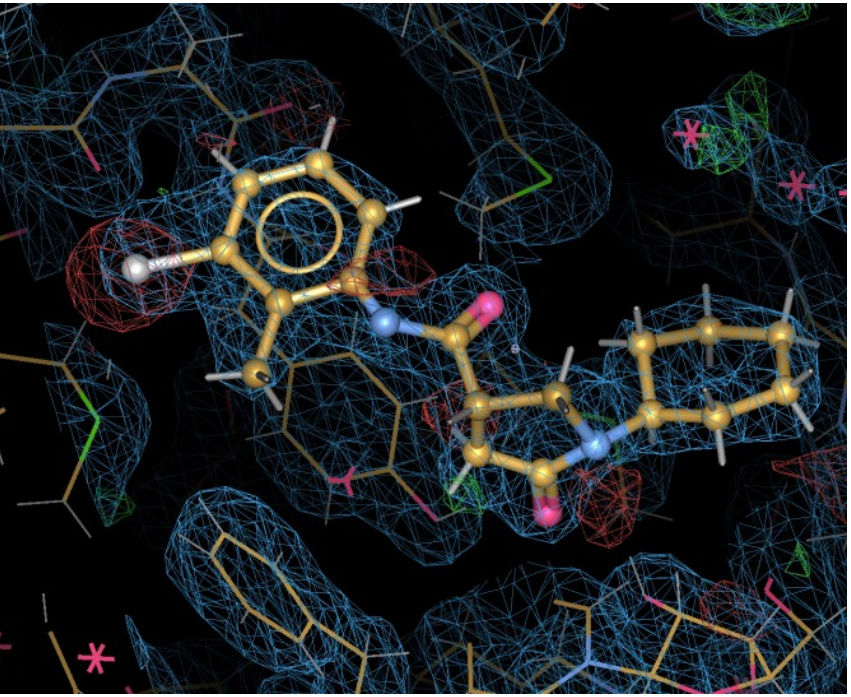
  green for positive difference: where the current model fails to place sufficient electrons

  Red for negative difference: where the current model places too many electrons

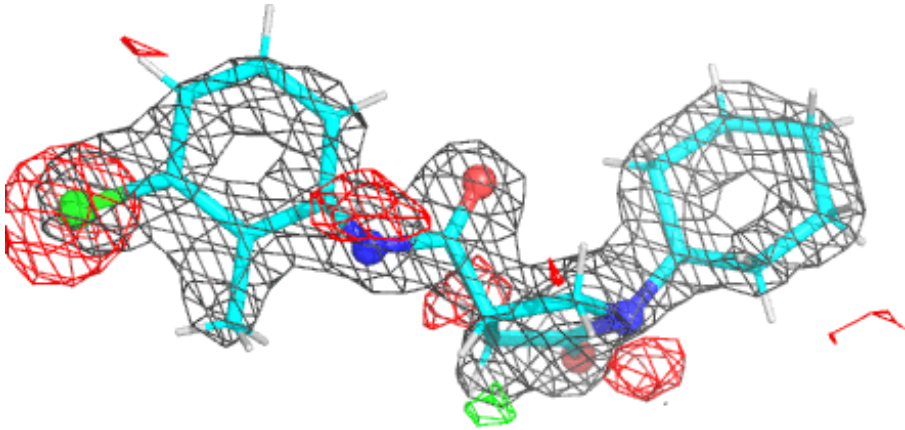


www.ebi.ac.uk/pdbe/entry/pdb/4z9l/bound/ANP

Electron density for a ligand with poor fit



2h7p.pdb: ED visualized in coot



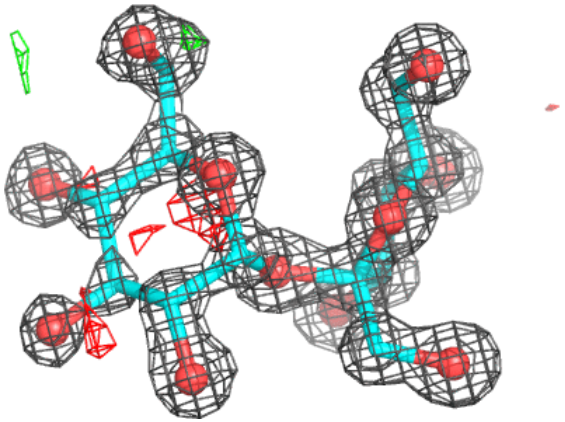
2h7p.pdb: ED around ligand, as visualized in buster-report

Notice difference density around ligand

2h7p has been obsoleted and replaced 4tzt with really superb ED and ligand fit

Data resolution affects electron density detail:

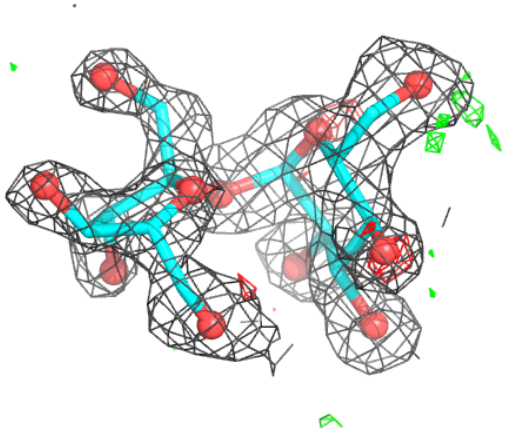
Well placed/refined sucrose ligand at different data resolutions:



1ylt 1.2Å resolution

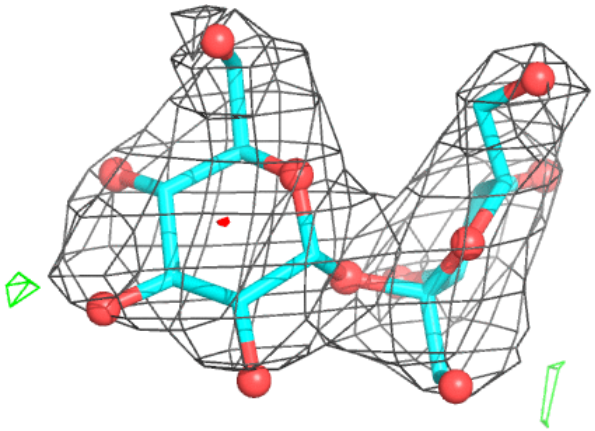
"atomic resolution"

2Fo-Fc ED shows atoms as Individual blobs. Need higher resolution for hydrogen atoms



2pwe 2.0Å resolution

Typical medium resolution for ligand studies. Can see ring pucker

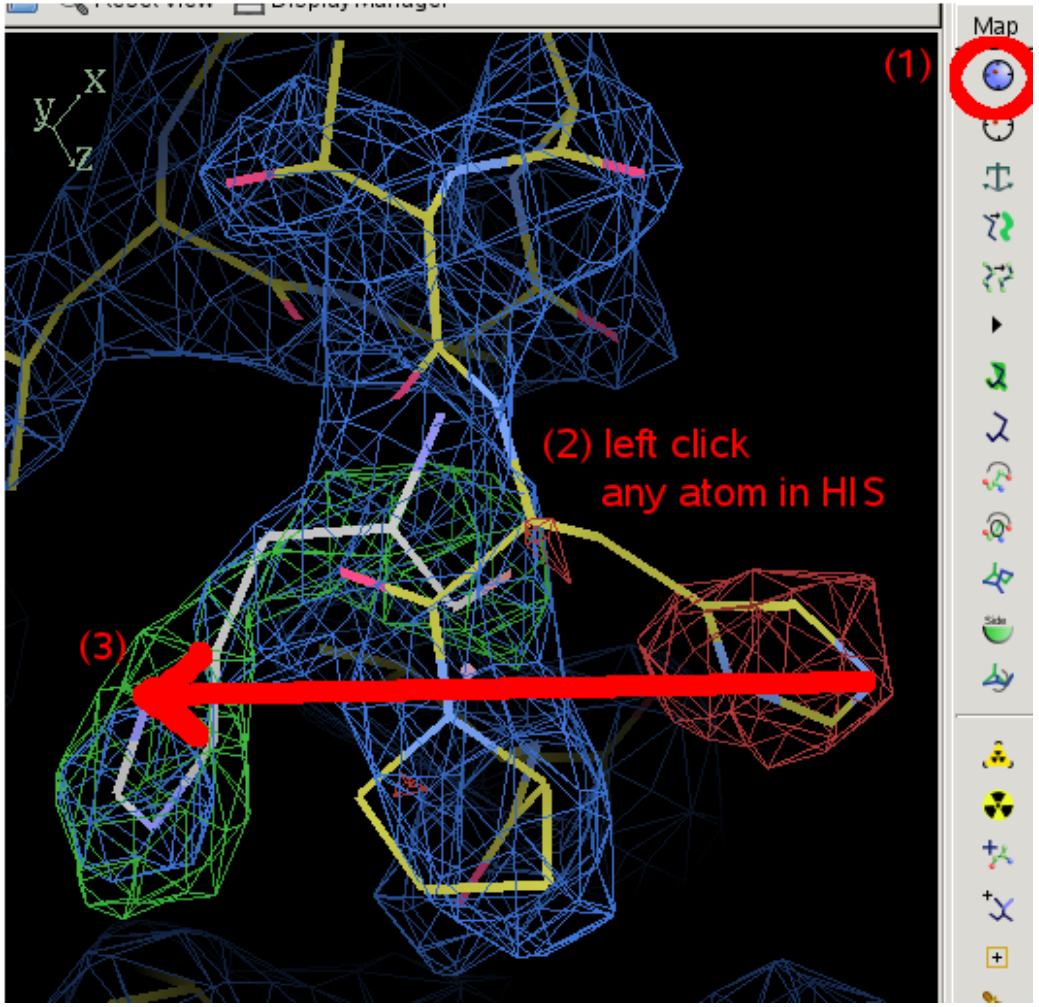


2qqv 3.0Å resolution

Low resolution. Ligand placement unambiguous but fine detail cannot be seen

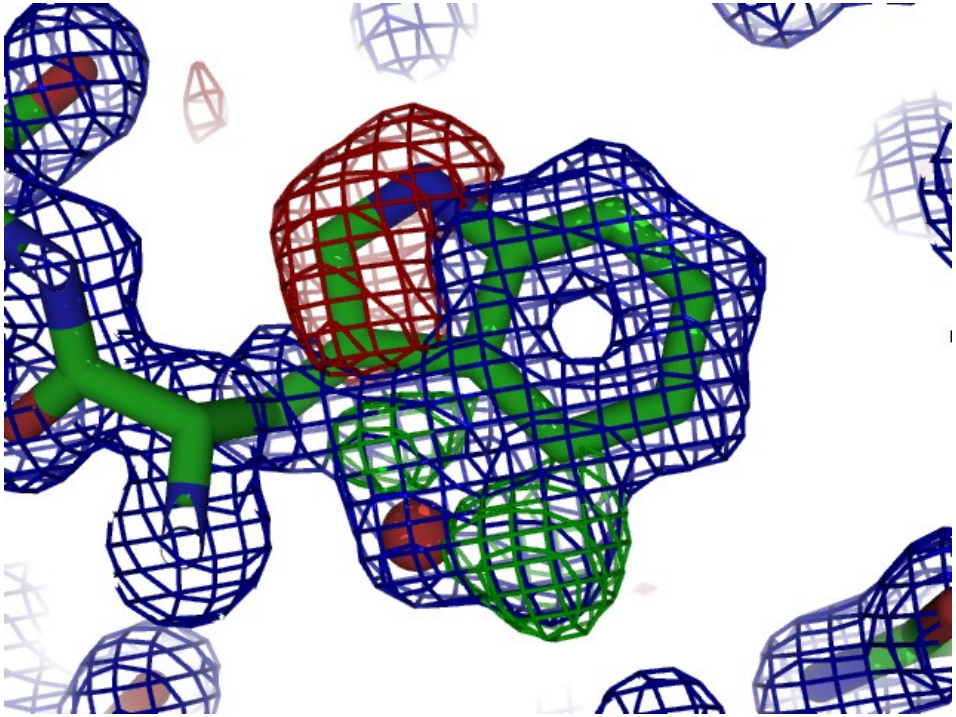
Model improvement

- Basically interpret electron density maps in real space to improve model
- Initially automated methods (warp/arpwarp) used
- But mostly manual corrections to the model are done using the Coot program
- Look at Fo-Fc difference map for both negative and positive features
- Build into 2Fo-Fc



Why do crystallographers make mistakes?

- Limitations to the data
  - Incomplete
  - Weak
  - Limited resolution
  - Space and time averaged

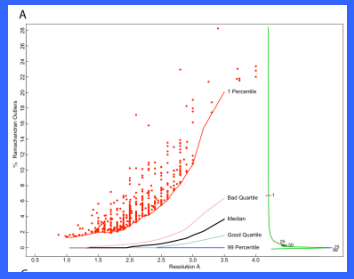


- The human factor
  - Subjectivity and bias involved in map interpretation and refinement (even at atomic resolution!)
  - Inexperienced people do the work, use of black boxes, …
  - Not everybody is a good chemist
  - Even experienced people make mistakes

wwPDB X-ray validation pipeline

External reference files (*e.g.*, Engh & Huber)

Distributions

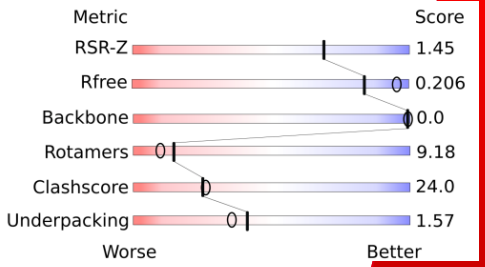


Validation XML file

Validation pipeline 1.0

MolProbity | Xtriage | EDS | Mogul | Percentiles | PDF maker

Deposited data (coordinates & reflections)



PDF report for depositor & referees - Statistics and plots for the entry, per chain, per residue, and list of unusual features

Gore *et al.*, *Acta Cryst.* **D68**, 478 (2012)

wwPDB validation reports

WORLDWIDE PDB PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Jan 31, 2016 – 06:45 PM GMT

PDB ID	:	1CBS
Title	:	CRYSTAL STRUCTURE OF CELLULAR RETINOIC-ACID-BINDING PROTEINS I AND II IN COMPLEX WITH ALL-TRANS-RETINOIC ACID AND A SYNTHETIC RETINOID
Authors	:	Kleywegt, G.J.; Bergfors, T.; Jones, T.A.
Deposited on	:	1994-09-28
Resolution	:	1.80 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

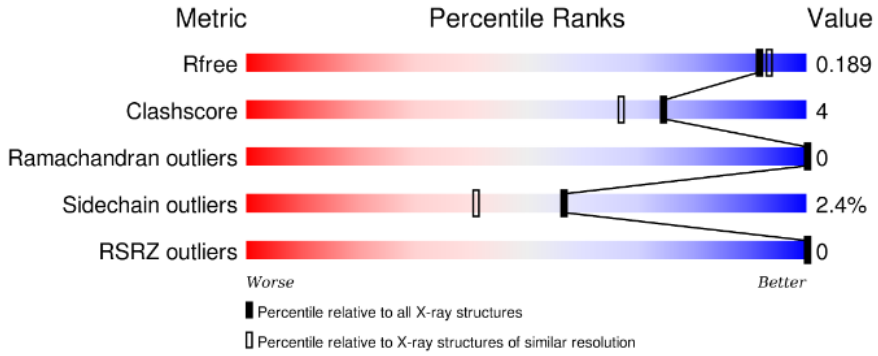
1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.80 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

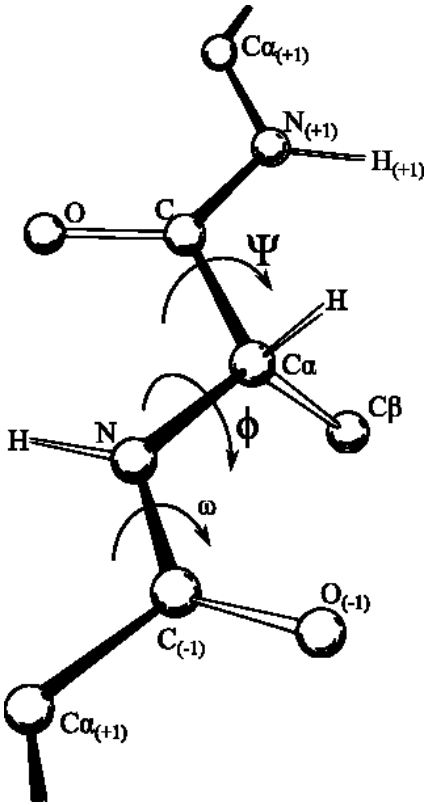
- Pipeline produces PDF report and XML output
- Slider graphic useful

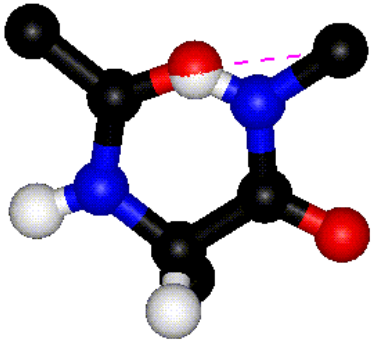


- Current PDF is “rather verbose”

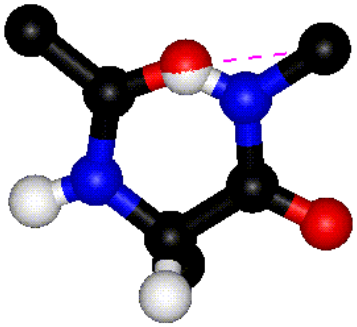
Ramachandran plot

- Look at main chain dihedral angles phi and psi
- Ramachandran et al. (1963) worked out only certain combinations of phi/psi cause clashes





Rotation around φ with ψ=0°



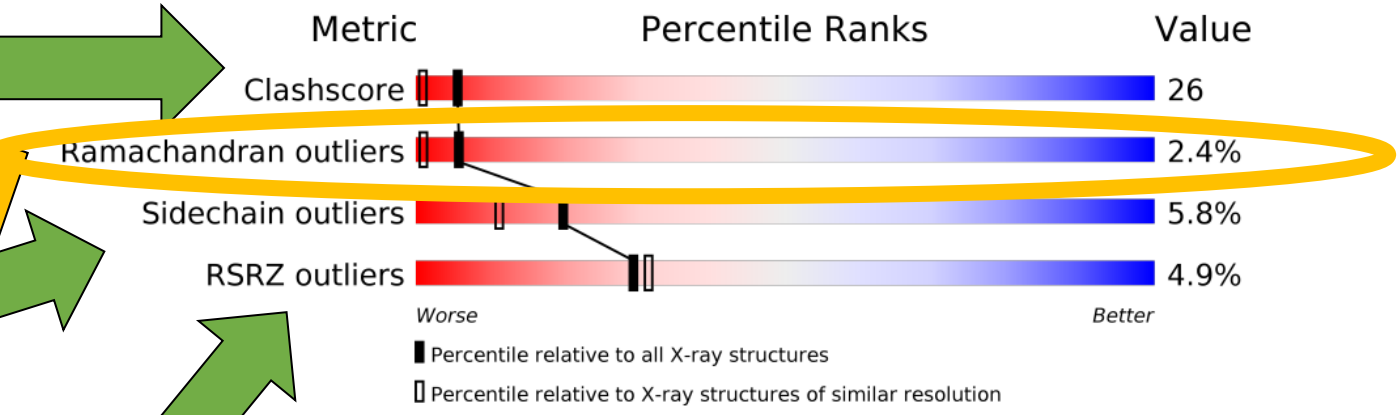
Rotation around ψ with φ=0°
(Images kindly provided by David Sanders, University of Saskatchewan.)

Summary ‘Sliders’ Validation information for users

Atoms bumping into each other

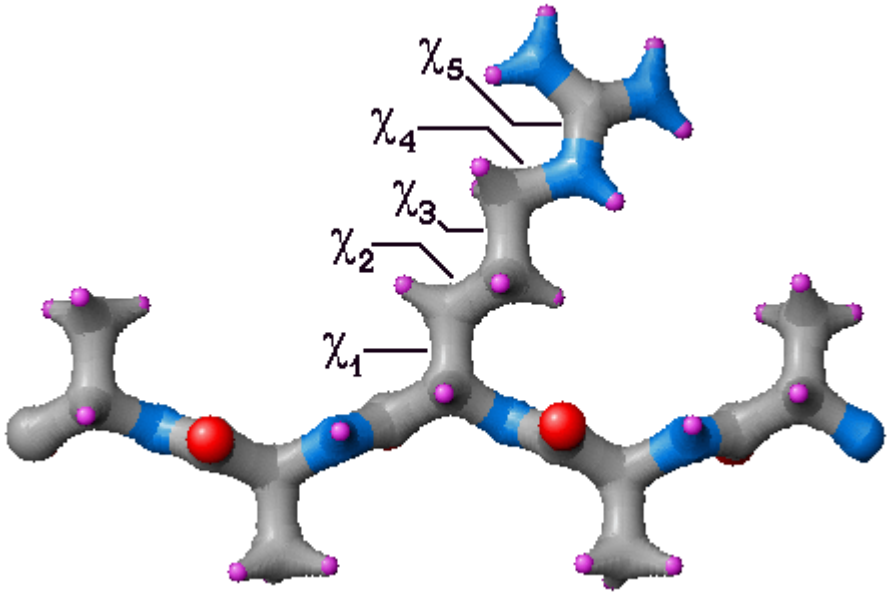
Surprising torsion angles

Atoms not in electron density



Sidechain outliers

- Just like the main chain phi and psi dihedral angles amino acid sides cha have chi angles with preferred and disallo regions



The 5 chi angles of an arginine side chain

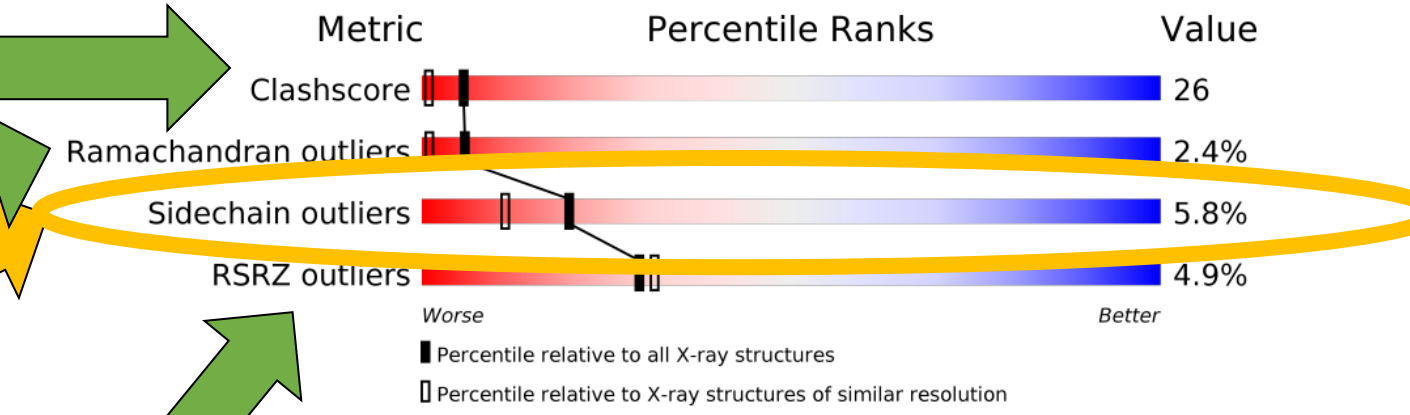
http://www.ccp14.ac.uk/ccp/web-mirrors/garlic/garlic-1.5/commands/dihedrals.html

Summary ‘Sliders’ Validation information for users

Atoms bumping into each other

Surprising torsion angle

Atoms not in electron density



MolProbity – clash score

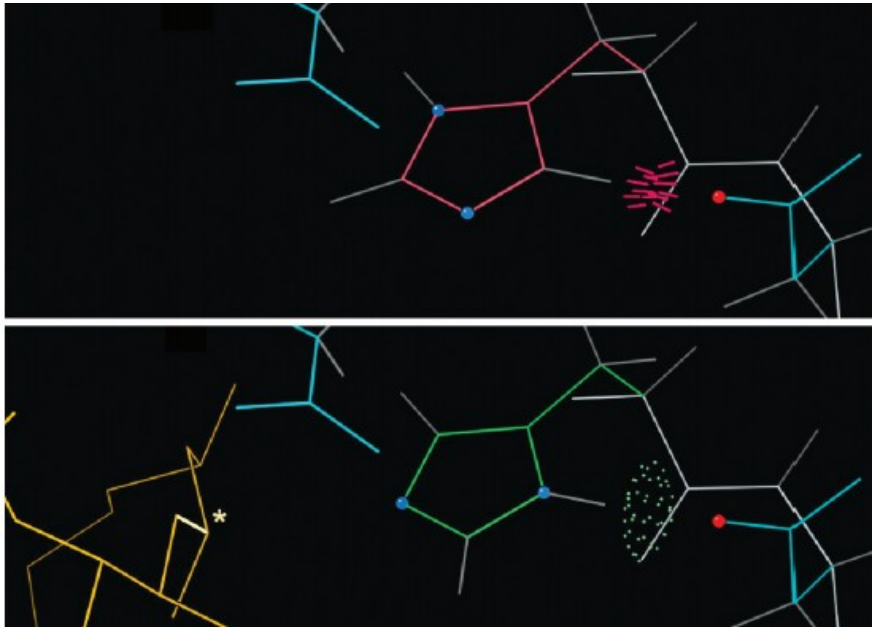
- Idea is to look for bad non-bonded contacts after hydrogen atoms have been added to the model
- Very powerful method
- Suggests NQH flips
- Included in wwPDB validation reports
- Or Use from:
  - Molprobity web site
  - Or within coot

Nucleic Acids Research, 2007, Vol. 35, Web Server issue **W375–W383**
doi:10.1093/nar/gkm216

MolProbity: all-atom contacts and structure validation for proteins and nucleic acids

Ian W. Davis[1], Andrew Leaver-Fay[2], Vincent B. Chen[1], Jeremy N. Block[1], Gary J. Kapral[1], Xueyi Wang[2], Laura W. Murray[1], W. Bryan Arendall III[1], Jack Snoeyink[2], Jane S. Richardson[1] and David C. Richardson[1,*]

[1]Department of Biochemistry, Duke University, Durham, NC, USA and [2]Department of Computer Science, UNC Chapel Hill, Chapel Hill, NC, USA

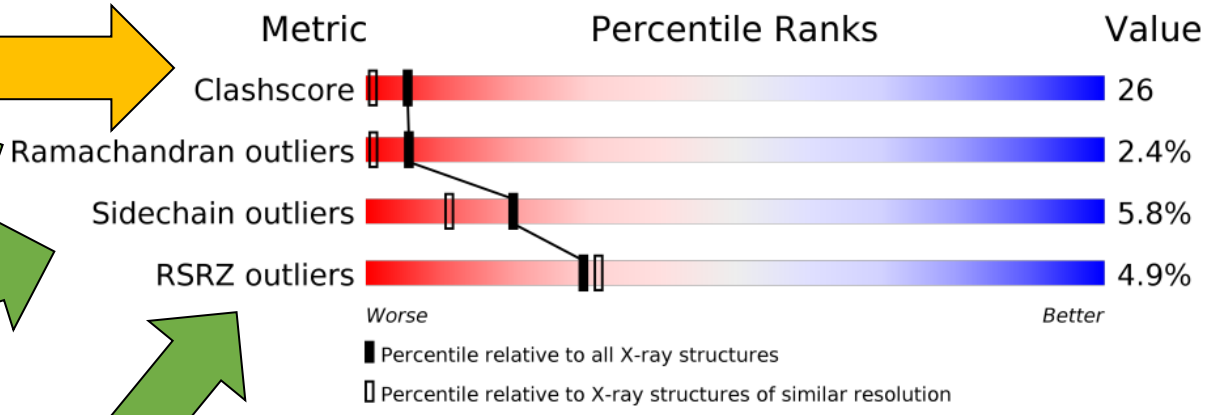


Validation information for users
Summary ‘Sliders’

Atoms bumping into each other

Surprising torsion angles

Atoms not in electron density

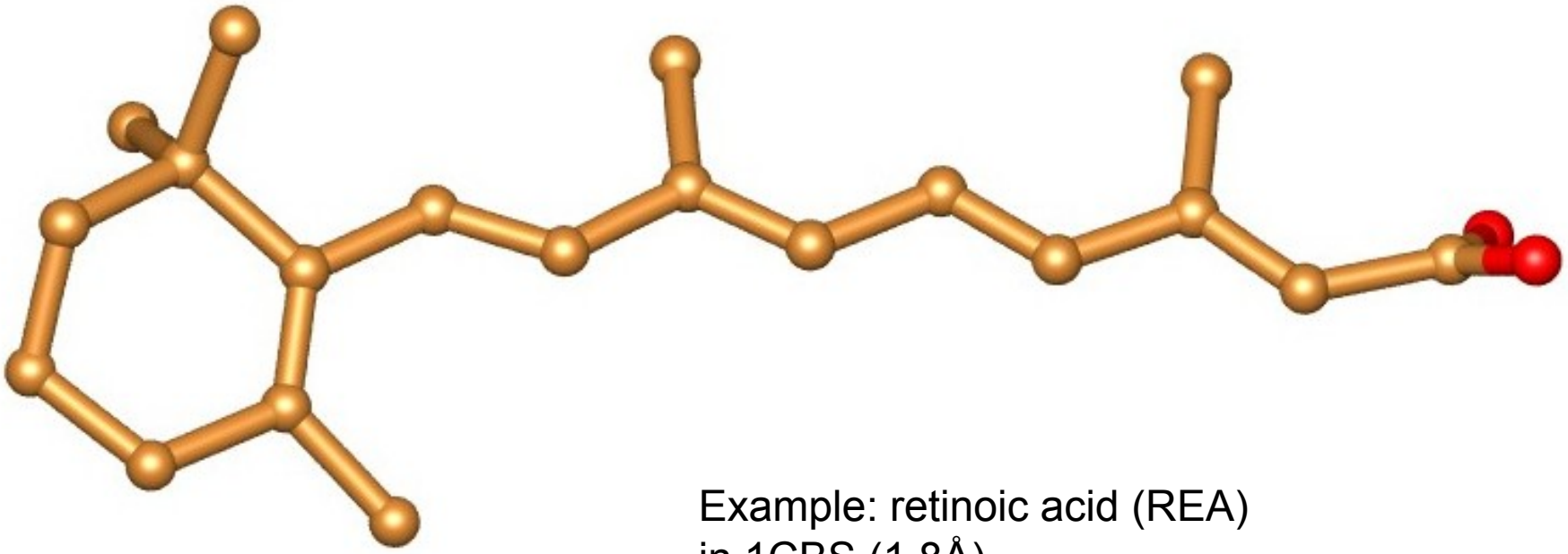


Real-space fit

- Quantitative, real-space measure of how well a residue fits its local density (Jones *et al.*, 1991)
- Express as R-value (RSR) or correlation coefficient (RSCC)
- $RSR = \Sigma \mid \rho_{obs} - \rho_{calc} \mid / \Sigma \mid \rho_{obs} + \rho_{calc} \mid$
- Sums extend over all grid points inside a mask around the residue

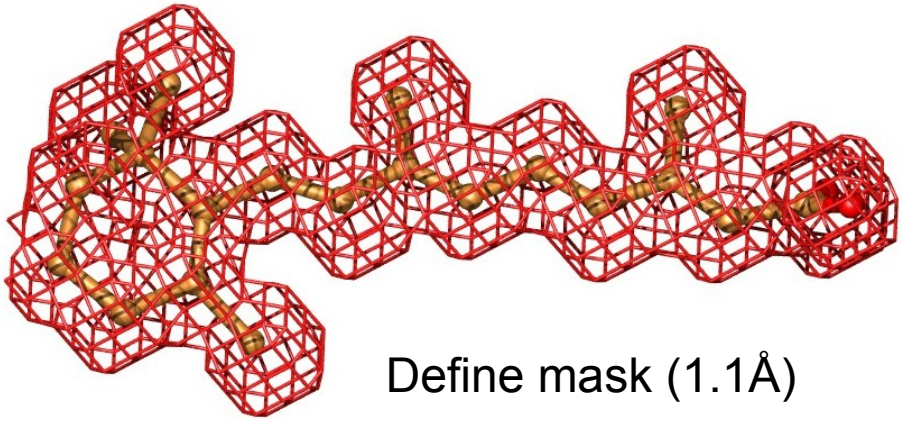
RSR - real-space R-value

- $RSR = \Sigma |\rho_{obs} - \rho_{calc}| / \Sigma |\rho_{obs} + \rho_{calc}|$

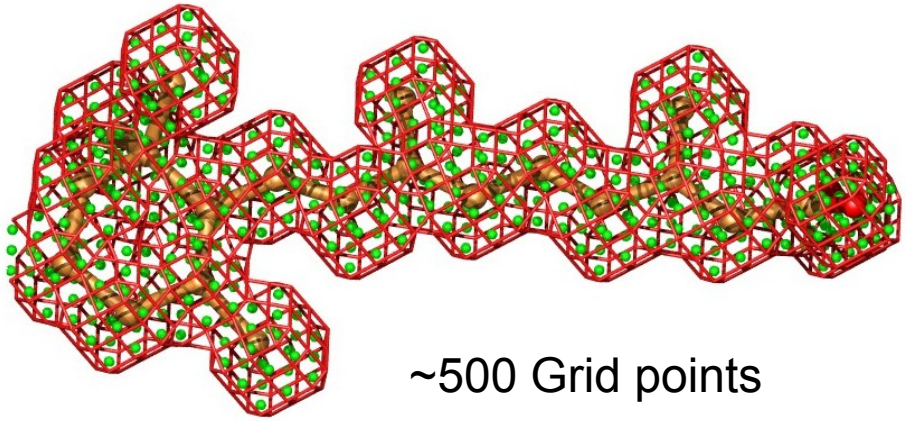


Example: retinoic acid (REA)
in 1CBS (1.8Å)

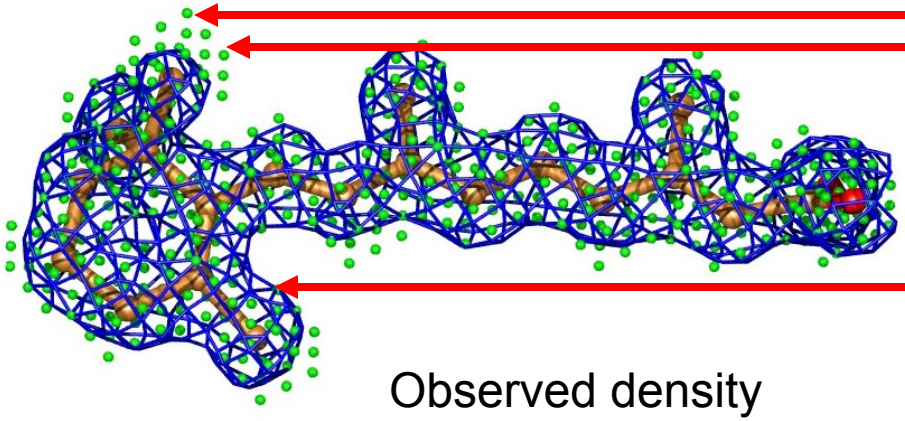
RSR - real-space R-value



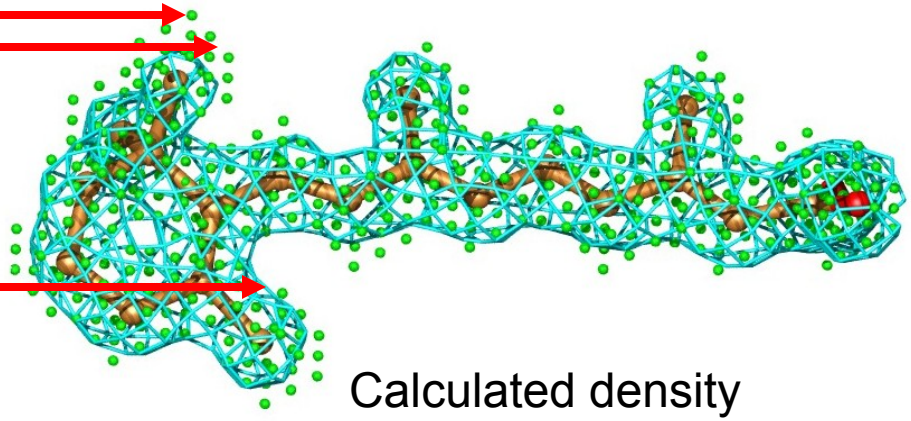
Define mask (1.1Å)



~500 Grid points



Observed density



Calculated density

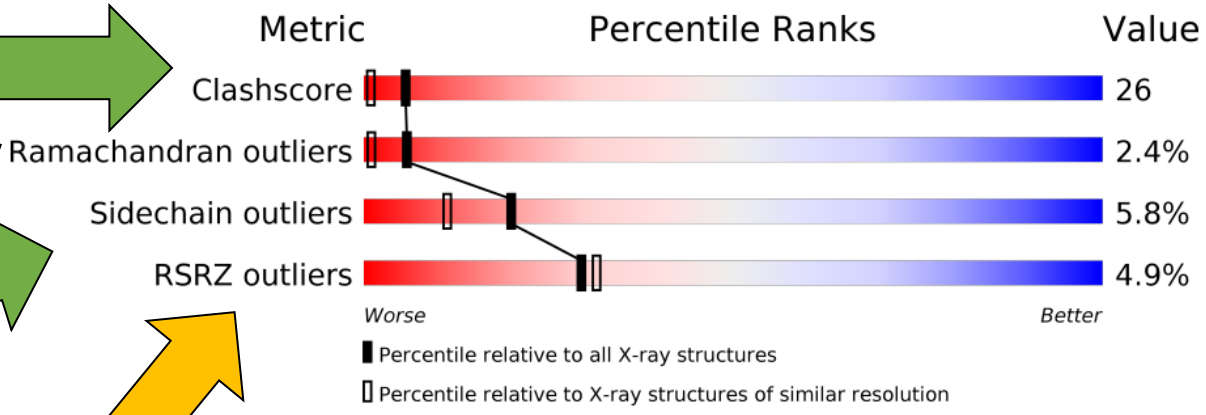
$$RSR = \Sigma \, |\rho_{obs} - \rho_{calc}| \, / \, \Sigma \, |\rho_{obs} + \rho_{calc}|$$

RSRZ is reported in **Summary 'Sliders'**

Atoms bumping into each other

Surprising torsion angles

Atoms not in electron density



PDBe simplification of validation sliders

(2) < 1 > Entry 1 to 2 of 2 Quality (desc)

1det RIBONUCLEASE T1 CARBOXYMETHYLATED AT GLU 58 IN COMPLEX WITH 2'GMP

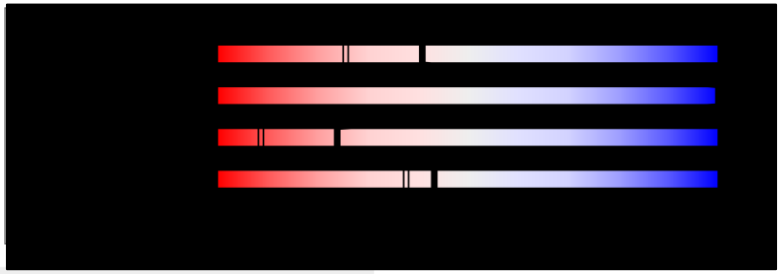
(2) Ishikawa K, Suzuki E, Tanokura M, Takahashi K
Biochemistry (1996) [PMID: 8679590]

Source organism: *Aspergillus oryzae*
Assembly composition: protein only structure

Interacting compounds: 2GP NA

(1) Add to basket Download files
(1)

X-ray diffraction
1.8Å resolution
Released: 11 Jul 1996
Model geom...
Fit model/da...



3syu Re-refined coordinates for pdb entry 1det - ribonuclease T1 carboxymethylated at GLU 58 in complex with 2'GMP

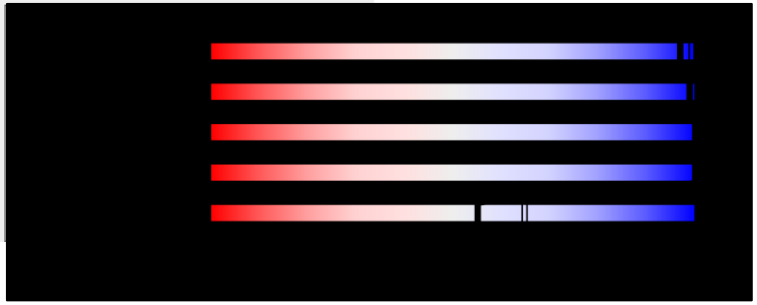
(2) Smart OS, Womack TO, Bricogne G
Acta Crystallogr. D Biol. Crystallogr. (2012) [PMID: 22505257]

Source organism: *Aspergillus oryzae RIB40*
Assembly composition: protein only structure

Interacting compounds: NA 2GP

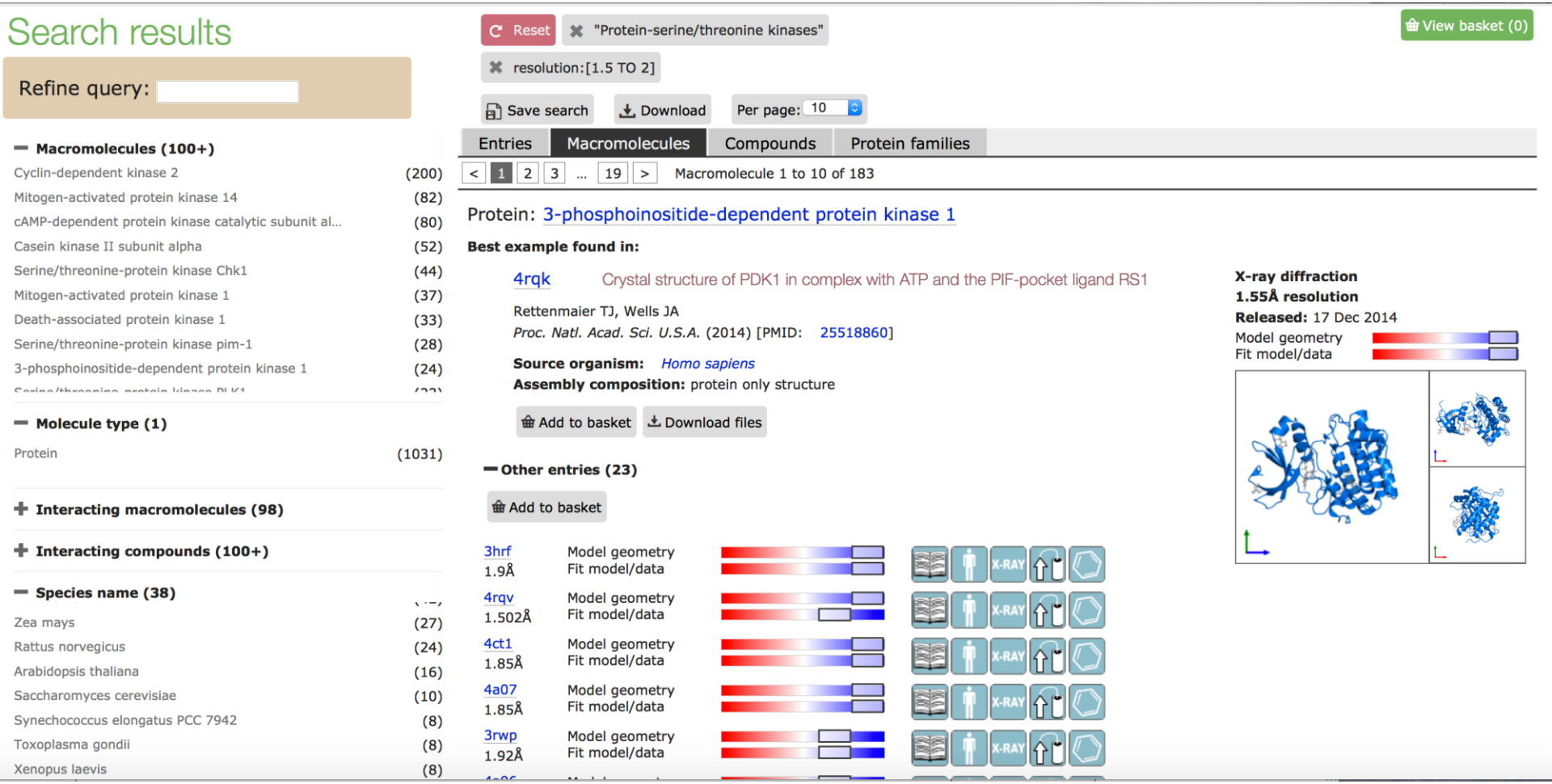
Add to basket Download files

X-ray diffraction
1.95Å resolution
Released: 28 Mar 2012
Model geom...
Fit model/da...

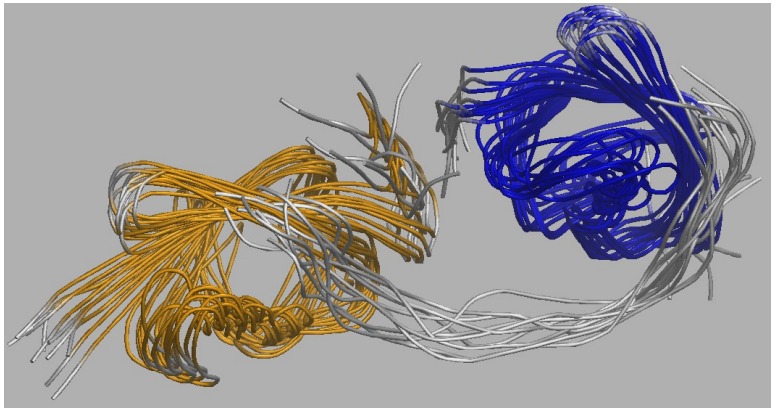


PDBe EMBL-EBI

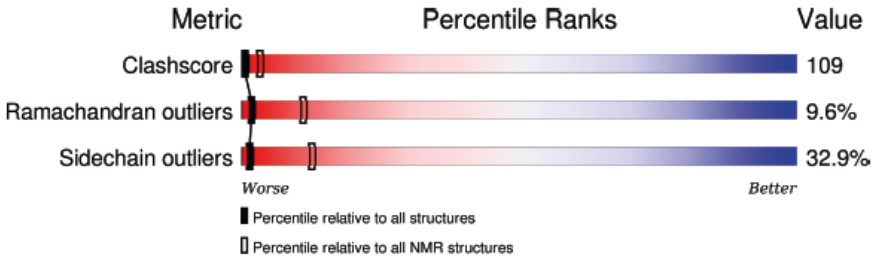
“Best molecule” – integration of validation information in PDBe query system



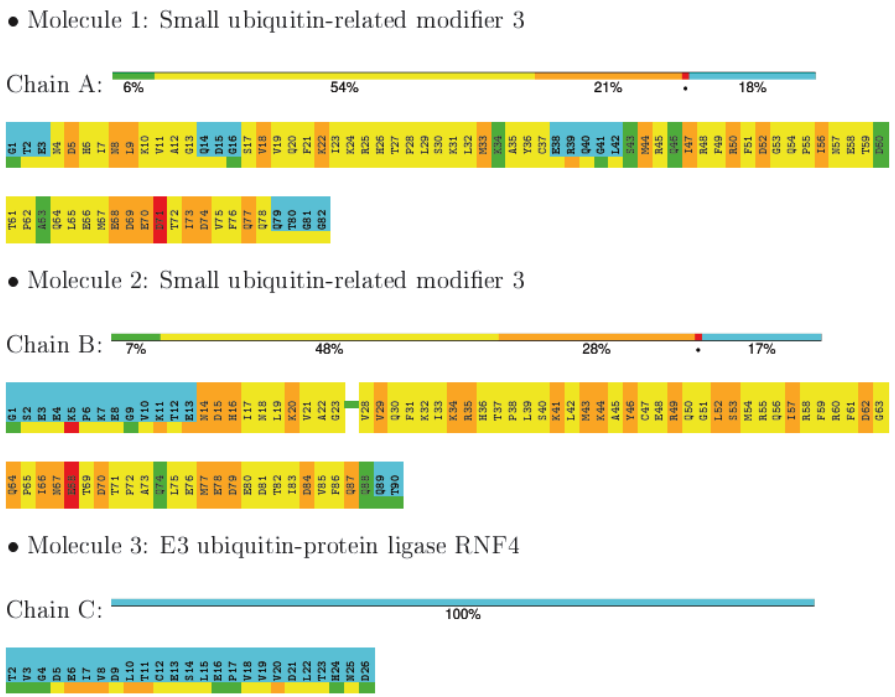
NMR validation



- NMR VTF recommendations published
- Global quality scores reported for“well-defined residues” only
  - As averages over the ensemble
  - Medoid model only



Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958



EM validation reports

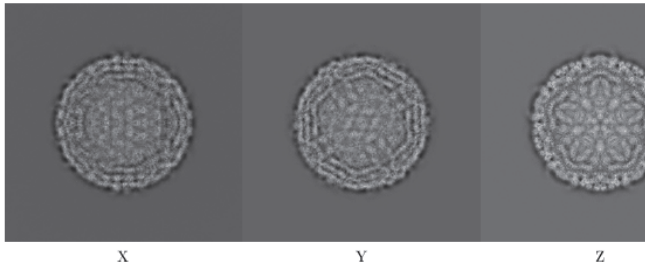
- Prototype EM map-validation reports
  - Most of the PDBe “Visual analysis” functionality implemented

4 Map analysis

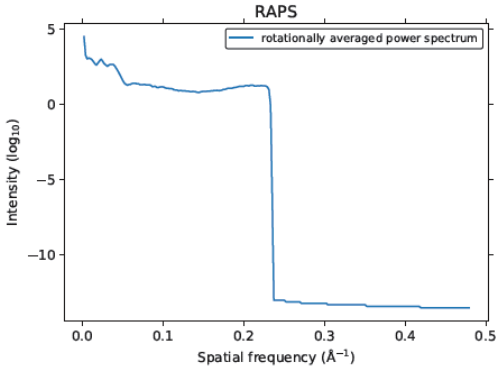
4.1 Map parameters

Property	Value
Endianness	little-endian
Pixel size	1.04
Axis order	XYZ
Number of pixels in X	768
Number of pixels in Y	768
Number of pixels in Z	768
Minimum density	-19.308
Maximum density	26.264
Average density	0.0267
Standard deviation of densities	1.129
Range of densities	45.572
Recommended contour	3.5

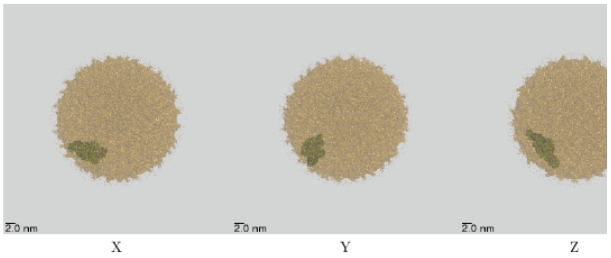
4.2 Orthogonal projections



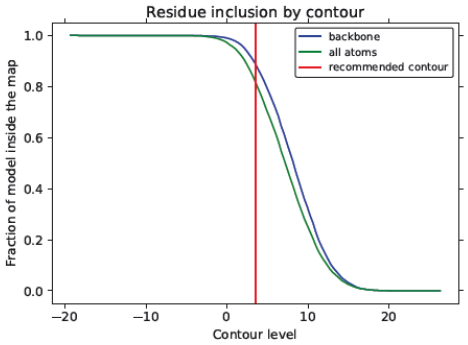
4.7 Rotationally averaged power spectrum



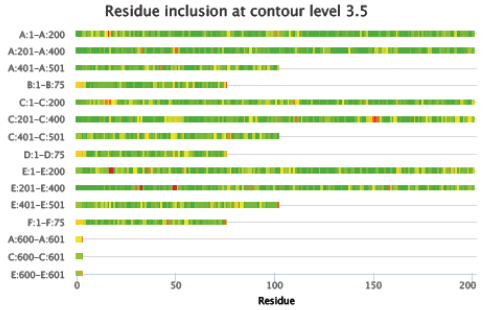
4.8 Model to map fitting



4.9 Residue inclusion by contour



4.10 Residue inclusion at recommended contour



Compare to: pdbe.org/emd-8116/analysis

Ligands in proteins

- So you have successfully navigated all the hazards so far have great data, well integrated, successful MR, refinement model building, Ramachandran analysis ……
- You have density in the active site and the whole point of the structure is to find how the interesting drug candidate ligand binds …….
- Here be dragons!