Discussion Group:

Model building, fitting & validation using high-resolution cryo-EM maps

Doreen Matthies, Ph.D.

Subramaniam Lab
Laboratory of Cell Biology, NCI/CCR
National Institutes of Health (NIH)

Motivation

Good news: Technological and methodological improvements in signal detection and data processing have made it possible to determine biomolecular structures at (pseudo)atomic resolution.

However: A general perception is that the quality of atomic models derived from cryo-EM reconstructions is typically suboptimal, when compared to crystal structures obtained from X-ray diffraction at similar resolution.

"XPLOR-NIH (Maki-Yonekura et al., 2010), CNS (Cheng et al., 2011) and Phenix.refine (Baker et al., 2013) have previously been used for refinement of models into cryo-EM data by adopting a pseudo-crystallographic approach. However, ***many structures deposited alongside high-resolution (4 Å or better) cryo-EM reconstructions have not been refined and consequently have worse stereochemistry than crystal structures*** *solved at similar resolutions." (Brown et al., accepted)*

Are the existing tools for model building and validation adequate?

Have these tools been optimized for EM density maps?

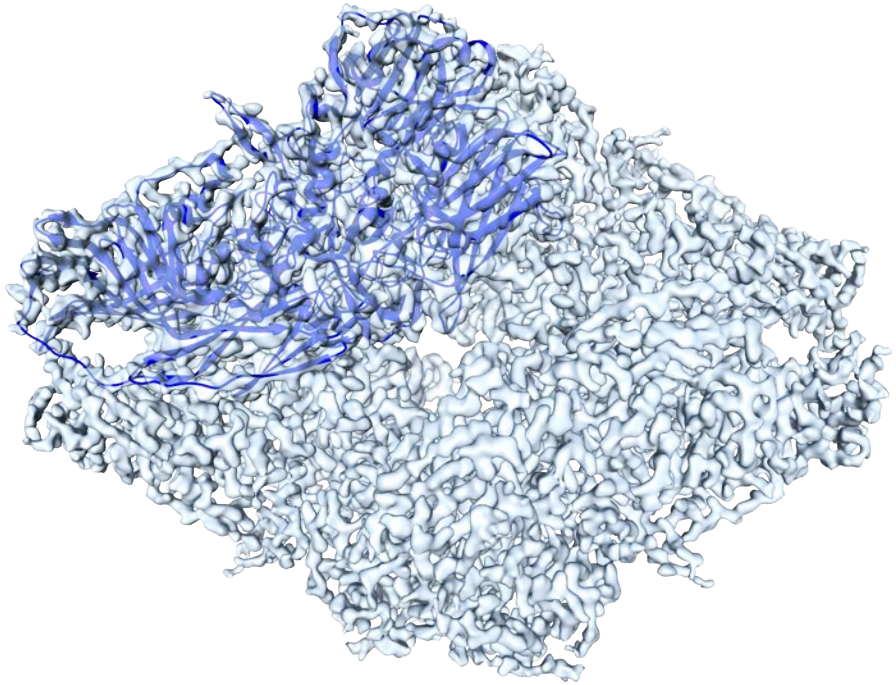
Is there a lack of awareness of what tools are already available?

Would it be desirable to create/adopt a standardized description of model quality?

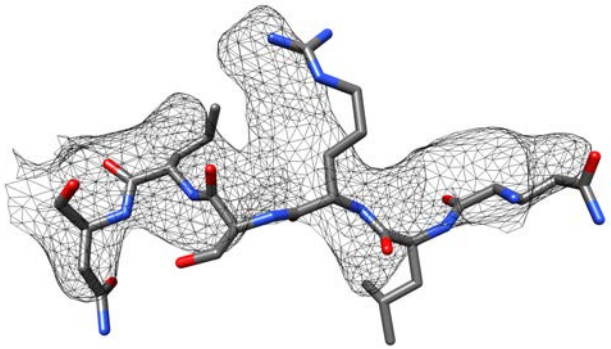
Overview of some published structures

Reference	Target	Resolution	Fitting	Model building	Refinement	Validation
Miyazawa et al., 2003	Acetyl-choline receptor pore (2D)	4.0 Å	Program O	Program O	Program O	PROCHECK
Ludtke et al., 2008	GroEL	4.2 Å	COOT	SSEHunter COOT	COOT	Comparison to X-ray structure
Cong et al., 2010	Mammalian chaperonin TRiC/ CCT	4.0 Å	Chimera COOT	MODELLER	COOT	COOT?
Yu et al., 2011	Cytoplasmic polyhedrosis virus (CPV)	3.1 Å	COOT	COOT REMO	CNS	(CNS force field)
Li et al., 2013	20S proteasome	3.3 Å	Chimera	-	MDFF	(MDFF force field)
Liao & Cao et al., 2013	Rat TRPV1 channel	3.3 Å		COOT	COOT	COOT?
Allegretti et al., 2014	F_{420} reducing hydrogenase	3.4 Å	Chimera COOT	COOT	COOT	COOT?
Amunts & Brown & Bai et al., 2014	Yeast mitochondrial large ribosomal Su	3.2 Å	Chimera? MOLREP COOT	RCrane I-TASSER	COOT REFMAC v.5.8 ERRASER-PHENIX	MolProbity FSC_{work} & FSC_{test}
Wong & Bai et al., 2014	Plasmodium falciparum 80S ribosome	3.2 Å	Chimera COOT	I-TASSER	REFMAC v.5.8 ERRASER-PHENIX	MolProbity, FSC_{work} & FSC_{test}
Voorhees & Fernandez et al., 2014	Mammalian ribosome in complex with Sec61	3.4 Å	Chimera COOT	COOT	REFMAC v.5.8	FSC_{work} & FSC_{test}
Bartesaghi & Matthies et al., 2014	β-galactosidase	3.2 Å	Chimera COOT	COOT	COOT	COOT MolProbity
Lu & Bai et al., 2014	Human γ-secretase	4.5 Å	COOT	COOT	REFMAC v.5.8	FSC_{work} & FSC_{test}

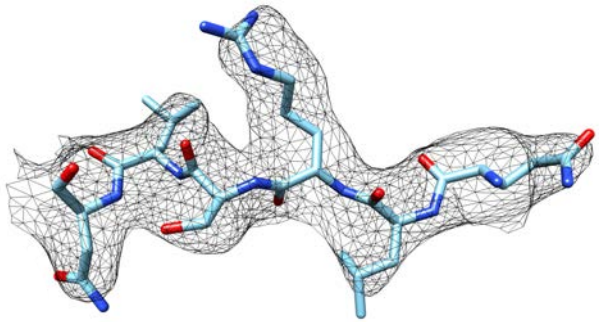
3.2-Å β-galactosidase - fitting



Rigid-body fitting of a single subunit of an X-ray structure using UCSF Chimera

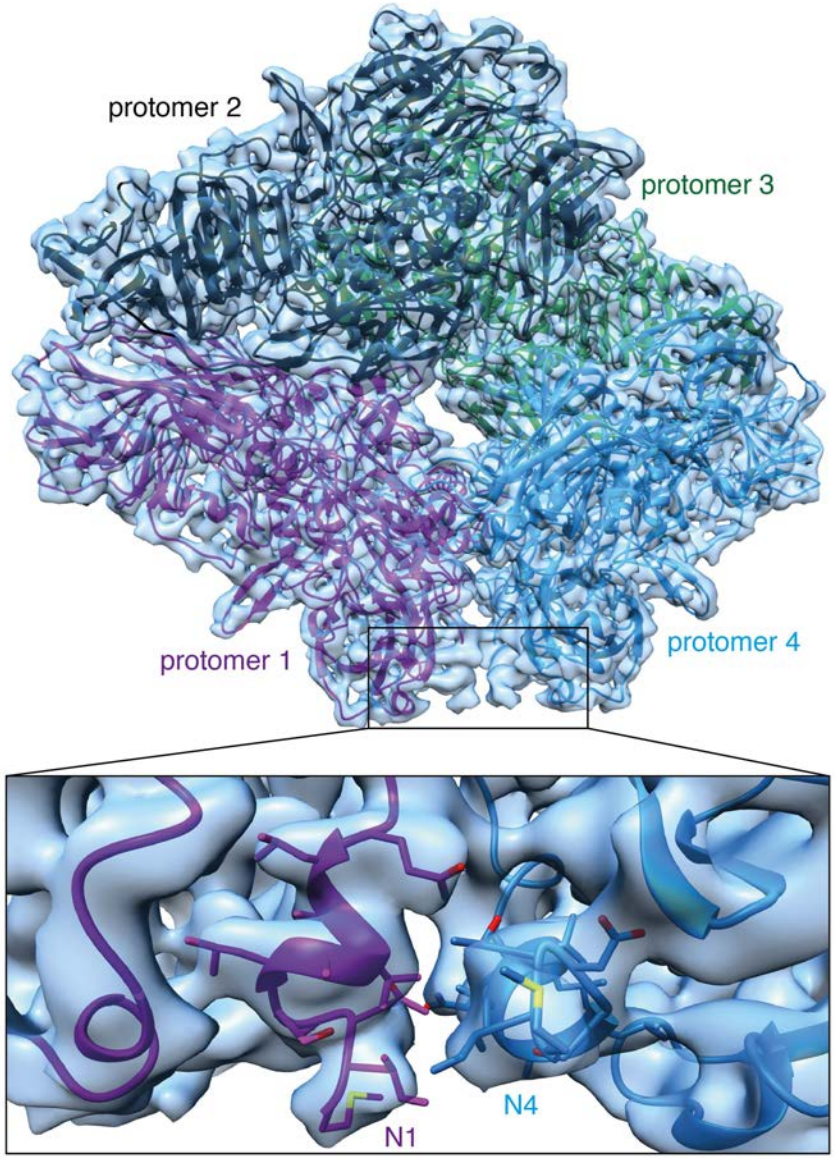


Flexible fitting and real space refinement using COOT



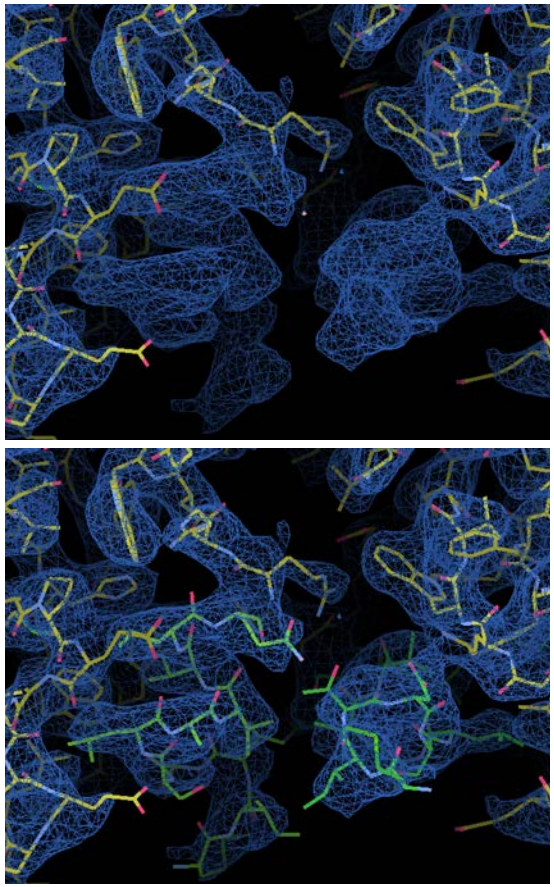
Bartesaghi & Matthies *et al.*, 2014

3.2-Å β-galactosidase - building



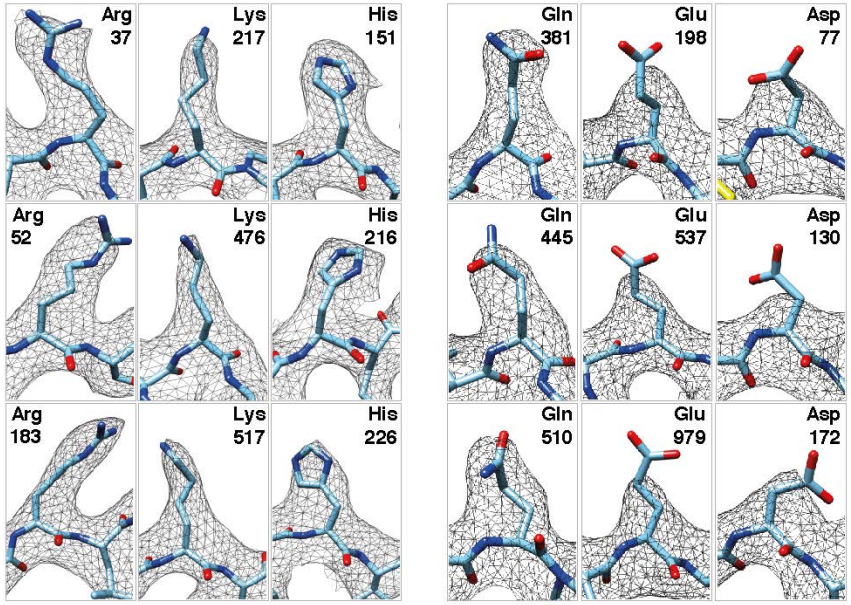
- N-terminal domain of unknown structure
- areas with low correlation were deleted and rebuilt

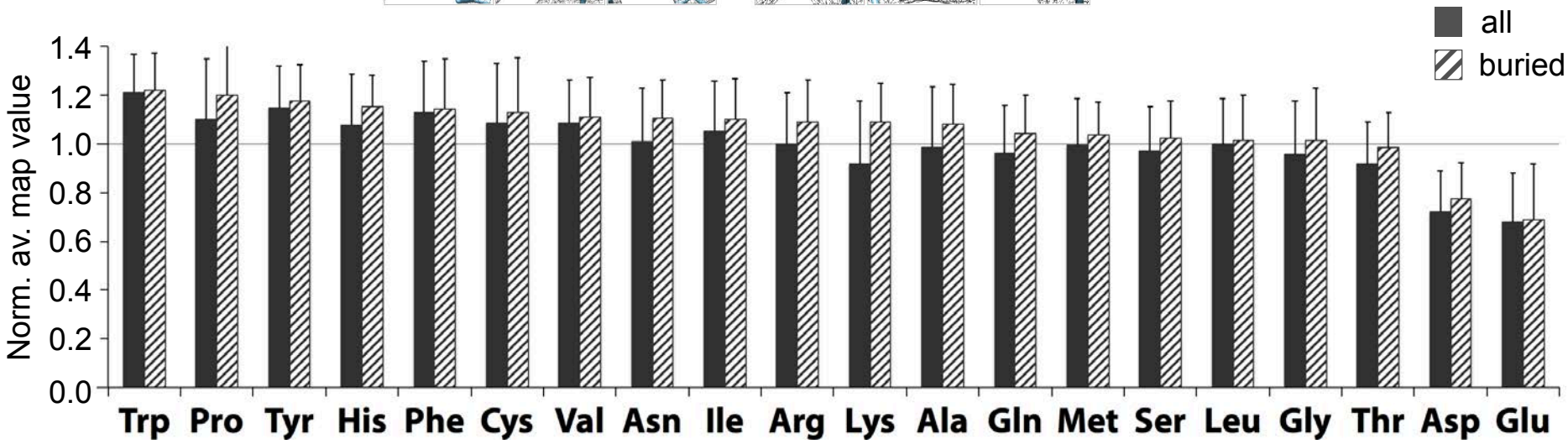
-> addition of N- or C-terminal residues to the model one by one in COOT followed by refinement



Bartesaghi & Matthies *et al.*, 2014

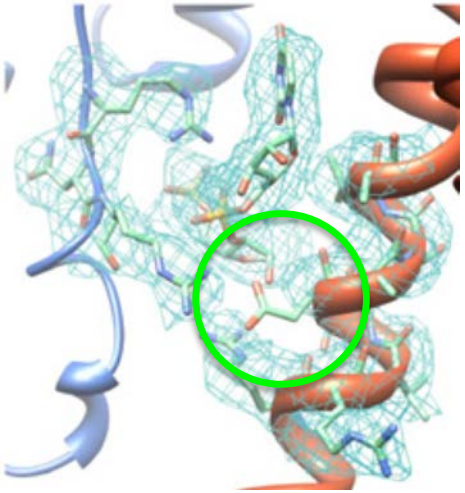
Lower density for glutamates and aspartates



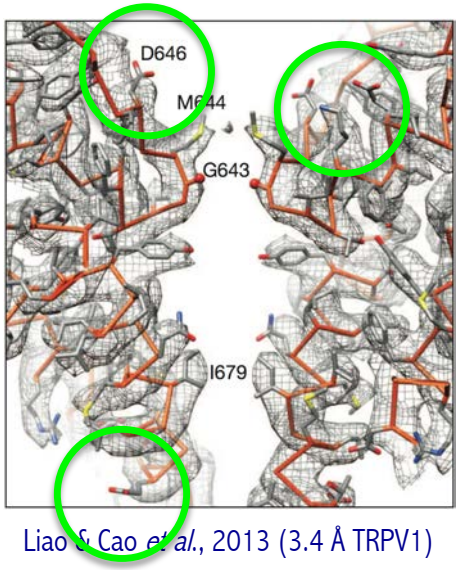


Bartesaghi & Matthies *et al.*, 2014

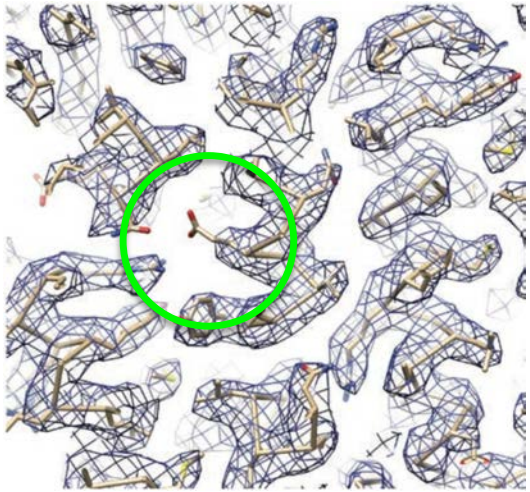
Lower density for glutamates and aspartates



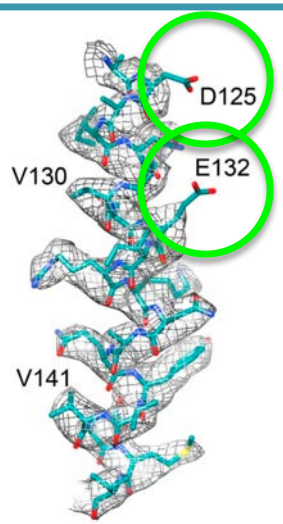
Ge & Zhou, 2011 (3.3 Å virus)



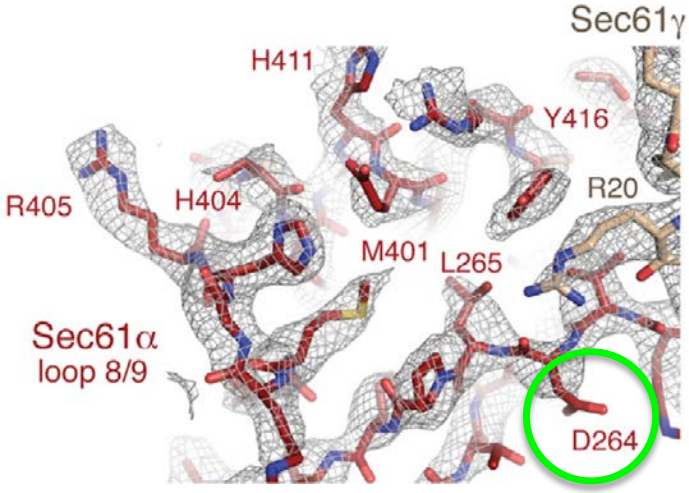
Liao & Cao *et al.*, 2013 (3.4 Å TRPV1)



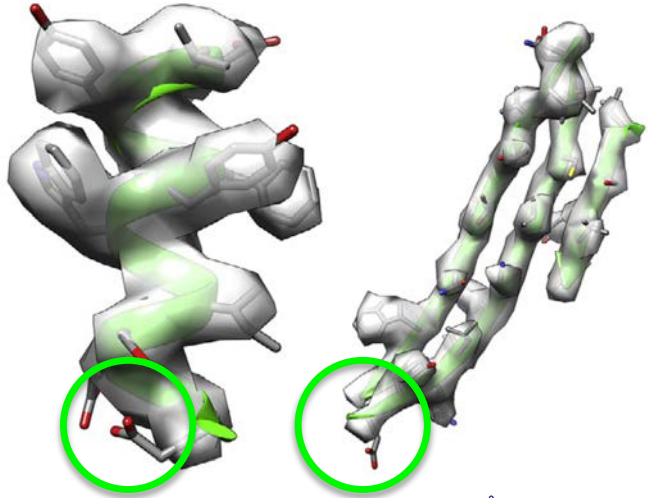
Li *et al.*, 2013 (3.3 Å 20S proteasome)



Allegretti *et al.*, 2014 (3.36 Å FRH)



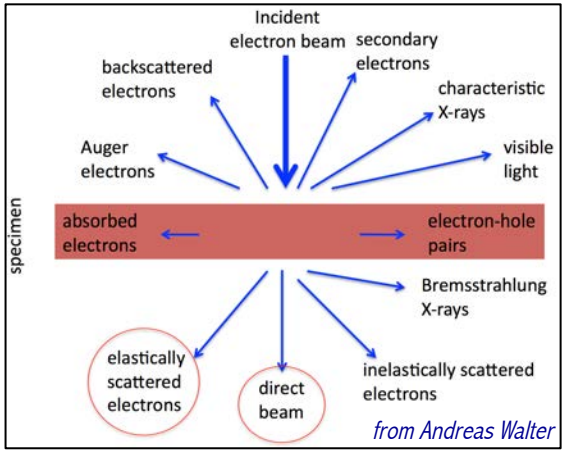
Vorhees *et al.*, 2014 (3.4 Å riboome-Sec61)

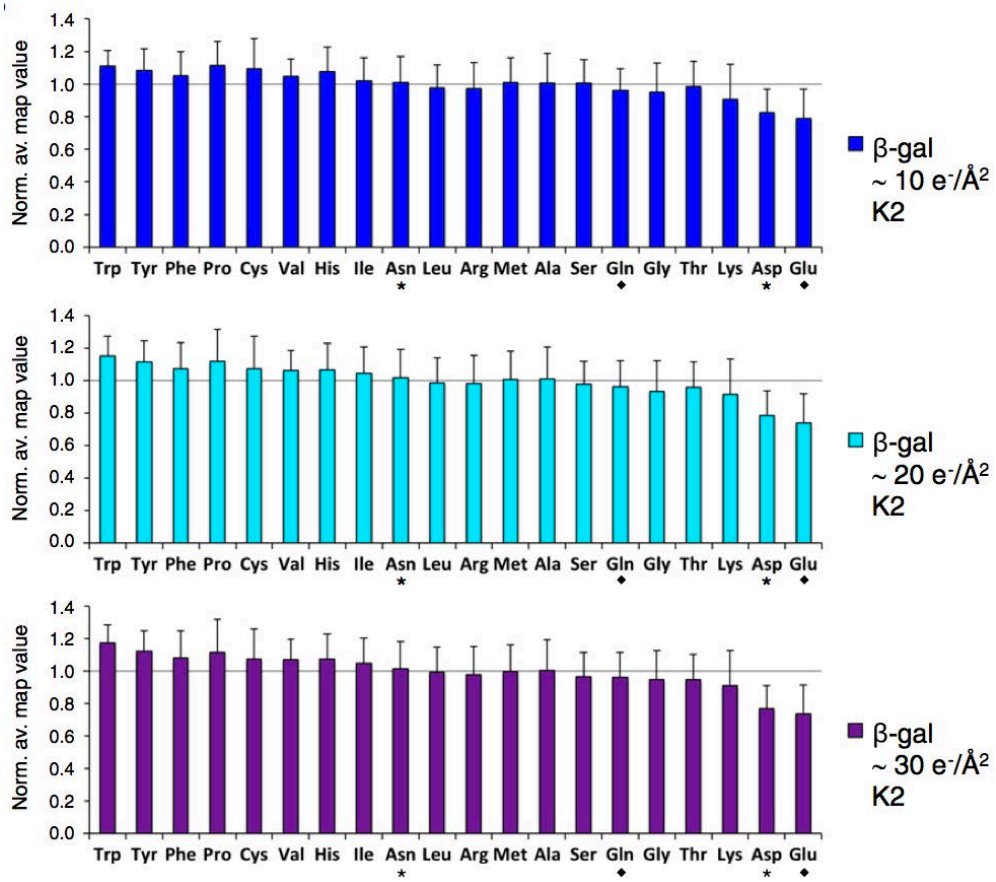


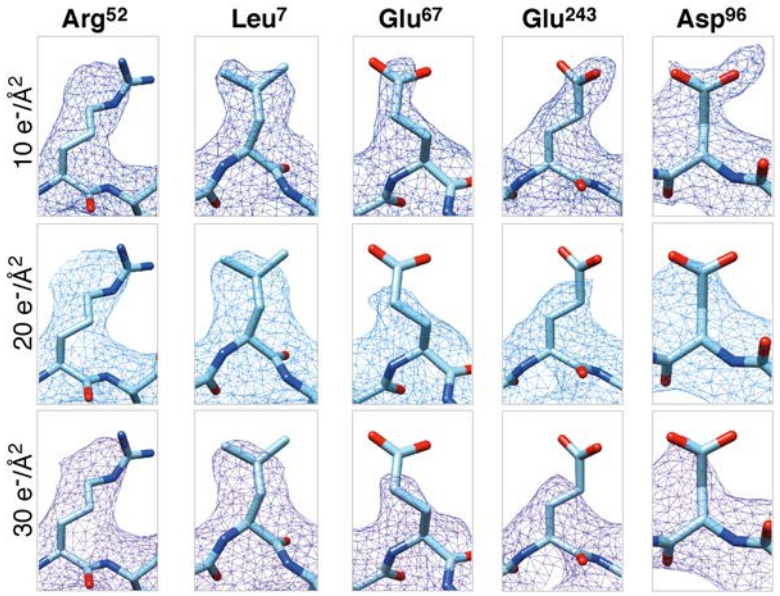
Campbell & Kearney *et al.*, 2014 (3.7 Å virus)

Residue-specific radiation damage?

- Radiation damage or effect of electron scattering, which is influenced by local electric charges and ionization states, or both, or is it something completely different?
- How do we model the side chains if we do not have enough signal?
- How many maps should we make available?
- Which maps/data should be used for validation & b-factor calculations?

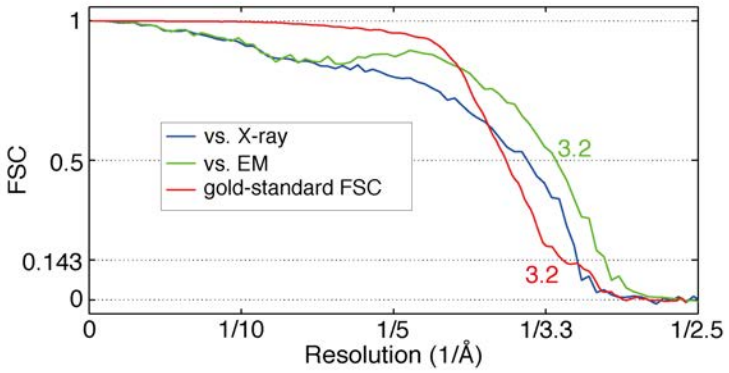


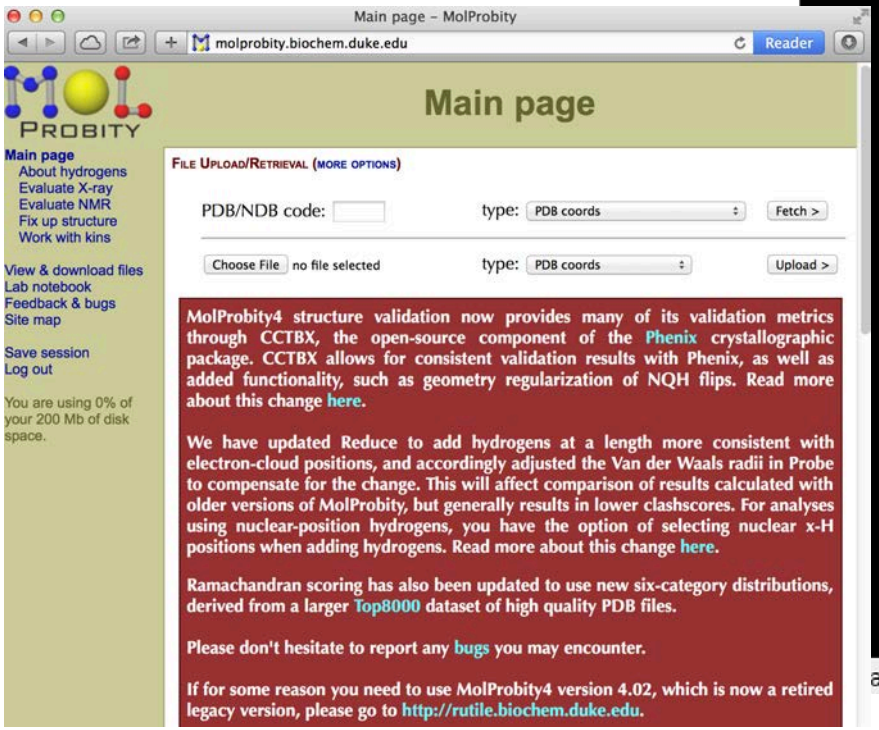


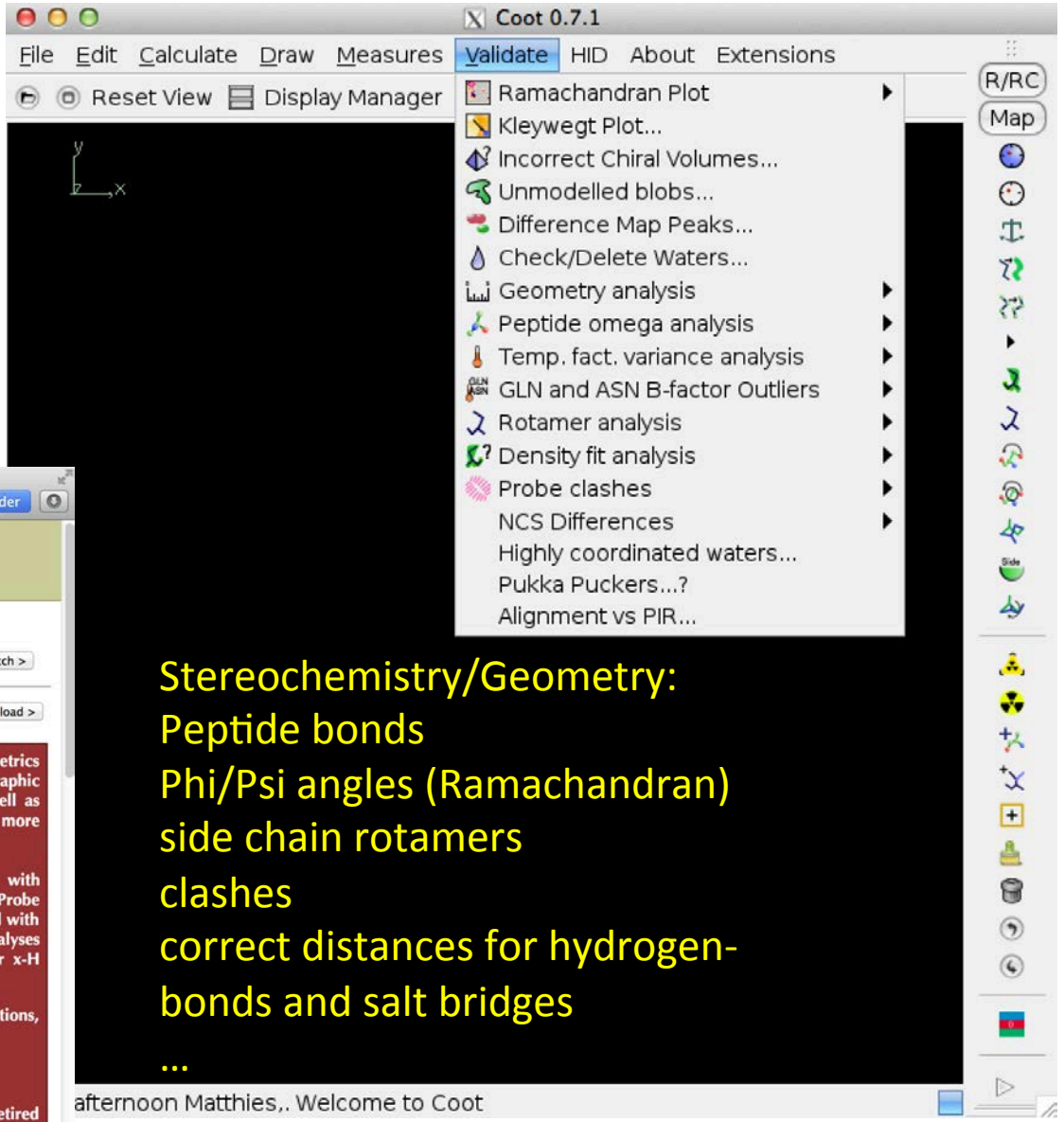


Bartesaghi & Matthies *et al.*, 2014

3.2-Å β-galactosidase - validation







Stereochemistry/Geometry:
Peptide bonds
Phi/Psi angles (Ramachandran)
side chain rotamers
clashes
correct distances for hydrogen-bonds and salt bridges
...

Software overview

Rigid-body fitting

- 3SOM
- ADPEM
- Attract-EM
- BCL::EM-Fit
- CoAn/CoFi
- EMatch
- Emfit
- EMLZerD
- GMFit
- IMP
- IQP
- MultiFit
- Situs
- UCSF Chimera
- X-PLOR
- …

Flexible fitting & Refinement

- CNS-DEN
- COOT
- DireX
- Emap (CHARMM)
- EM-IMO
- Flex-EM / RIBFIND
- FRODA
- iMODFIT
- IMP
- MapSGLD (CHARMM)
- MDFF (NAMD)
- MDfit
- NMFF
- NORMA
- Phenix.refine
- REFMAC
- ROSETTA
- RSRef
- S-flexfit
- YUP.SCX
- XPLOR-NIH
- …

Integrative Modeling

- IMP
- ROSETTA

Model building

- COOT
- EM-fold
- MODELLER /Mod-EM
- ROSETTA
- TASSER
- Assemble2
- Rcrane
- …

Villa & Lasker, 2014; …

Model validation

- ADP-EM
- COOT
- MolProbity
- ProQM
- Situs
- …

Cross-validation methods

- (Amunts & Brown & Bai & Llacer *et al.*, 2014) (Wong & Bai & Brown *et al.*, 2014)
- (DiMaio *et al.*, 2013)
- (Shaikh *et al.*, 2003)
- (Falkner & Schröder, 2013)

Discussion Topic 1: Model Building

- Homology modeling vs. *de novo* building:
  When is the sequence identity too low?
  - *uncertainties in sequence alignments*
  - *validity of homology-modeling premise*

Some software tools

- Homology modeling
  - MODELLER
  - TASSER
- *De novo* modeling
  - COOT
  - Rosetta
  - EM-fold (helices)
  - Assemble2 (RNA)
  - COOT/Rcrane (RNA)

- Are automated *de novo* methods ready to replace manual approaches?
  - *sufficiently fast for large systems?*
  - *easy to use? (installation, documentation, usage)*
  - *specialized high-performance computing equipment?*

- What are the users' experiences?

Discussion Topic 2: Flexible Fitting & Refinement

FF used to alter atomic model to conform to the EM map

Challenge: maps contain errors and uncertainties difficult to quantify and are transferred into the final model;
FF algorithms can also introduce errors from limitations in sampling and scoring, particularly at mid-resolutions

- Manual vs. automated fitting, or both?
- Real-space vs. reciprocal-space refinement, or both?
- What are the advantages and disadvantages of
  - traditional methods used for X-ray data
  - methods based on MD/Monte-Carlo simulations
  - molecular-modeling methods
- Is the choice of method dependent on resolution?
- What are the users' experiences?

Some software tools

- Traditional X-ray tools
  - CNS
  - XPLOR
  - COOT
  - Phenix.refine
  - REFMAC
  - RSRef (CNS)
- Molecular modeling tools
  - Rosetta
  - DireX
  - Flex-EM (MODELLER)
  - IMP
  - EM-IMO
  - S-flexfit
- MD/MC Simulation-based methods
  - MDFF (NAMD)
  - EMAP (CHARMM)
  - MDFit
  - FRODA
- Normal-modes/elastic-network
  - iMODFIT
  - NMFF
  - NORMA
  - YUP.SCX

Discussion Topic 3: Model validation

I. Stereochemistry & geometry

Peptide bonds, Phi/Psi angles (Ramachandran), side chain rotamers & clashes, correct distances for hydrogen-bonds and salt bridges, etc.

II. Structural model vs. data, i.e. cross-validation

Need for generally accepted cross-validation method, e.g. FSC_{work} vs. FSC_{test}, as well as a local descriptor of confidence in coordinate assignments (such as B-factors in X-ray crystallography)

III. Independent experimental validation

Crosslinking, Cys-accessibility measurements, EPR/FRET, mutants, …

- What are the users' experience and opinions?

Some software tools

- ADP-EM
- COOT
- MolProbity
- ProQM (membrane proteins)
- Situs

Cross-validation

- FSC_{work} vs. FSC_{test} (Amunts & Brown & Bai & Llacer *et al.*, 2014; Wong & Bai & Brown *et al.,* 2014; Fernandez *et al.*, 2014)
- Splitting data in half and only use one half for model building and refinement (DiMaio *et al.*, 2013)
- Omitting data from the high spatial frequency range (Falkner & Schröder, 2013)
- Exclusion of resolution shells in reciprocal space (Shaikh *et al.*, 2003)