Molecular Dynamics Flexible Fitting

Ryan McGreevy Research Programmer

University of Illinois at Urbana-Champaign NIH Resource for Macromolecular Modeling and Bioinformatics

Molecular Dynamics Flexible Fitting (Ribosome-bound YidC)

Electron Microscope

EM density

map

Match through MD

Supercomputer

crystallographic structure

APS

Synchrotron

Molecular Dynamics Flexible Fitting -Theory

Two terms are added to the MD potential

 $U_{total} = U_{MD} + U_{EM} + U_{SS}$

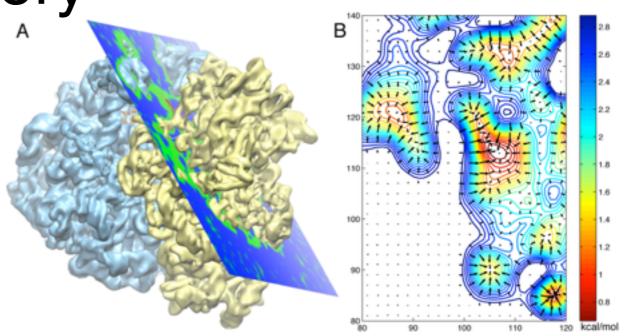
An external potential derived from the EM map is defined on a grid as

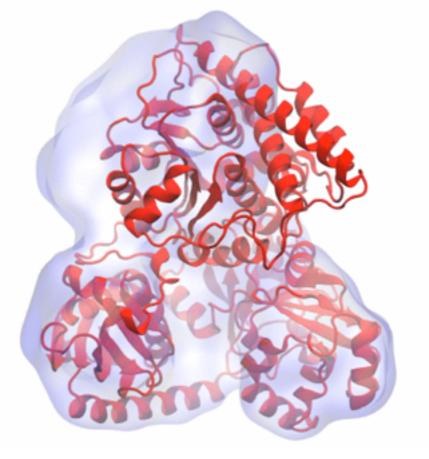
$$U_{EM}(\mathbf{R}) = \sum_{j} w_{j} V_{EM}(\mathbf{r}_{j})$$
$$V_{EM}(\mathbf{r}) = \begin{cases} \xi \left(1 - \frac{\Phi(\mathbf{r}) - \Phi_{thr}}{\Phi_{max} - \Phi_{thr}}\right) & \text{if } \Phi(\mathbf{r}) \ge \Phi_{thr}, \\ \xi & \text{if } \Phi(\mathbf{r}) < \Phi_{thr}. \end{cases}$$

A mass-weighted force is then applied to each atom

$$\mathbf{f}_i^{EM} = -\nabla U_{EM}(\mathbf{R}) = -w_i \partial V_{EM}(\mathbf{r}_i) / \partial r_i$$

[1] Trabuco et al. *Structure* (2008) 16:673-683.[2] Trabuco et al. *Methods* (2009) 49:174-180.





Acetyl – CoA Synthase

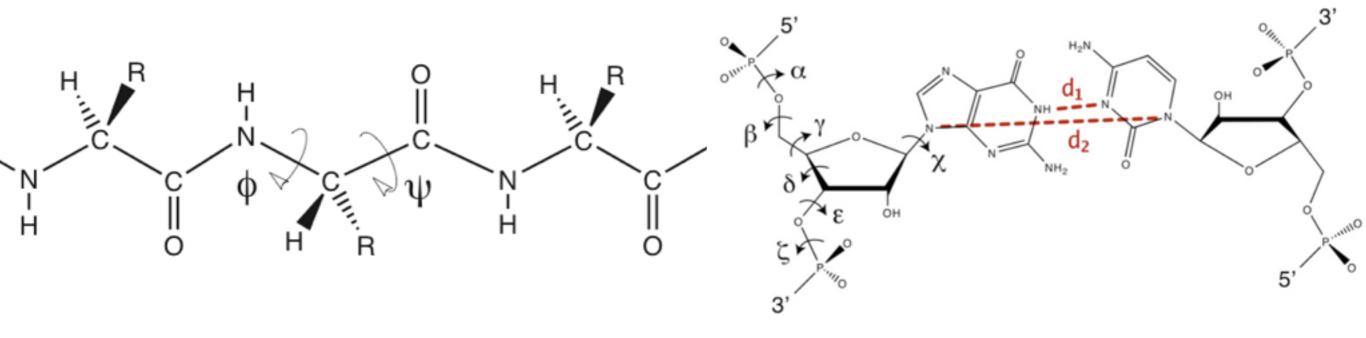
Secondary structure restraints

Harmonic restraints are applied to preserve secondary structure of proteins and nucleic acids, avoiding "overfitting."

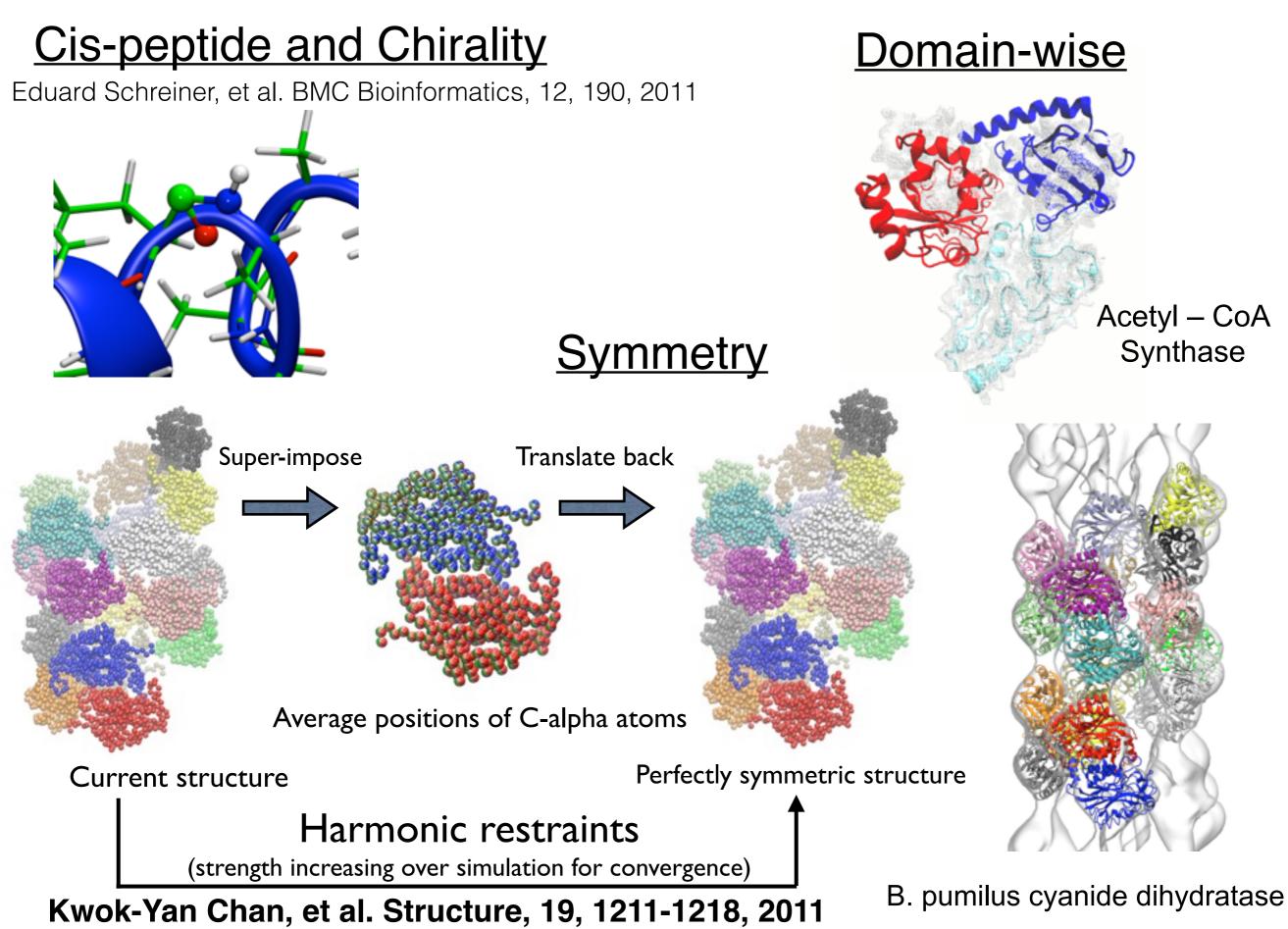


For proteins, ϕ and ψ dihedral angles of residues within helices or beta strands are restrained.

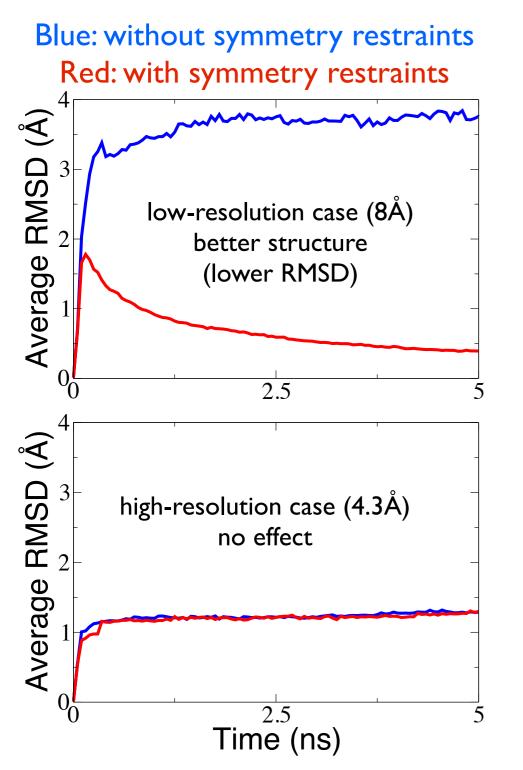
For nucleic acids, distance and dihedral restraints are applied to a selected set of base pairs.

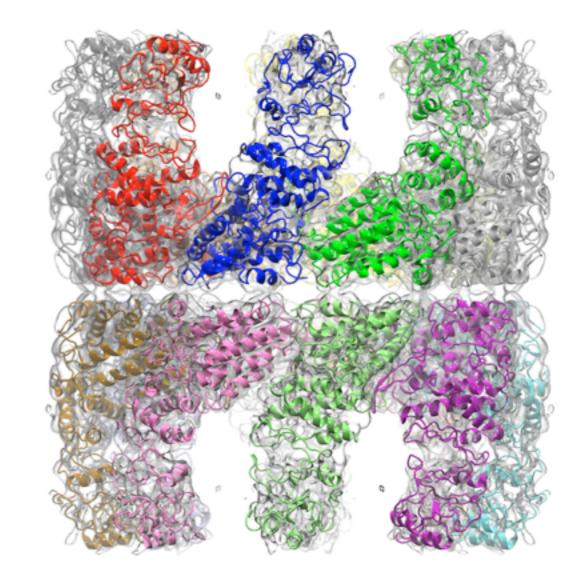


Additional Restraints



Symmetry restrained MDFF - Test Case 1 Improve quality of fit for low-resolution data

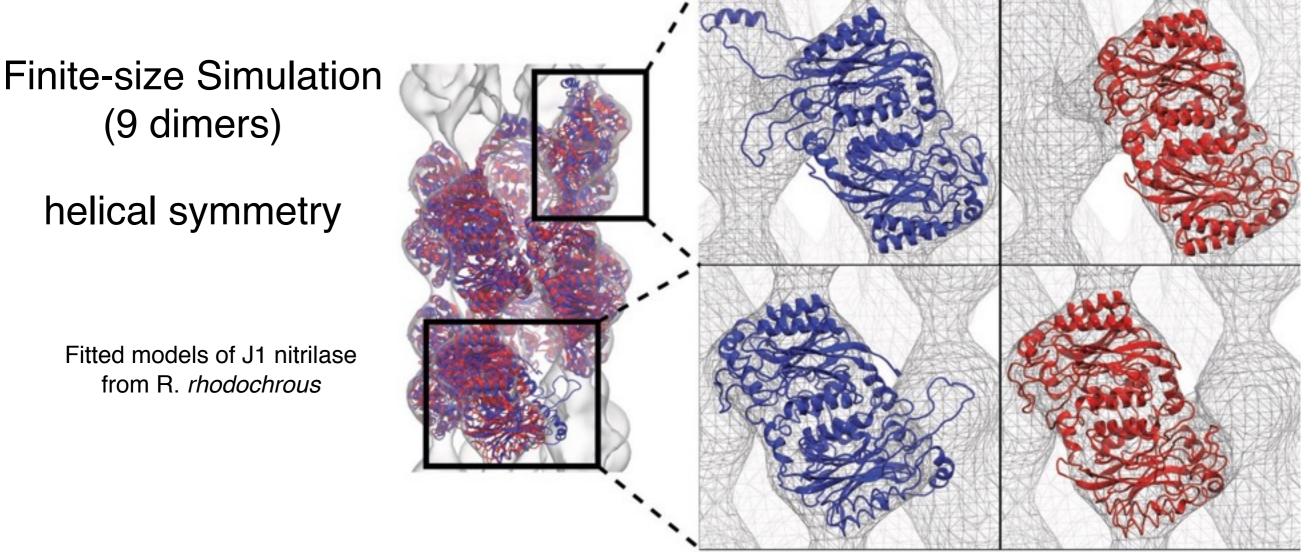




Archaeal group II chaperonin from *M. maripaludis* (Mm-cpn) 8-fold rotational + 2 fold reflection symmetry homology model (based on PDB 3LOS) fitted into EM map (EMDB 5140)

Chan et al. *Structure* (2011) 19:1211-1218.

Symmetry restrained MDFF - Test Case 2 Prevent "edge distortion effect"



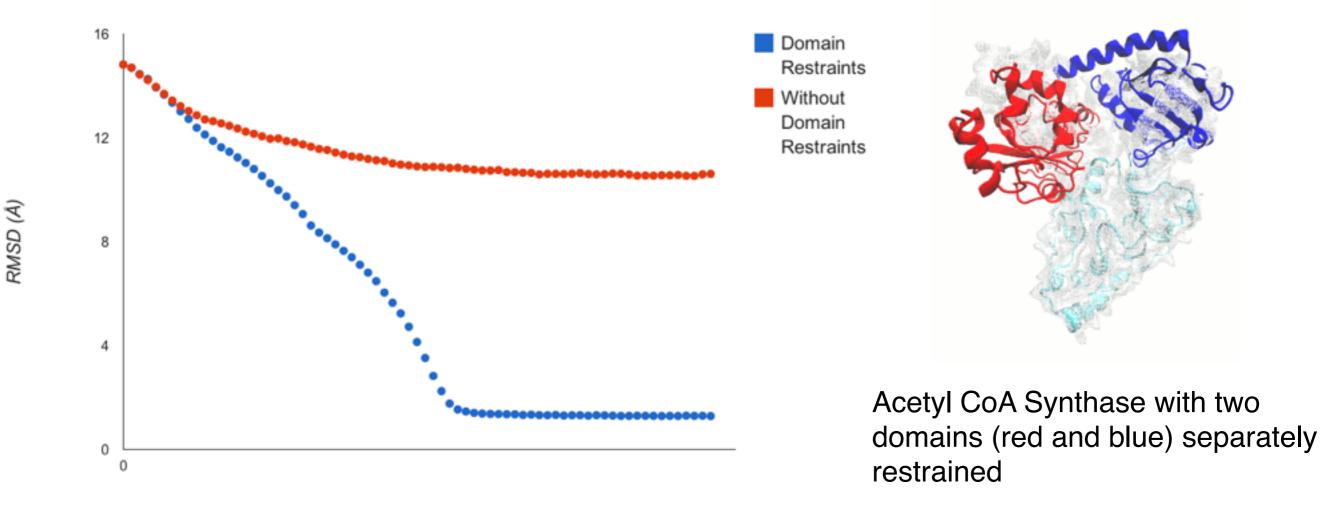
homology model and EM map (EMD 1313) from collaborator T. Sewell, U. of Cape Town

Without Symmetry Restraints With Symmetry Restraints

Chan et al. Structure (2011) 19:1211-1218.

Domain restrained MDFF

Use Targeted MD (TMD) feature of NAMD to restrain nonoverlapping groups of atoms to maintain rigid domains

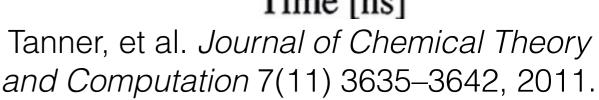


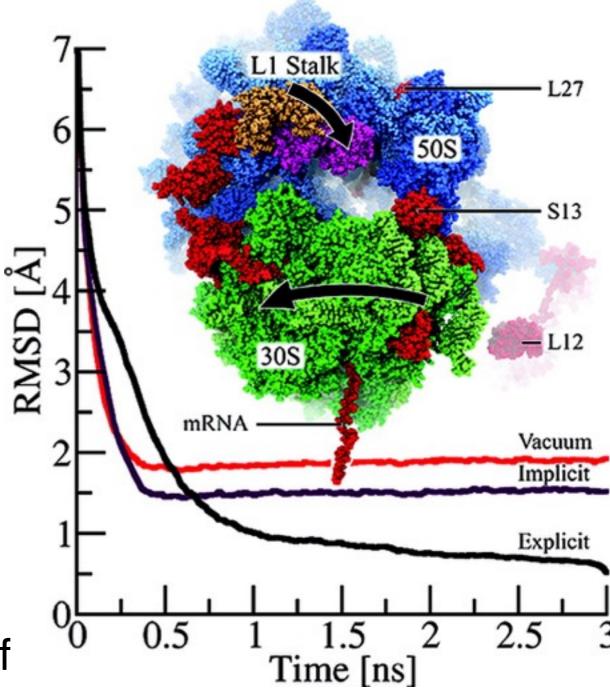


Simulation Environment

MDFF can be run in different environments:

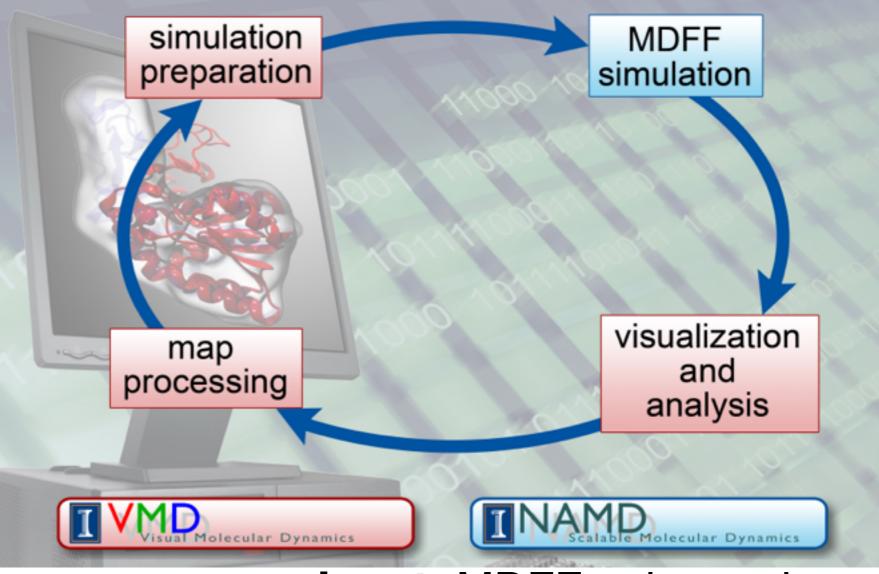
- 1. <u>Vacuum</u>
 - No water molecules
 - Fastest but potentially inaccurate
- 2. Explicit Solvent
 - Explicit atomic detail water molecules
 - Computationally slow and introduces effects of viscous drag
- 3. Implicit Solvent
 - generalized Born approximation of electrostatics
 - Compromise between speed and accuracy





MDFF Software Suite

- NAMD and VMD • used together to run **MDFF**
- **Every NAMD and** • VMD feature is available in MDFF



Fitting time is dependent on:

- system size
- map and structure quality
- YidC: 6 months to model; 6 hours MDFF (20ns); workstation

Input: MDFF only requires a PDB, PSF, and density map

Output: produces simulation trajectory from which an ensemble of structures can http://www.ks.uiuc.edu/Research/mdff/be extracted

MDFF Software Suite

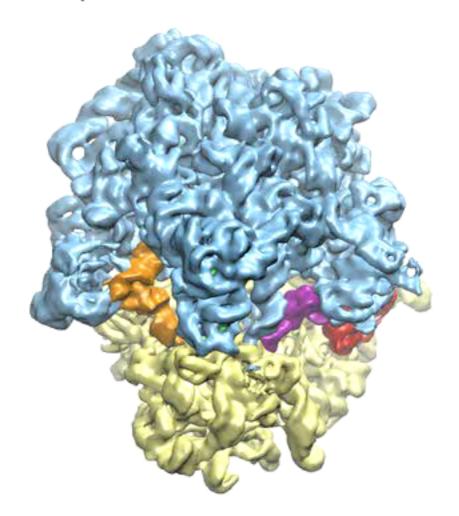
- system sizes up to 100 million atoms (viruses, chromatophore)
- maps from 3 to 15 Å
- runs on laptops to petascale computing resources (Blue Waters, Titan)
- New MDFF GUI (VMD 1.9.2) makes setting up, running, and analyzing fitting simulations even easier

File ▼ MDFF Setup MDFF Connect Submit Connect Pause Finish MD Status: Step 50200 ▼ Cross Correlation Analysis ★ Calculate Real-Time Cross Correlation Experimental Density (Mol ID): O Selection: Use Threshold Threshold Elle Feed-Time Cross Correlation Integractive GPU-accelerated Cross Correlation 0.825	MDFF GUI _ D 3				
Submit Connect Pause Finish MD Status: Step 50200 ▼ Cross Correlation Analysis ▼ X Calculate Real-Time Cross Correlation Experimental Density (Mol ID): 0 0 0 Selection: protein and noh map Resolution: 5 0 Use Threshold:	File 🔻				
MD Status: Step 50200 ▼ Cross Correlation Analysis × Calculate Real-Time Cross Correlation Experimental Density (Mol ID): 0 Selection: protein and noh Map Resolution: 5 Use Threshold Threshold: Elle Real-Time Cross Correlation 0.925 Cross Correlation 0.775 0.725 0.725 0.725 0.725	MDFF Setup	IMDFF Connec	t		
Cross Correlation Analysis * Calculate Real-Time Cross Correlation Experimental Density (Mol ID): 0 Selection: protein and noh Map Resolution: 5 Use Threshold Threshold Elle Real-Time Cross Correlation Cross Correlation Ones: Correlation Cross Correlation Cross Correlation Cross Correlation Cross Correlation Ones: Correlation Ones: Correlation Ones: Correlation Ones: Correlation Ones: Correlation Ones: Correlation	Submit		Connect	Pause	Finish
* Calculate Real-Time Cross Correlation Experimental Density (Mol ID): 0 Selection: protein and noh Map Resolution: 5 Use Threshold File File Real-Time Cross Correlation Cross Correlation Cross Correlation 0:825 Cross Correlation 0:825	IMD Status: Ste	p 50200		·	
Experimental Density (Mol ID): 0 Selection: protein and noh Map Resolution: 5 Use Threshold Threshold: Elle Integractive GPU-accelerated Cross Correlation 0.825 Cross Correlation 0.775 0.725 0.675	▼ Cross Corre	lation Analysis			
Selection: protein and noh Map Resolution: 5 Use Threshold Threshold: Eile Real-Time Cross Correlation Integractive GPU-accelerated Cross Correlation 0.825 Cross Correlation 0.775 0.725 0.725 0.725		_			
Map Resolution: 5 Use Threshold Threshold: Eile Interfactive GPU-accelerated Cross Correlation					
Use Threshold Threshold: Elle Real-Time Cross Correlation Interactive GPU-accelerated Cross Correlation 0.825 Cross Correlation 0.775 0.725 0.775		-			
Eile	,				
Elle		ild _			
Real-Time Cross Correlation					
Interactive GPU-accelerated	Elle				
5000 15000 25000 35000 45000 Timestep	on				

http://www.ks.uiuc.edu/Research/mdff/

Molecular Dynamics Flexible Fitting -Example

Cryo-EM map of the E. coli ribosome at 6.7-Å resolution

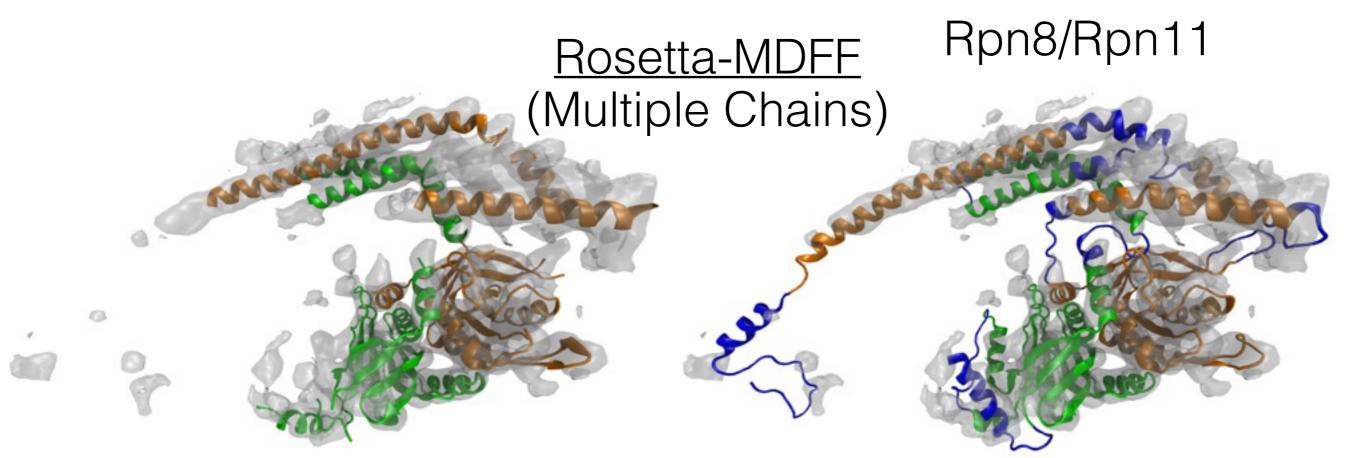


Obtaining Initial Structures

- 1. X-ray crystallography or NMR structures
- 2. Refine structures from low-res X-ray data with xMDFF

Ryan McGreevy*, Abhishek Singharoy*, et al. Acta Crystallographica D70, 2344-2355, 2014

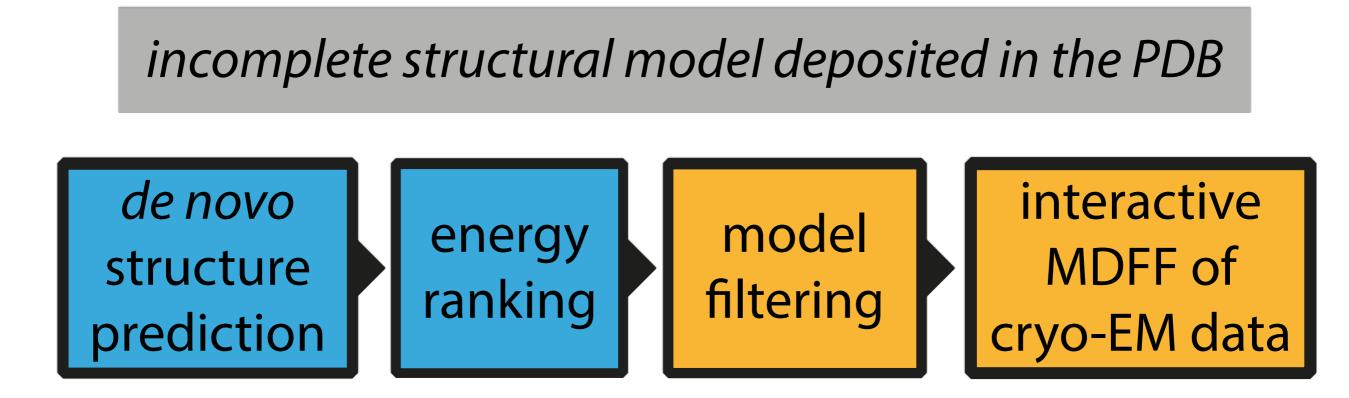
3. Homology or ab initio modeling with Modeller, Rosetta, MUFOLD (Ci-VSP, YidC, Holotranslocon)



Rosetta structure prediction to fill **missing pieces** and MDFF to filter, refine and validate candidate structures

Rosetta/MDFF interactive Modeling

Combining structure prediction with the user's expertise to interpret densities



complete structural model that fits cryo-EM data

Rosetta

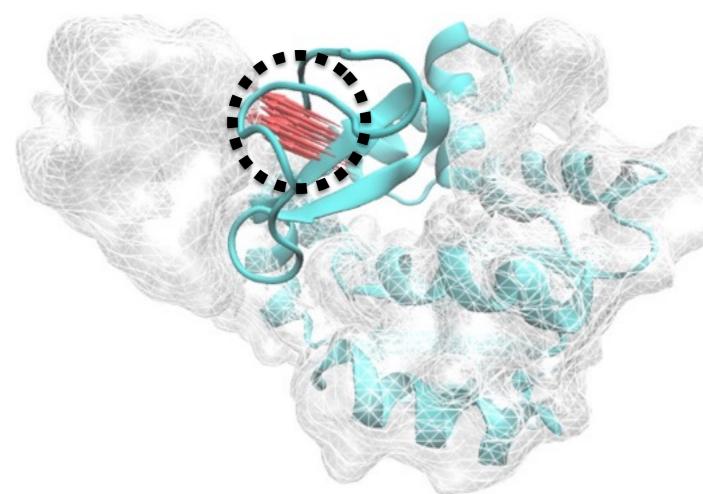
Leaver-Fay *et al.* Methods Enzymol. 2011 Porter *et al.* PLoS One 2015



Humphrey *et al.* J. Mol. Graph. 1996 Philips *et al.* J. Comput. Chem. 2005

Interactive Modeling with MDFF GUI

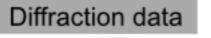
- Apply forces to manually manipulate structure into the density
- Useful for difficult to fit structures with large conformational changes

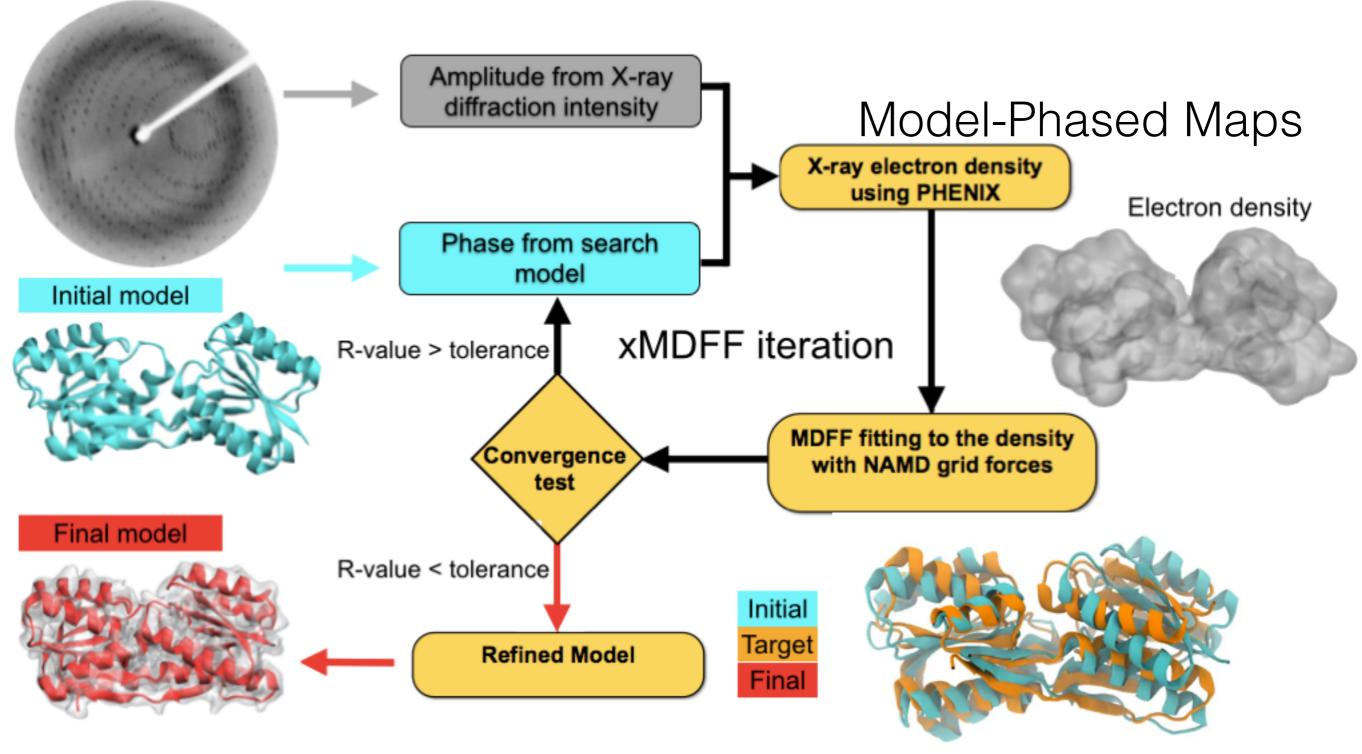


- New MDFF GUI in VMD 1.9.2
- Set up and run interactive (or traditional) MDFF/xMDFF simulations

File MCFF False Working Decchary: Annehystercgreevy/Downloads/htt#Mortal-files/2-indf-vacuo PSF File: Load PSF/PCB Density Map: Load PSF/PCB Ordpd: selection lister: protein and noh Filed PCB selection lister: protein and noh Filed PCB colone Cocoparity Serviced PCB selection lister: protein and noh Filed PCB Colone Cocoparity Serviced PCB selection lister: protein and noh Filed PCB Colone Cocoparity Serviced PCB selection lister: protein and noh Filed PCB Colone Cocoparity Serviced PCB selection lister: protein and noh Filed PCB colone Cocoparity Serviced PCB selection Christly Restants: Projectivinsdipublichroud4/Altivend*Stipuigronhoarch/scineachammpartizipar, stip*_pird_lipid_ina.mp Image: stip MCDFF: Image: stip Image: stip <td< th=""><th>- 1</th></td<>	- 1
MOFFF Fies Working Directory: Anonshyserecgreevy/Downloads/htdfh.ttorial-files/2-rtdff-vacuo PSF File: Load PSF/PDB Load PSF/PDB Load PSF/PDB Density Map: Coopancy Structure of C	
Working Deectary: Aometrysencgreessy/DownloadJungFhdorlah-Hes/2-mdF-vacuo PSF Fie: PDB Fie: Load PSF/PDB Density Map: Gridpb selection text: protein and non Flored PDB Column: Eccupancy Sitevatation Output Name: end Sitevatation Parameters Multiple: MDEF	
PSF File: PDB File: Load PSF/PDB Density Map: Gridpb selection text: Prove PDB Celum: Excerption Exection text: Prove PDB Celum: Selection text: Prove PDB Celum: Coccupancy Structation Output Name: Prove Celuming publication of the selection of the	
POB File: Load PSF/PDB Density Map: Grippids selection text: Fixed PDB celction text: none Fixed PDB celction text: none Fixed PDB celction: cocceptory: Structation Output Step: 1 Restarts: Secondary Stucture Restraints: Citypeptide Restraints Project/vind/put/fixu/d4/fit/vind132/plugin/hoarch/tc/treadcharmspart 2/par_at27_prot_Upid_na.mp Situation Parameters WMDFF MD Parameters MDFF MD Parameters MD Parameters MDFF MD Parameters MD Parameters M	Brows
Lead PSF/PDB	Brows
Density Map: Gridpab selection text: Fixed PDB calume: Coccupancy Simulation Output Step: 1 Restraints: Secondary Stucture Restraints: Projectiv/mdput/fixed4/Abb/md132/plugins/ncarch/cureadchammpar12/par_et22_prot_lpid_na.mp Simulation Parameters Simulation Parameters Simulation Parameters MOD Parameters Generate NAMD files State State State State State State State State Simulation Parameters MOD Parameters MOD Parameters State	Erows
Gridpab selection text: protein and non Fixed PDB selection text: exceptory Simulation Output Name: exit Simulation Output Name: extention Projectivemetry Structure Restraints Chinality Restraints Citipeptide Restraints Parameter Files: Projectivemetry Structure Restraints Chinality Restraints Citipeptide Restraints Parameter Files: Projectivemetry put innued d/strivemetric Disput grish near Chinality Restraints Simulation Parameters MoDFF_ MoDFF_ MoD Parameters Rest-Time Cross Correlation Secondary Structure Restraints Rest-Time Cross Correlation Secondary Structure Restraints Rest-Time Cross Correlation Secondary Structure Output Name Secondary Structure Cross Correlation Secondary Structure Cross Correlation Secondary	
Gridpib selection text protein and non Fload PDB selection text rone Fload PDB selection te	
Placed PDB selection text inone Placed PDB celluteries Simulation Culput Name, indf Simulation Culput Starture Restraints Chirality Restraints Cispeptide Restraints Parameter Files: ProjectivendlyDoffmudd4/stavendlS2/plugmc/ncarch/cistreadcharmepart 2/par_st2?_prof_lipid_na.mp Simulation Parameters MOFF_ N/DOFF_ N/DOFF_ N/D Parameters Moffer_ N/D Parameters Restraints Cross Correlation Poten and non Simulation Parameters Restraints Cross Correlation Poten and non Signature Profectione Correlation Profectione Corr	Brows
Pland PDB Caluer: Struitation Output Step: 1 Restaute: Secondary Structure Restaute: Chirality Restaute: Clipeptide Restaute: Parameter Files: Projectivend judd insudd distivered SQ plangers/noarch/tcs/readcharmspart 2/par_at2?_prot_lipid_na.np • Smutation Parameters • MDFF • MD Parameters • MDFF • MD Parameters • MD Parameters	
Simulation Output Name Simulation Output Step: 1 Restraint: Secondary Structure Restraints Chirality Restraints Projects/vmd/publifmudd4/stb/vmd132/phugms/hoarch4cs/needchammepar12/par_at2?_prot_lput_na.mp Simulation Parameters. VMDFF_ MD Parameters. Generate NAMD files The Cross Correlation Protein and non S Anal Vise Cross Correlation Protein and Name Cross Correlation Protein Anal Vise Cross	
Restraints Secondary Structure Restraints Chipeptide Restraints Projects/widputs/inuod4/lib/wed132/phugins/hoarch/cs/readchamepart 2/par_at27_prot_lipid_namp Simulation Parameters Mo Parameters Mo Parameters Generate NAMD files Footon Analysis FredeTimee Cross Correlation Protein and noh S Coss Correlation Protein and noh Coss Correlation Coss Correlation Protein and noh Coss Correlation Coss Coss Coss Correlation Coss Coss Coss Coss Coss Coss Coss C	
Projectu/vedpublic/ved192/psugerc/ncarch/tstreadcharmepar12/psr_sk27_prot_tpod_namp Simulation Parameters MADFF MO Parameters Generate NAMD files Socoo atom Analysis AFTee Cross Correlation Follen and noh S G G Socoo Anal Vze interactive	
Projects/ved/pub/inued44/sb/ved/32/plugino/hoarch/dc/ineedchamemparl 2/par_al2?_prof_lipid_na.np Sexulation Parameters MADEF. MO Parameters Generate NAMD files Sozoo Sozoo Sozoo Resel-Time Cross Correlation Frotein and non S Coss Correlation Resel-Time Cross Correlation Frotein and non S Coss Correlation Resel-Time Cross Correlation Frotein and non S Coss Correlation Coss Correlation Frotein and non S Coss Correlation Frotein and non Coss Correlation Frotein and non Coss Coss Correlation Frotein and non Coss Coss Correlation Frotein and non Coss Coss Coss Coss Coss Coss Coss	
Structure Parameters MADEF. MO Parameters MO Parameters Generate NAMD files S0200 aton Analysis at Time Cross Correlation protein and non S Model Time Cross Correlation Maging Structure Correlation Maging Structure Correlation Maging Structure Correlation	
bit of the second secon	Add
MADEFGenerate NAMD files In the second sec	Remo
MADFFGenerate NAMD files Generate NAMD files S0200 atom Analysis at-Time Cross Correlation protein and noh S Peat-Time Cross Correlation Peat-Time Cross Correlation S Anal yze interactive	
Analyze interactive	
B B B B B B B B B B B B B B	
Analyze interactive	
mulations in real-time	
0.725-	
0.075	
5000 15000 25000 45000 Timestep	

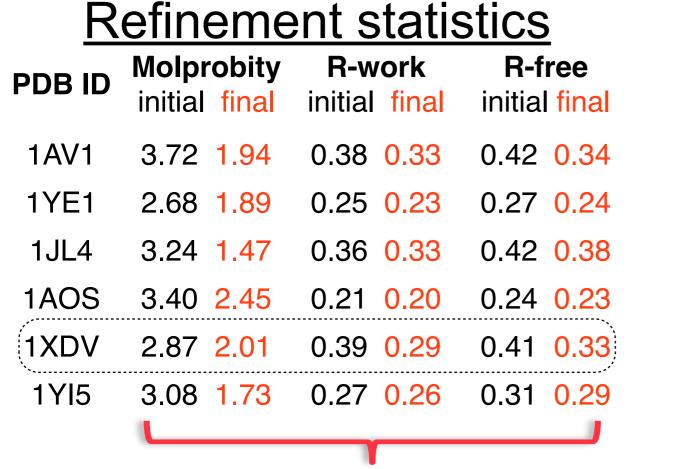
xMDFF: MDFF for low-resolution x-ray crystallography

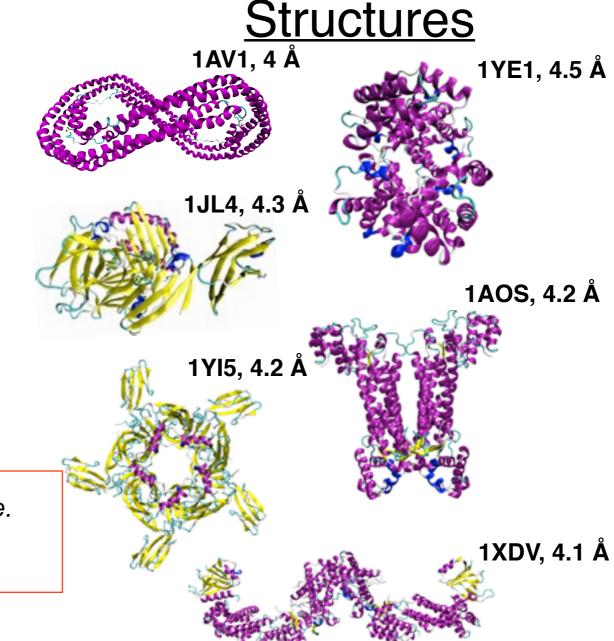




Ryan McGreevy*, Abhishek Singharoy*, et al. Acta Crystallographica D70, 2344-2355, 2014

xMDFF Improves Structures Posted at the Protein Data Bank



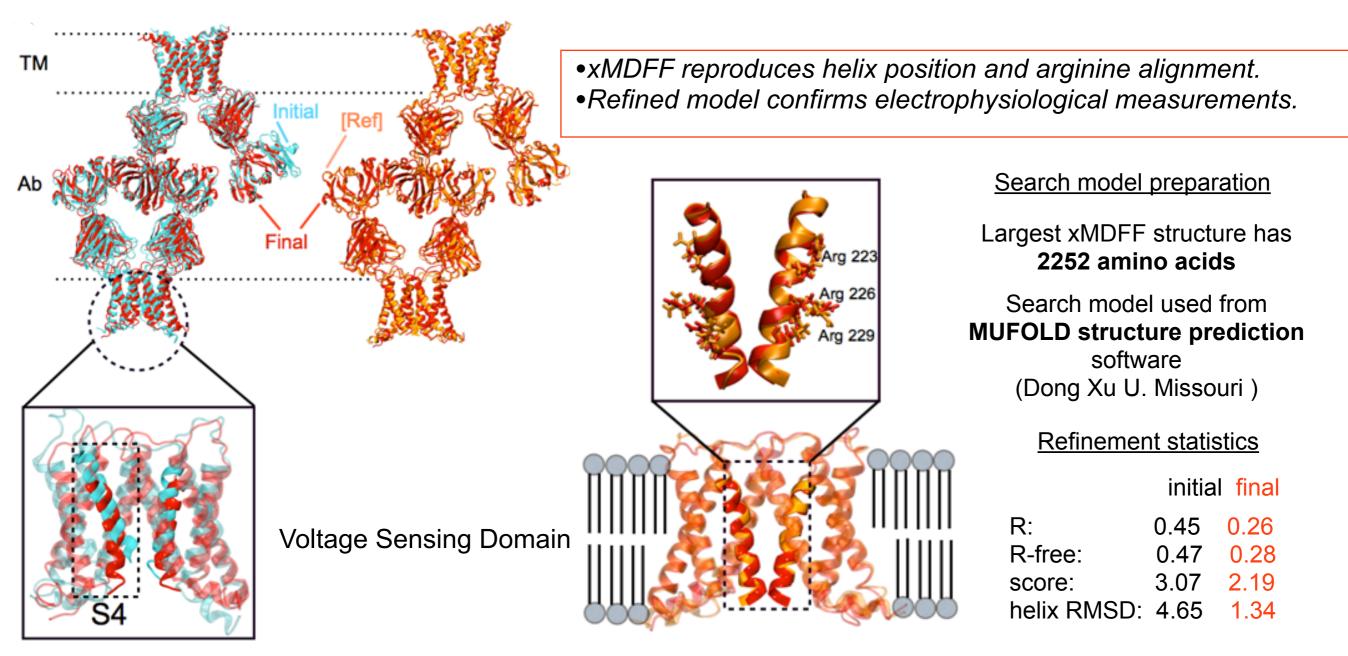


- Better R-work and R-free values than published before.
- Close R-work and R-free implies less over-fitting.
- Improved geometry implied by low Molprobity score.

McGreevy, Singharoy, et. al. Acta Cryst. D70 2344-2355, 2014.

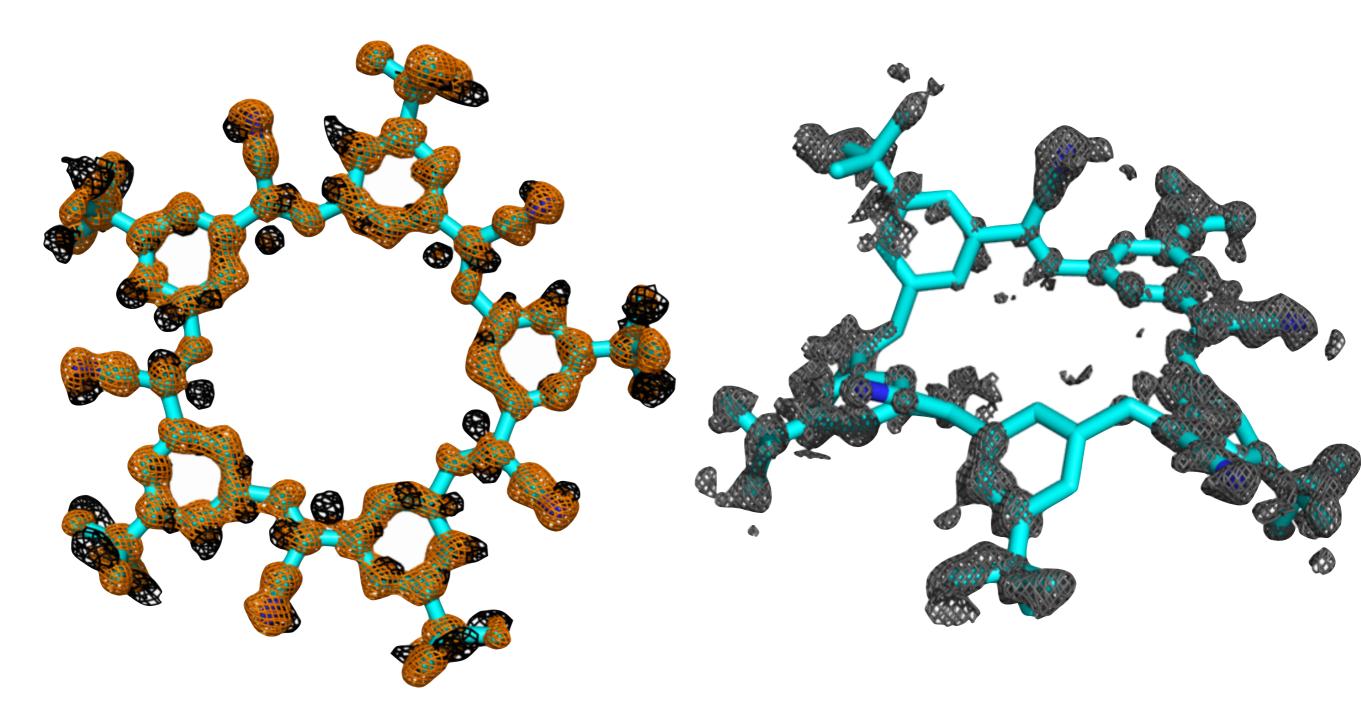
xMDFF Solves Voltage Sensor Protein Structure at 4 Å Resolution

Collaboration with E. Perozo (U. Chicago)



Qufei Li, et. al. Nature Structural & Molecular Biology, 21:244-252, 2014

xMDFF for Abiological Materials Cyanostar (2Å)



xMDFF-Phenix

(dual occupancy of CS shown in black and orange)

Phenix-only

Singharoy, et al. J. Am. Chem. Soc.137 (27), pp 8810–8818, 2015.

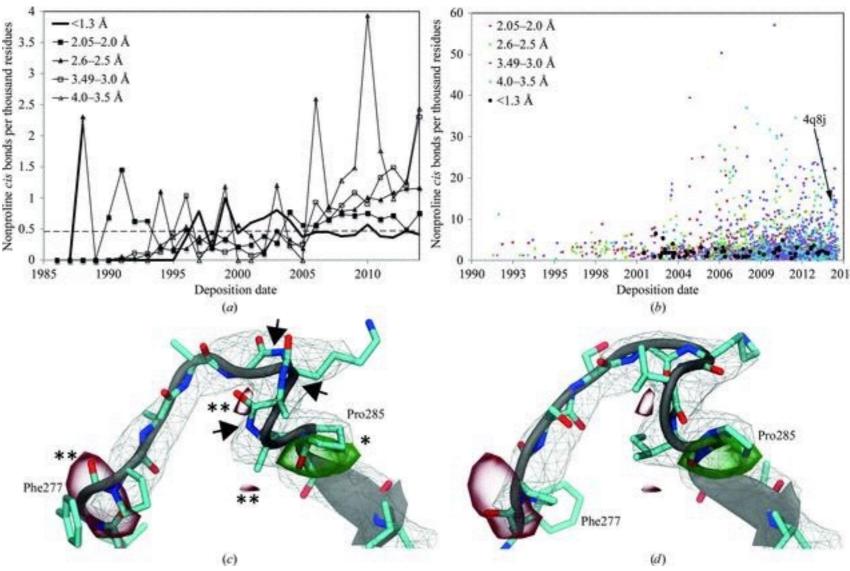
Importance of Checking Initial Structure

<0.05% non-proline bonds found in the cis conformation natively, however:

The frequency of non-proline cispeptide bond errors has been increasing for low-resolution

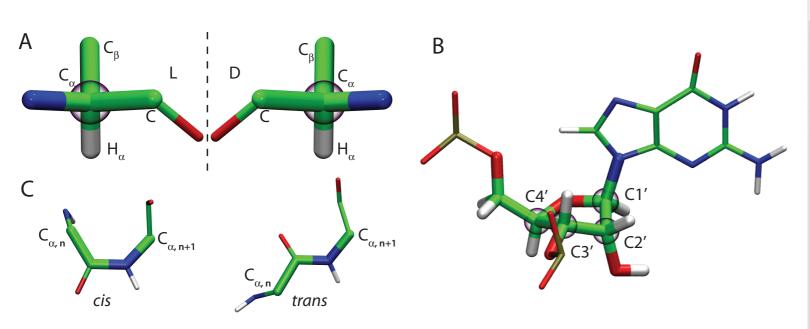
•

 These errors can hide issues in other parts of the structure



Tristan Croll. Acta Crystallographica D71, 706-709, 2015.

Structure Checking Plugins in VMD



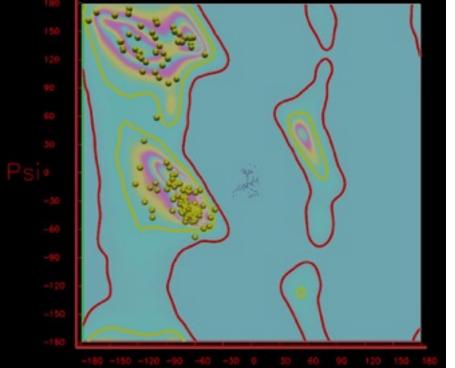
Eduard Schreiner, et al. BMC Bioinformatics, 12, 190, 2011

Wrong chirality, cis-peptide bonds, and torsion angle outliers may arise during modeling

•

- VMD provides tools to check, visualize, and correct these errors
- These tools, together with MD force fields, produce models with good structural geometry

Eile					
Initialise and upd	ate plots				
Molecule:	0	Selection:	all		
Initialise plot	plot torsions for current frame	Torsion to inspe	ect		0
	Reset Torsion	nPlot and remove	all visuaasabo	175	
Outlying and man	ginal residues				
Outliers (<0.02%			targinal (0.02-2	%)	
Chain I Segnam	e i Resid - i Score (per 10k re	esidues)	P P1	e I Resid 97	1 Score (#) 0,301
		A	P P1 P P1	98 159	0.454
	Show this residue			Show	this residue
	Select this residue for IMD			Select this	residue for IME
	Select this residue for IMDFF		:	Select this re	sidue for IMDF
Settings for intera		residues before a		ues after sel gstroms of se	ection elected strand.



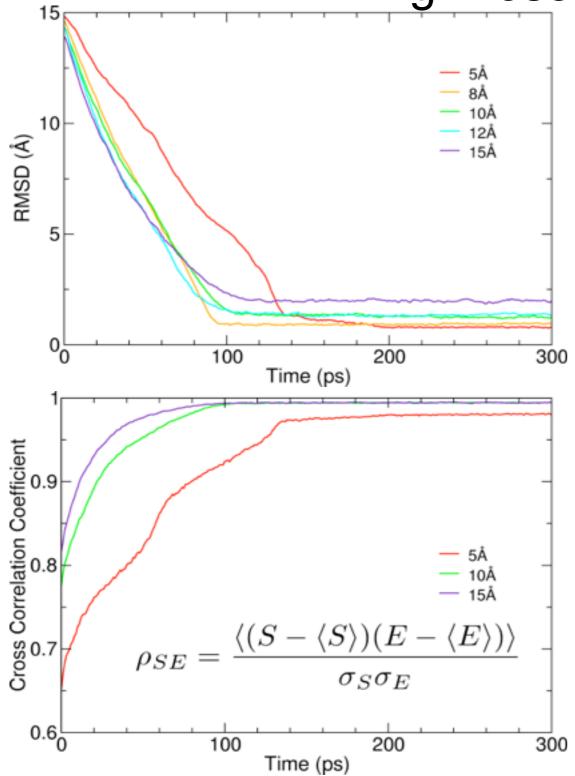
TorsionPlot Plugin new in VMD 1.9.3

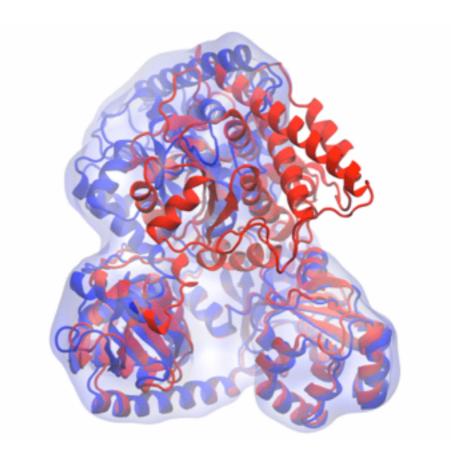
Cascade MDFF for high-resolution cryo-EM:				
Successively higher resolution maps				
5 Å Smo	othing	<u>3 Å Smc</u>	pothing	No Smoothing
Protocol	Global Cross- correlation	RMSD (Å)	Aile C	
Reference	0.732	-		
Direct	0.699	12.41		
Cascade	0.724	2.30		RPV1 (3.4 Å) D. Julius, Y. Cheng, Nature 504, 201

M. Liao, E. Cao, D. Julius, Y. Cheng, Nature 504, 2013.

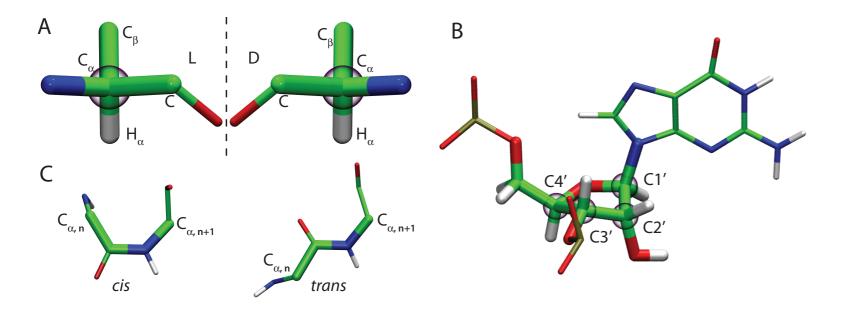
Analyzing MDFF Model Quality 0: Known Structures

MDFF has been validated against a wide-ranging set of known high-resolution structures





Analyzing MDFF Model Quality 1: Structure Checking



Eduard Schreiner, et al. BMC Bioinformatics, 12, 190, 2011

•

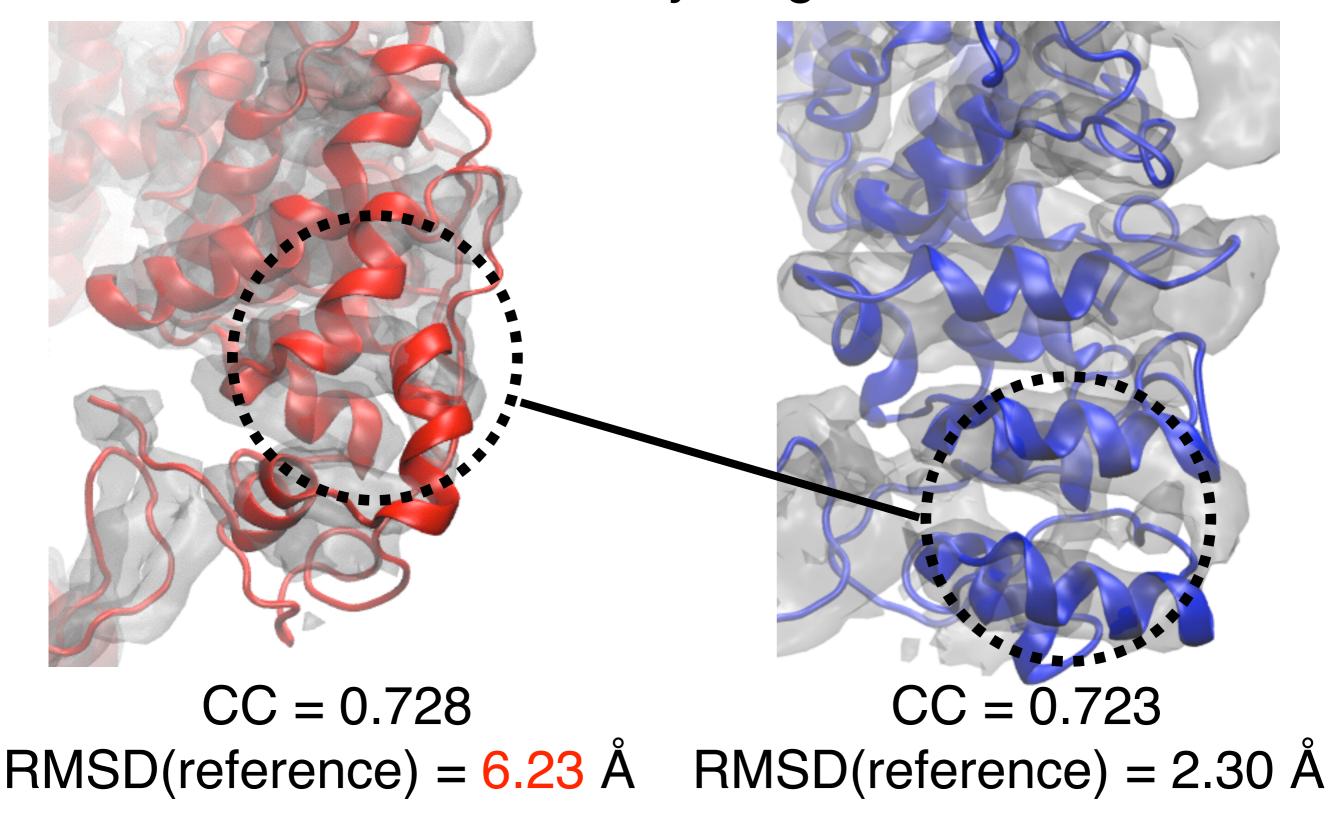
- Wrong chirality, cis-peptide bonds, and torsion angle outliers may arise during modeling
- VMD provides tools to check, visualize, and correct these errors
- These tools, together with MD force fields, produce models with good structural geometry

Ryan McGreevy*, Abhishek Singharoy*,et al. Acta Crystallographica D70, 2344-2355, 2014

<u>xMDFF refined</u> <u>structures</u>

PDB ID	Molprobity initial (published)	final
1AV1	3.72	1.94
1YE1	2.68	1.89
1JL4	3.24	1.47
1AOS	3.40	2.45
1XDV	2.87	2.01
1YI5	3.08	1.73

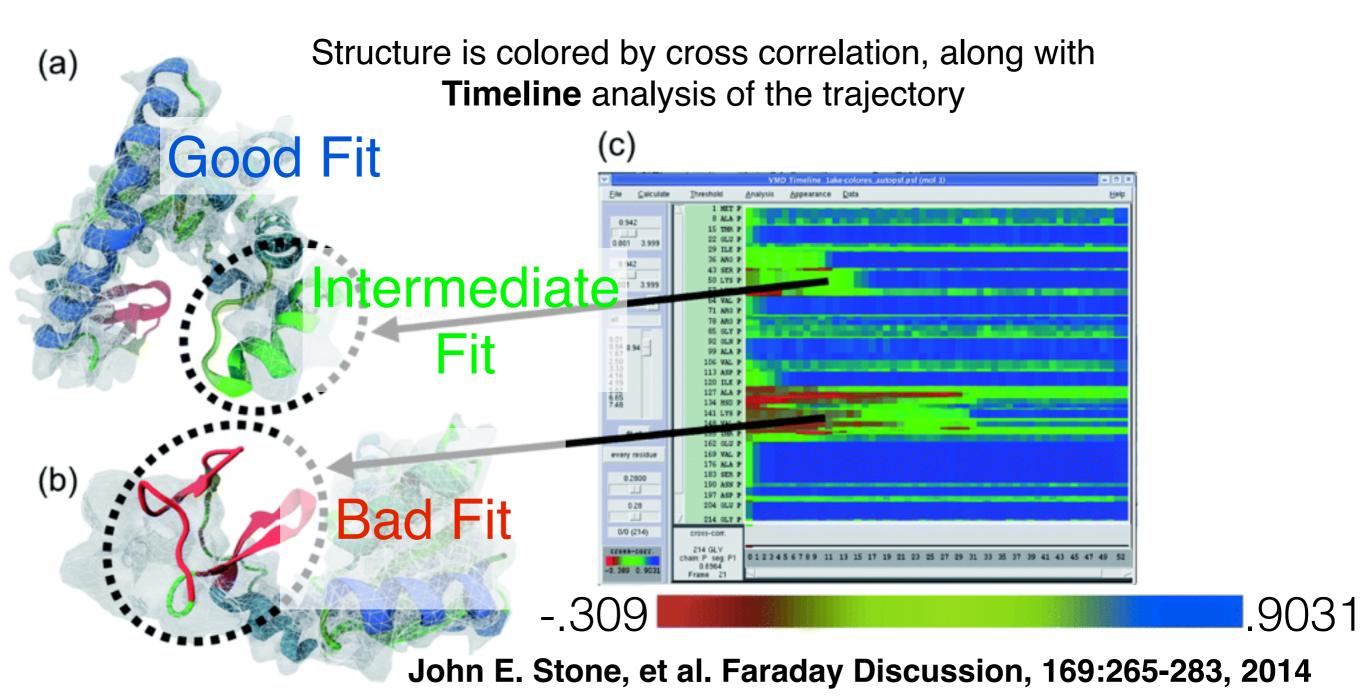
Analyzing MDFF Model Quality 2: **Global Cross Correlation** Global CC is not always a good indicator of fit



TRPVI EM map and structure from M. Liao, E. Cao, D. Julius, Y. Cheng, Nature 504, 2013.

Analyzing MDFF Model Quality 2: Local Cross Correlation

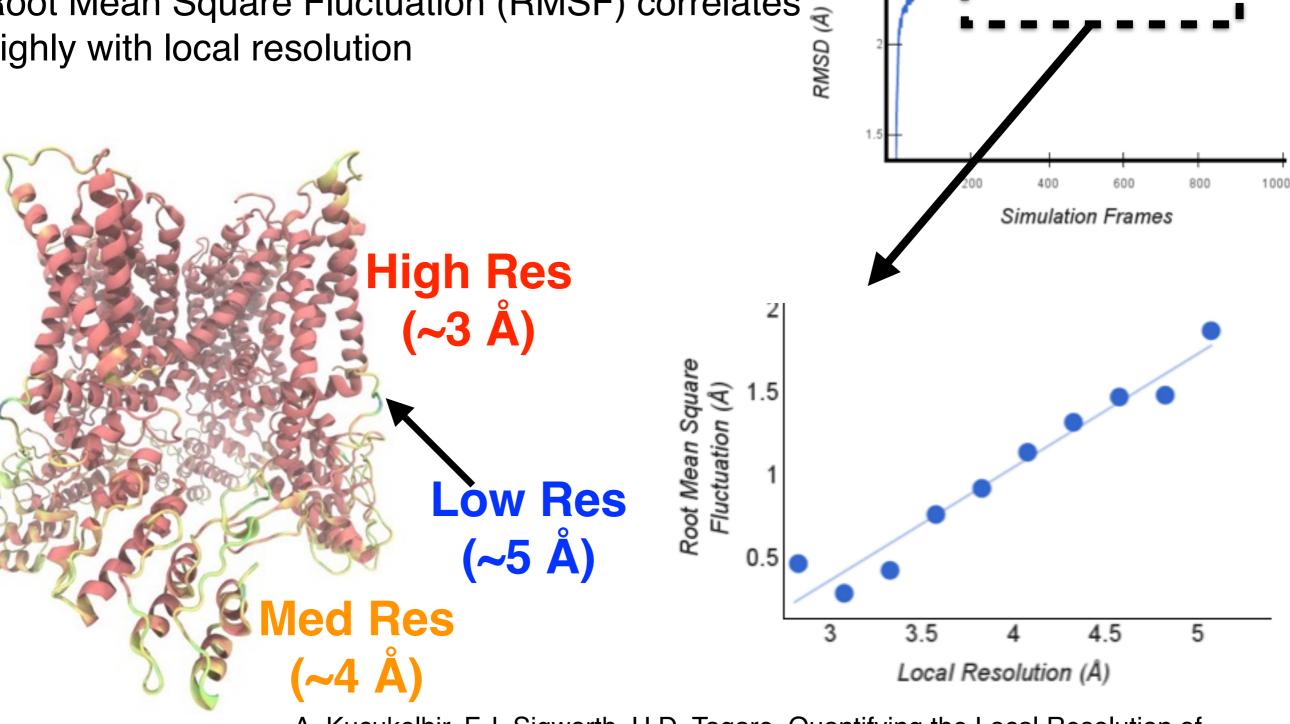
- Local cross correlation indicates quality of fit of specific regions across the entire structure
- New parallel CPU and GPU algorithms provide significant speed up (25-50x speedup over Chimera), allowing for fast computation along fitting trajectories



Analyzing MDFF Model Quality 3: Local Resolution Analysis

Local resolution of the experimental density from ResMap for error analysis and simulation parameterization RMSD (Å) During MDFF

Root Mean Square Fluctuation (RMSF) correlates highly with local resolution

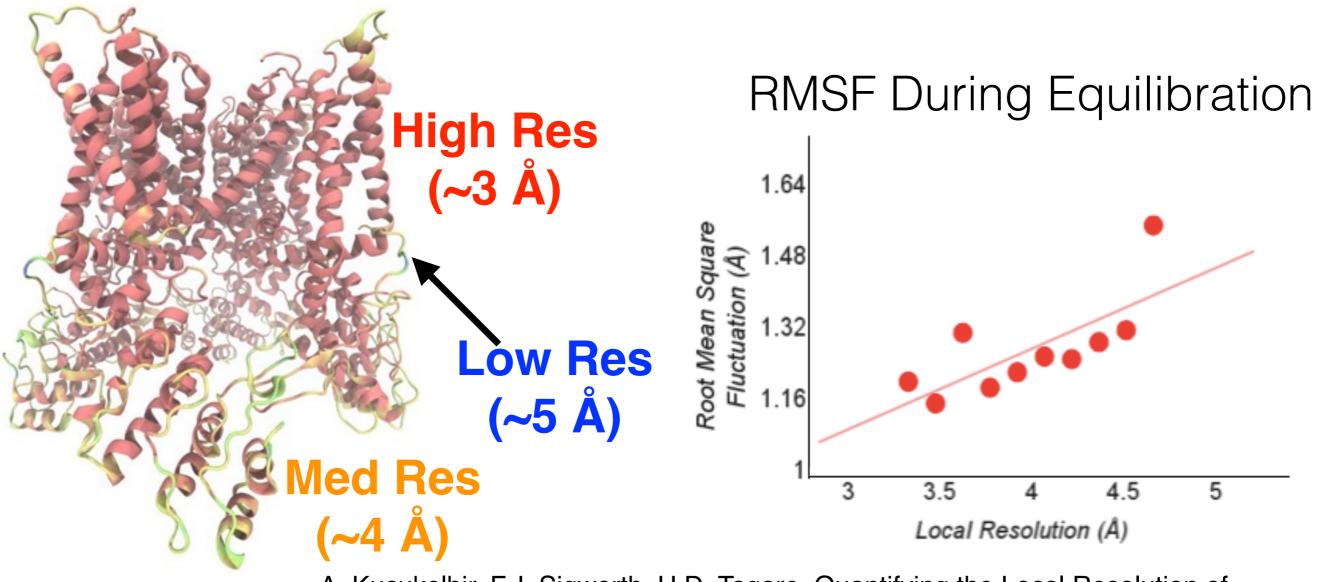


A. Kucukelbir, F.J. Sigworth, H.D. Tagare, Quantifying the Local Resolution of Cryo-EM Density Maps, Nature Methods, Volume 11, Issue 1, Pages 63-65, 2014.

Analyzing MDFF Model Quality 3: Local Resolution Analysis

Local resolution of the experimental density from ResMap for error analysis and simulation parameterization

Root Mean Square Fluctuation (RMSF) correlates highly with local resolution



A. Kucukelbir, F.J. Sigworth, H.D. Tagare, Quantifying the Local Resolution of Cryo-EM Density Maps, Nature Methods, Volume 11, Issue 1, Pages 63-65, 2014.

Analyzing MDFF Model Quality 4: Cross-validation correlation

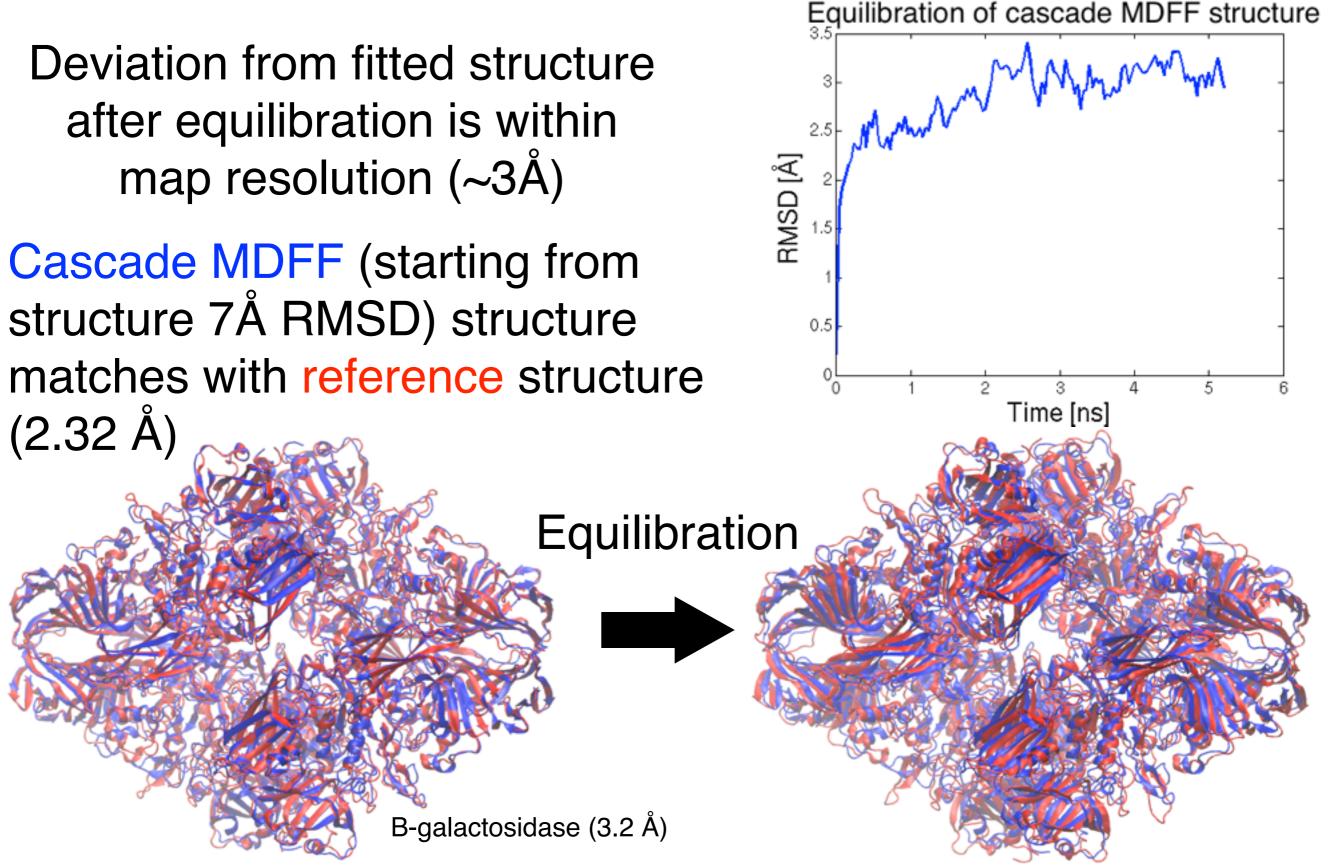
Cascade and direct fitting structure to one half map and calculating the cross correlation to the other

CC w.r.t. Fit to	Halfmap I	Halfmap II
Halfmap I	0.715 (<mark>0.686</mark>)	0.714 (<mark>0.685</mark>)
Halfmap II	0.716 (<mark>0.688</mark>)	0.716 (<mark>0.688</mark>)

CC of reference structure w.r.t. each half map was 0.719

TRPVI EM map and structure from M. Liao, E. Cao, D. Julius, Y. Cheng, Nature 504, 2013.

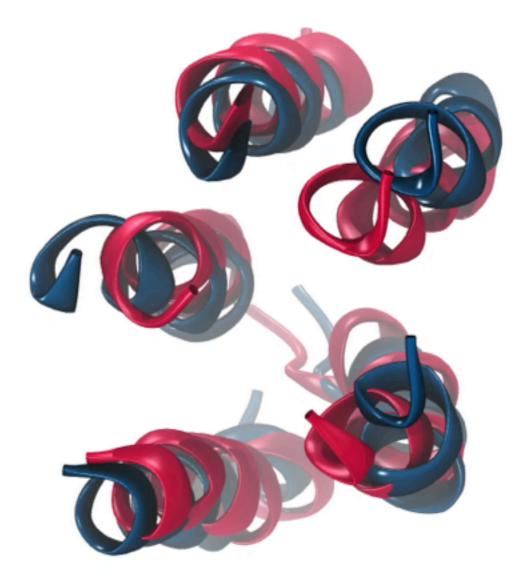
Analyzing MDFF Model Quality 5: **MD post-processing** Stability of structure during equilibration



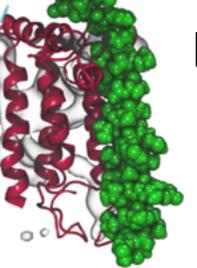
A. Bartesaghi, D. Matthies, S. Banerjee, A. Merk, S. Subramaniam, Proc. Natl. Acad. Sci. 111, 2014.

Analyzing MDFF Model Quality 6: Agreement with Experiment

Ribosome-bound structure predicted by MDFF from cryo-EM map ~ 7.5 Å



Beckmann, Schulten *et al.* eLife; 3:e03035 (2014)



Ribosome + Fo-c+YidC complex

Crystal Structure (3WVF) 3.2 Å Kumazaki *et al*. Nature (2014)

Nascent chain confirmed also by chemical cross-linking, gel filtration chromatography and mass spectroscopy.

MDFF Has a Wide Range of Applications

30S

Over 60 reported MDFF applications:

• By intramural Researchers:

Qufei Li et al. *Nat. Struct. Mol. Biol.* (2014): Structural mechanism of voltagesensing protein

Wickels et al. *eLife* (2014): Ribosomal insertase YidC
Zhao et al. *Nature* (2013): All-atom structure of HIV-1 Capsid
Agirrezabala et al. *PNAS* (2012): Ribosome translocation intermediates
Frauenfeld et al. *Nat. Struct. Mol. Biol*. (2011): SecYE ribosome complex

• By extramural Researchers:

Gogala et al. *Nature* (2014): Ribosome Sec61 complex Unverdorben et al. *PNAS* (2014): 26S proteasome Bharat et al. *PNAS* (2014): Tubular arrays of HIV-1 Gag Park et al. *Nature* (2014): SecY channel during initiation of protein translocation Hashem et al. *Nature* (2013): *Trypanosoma brucei* ribosome Becker et al. *Nature* (2012): Ribosome recycling complex Lasker et al. *PNAS* (2012): Proteasome Strunk et al. *Science* (2011): Ribosome assembly factors Wollmann et al. *Nature* (2011): Mot1–TBP complex Becker et al. *Nat. Struct. Mol. Biol.* (2011) Dom34–Hbs1 stalled ribosome complex^{YidC} Guo et al. *PNAS* (2011): RsgA GTPase on ribosomal subunit Armache et al. *PNAS* (2011): Translating eukaryotic ribosome

MDFF/xMDFF Methodological Articles:

Trabuco et al. *Structure* (2008) 16:673-683.

Trabuco et al. *Methods* (2009) 49:174-180.

Ryan McGreevy*, Abhishek Singharoy*, et al. Acta Crystallographica (2014) D70, 2344-2355 Abhishek Singharoy, et al. Journal of the American Chemical Society (2015). In Press.

8 Å resolution cryoEM density of ribosome-YidC complex. Wickels et al. *eLife.* 2014

NC

tRNA

50S

Acknowledgements and Further Information

Find out more about MDFF including:

- software downloads
- publications
- documentation
- tutorials

http://www.ks.uiuc.edu/Research/mdff/



Abhi Singharoy



Ivan Teo