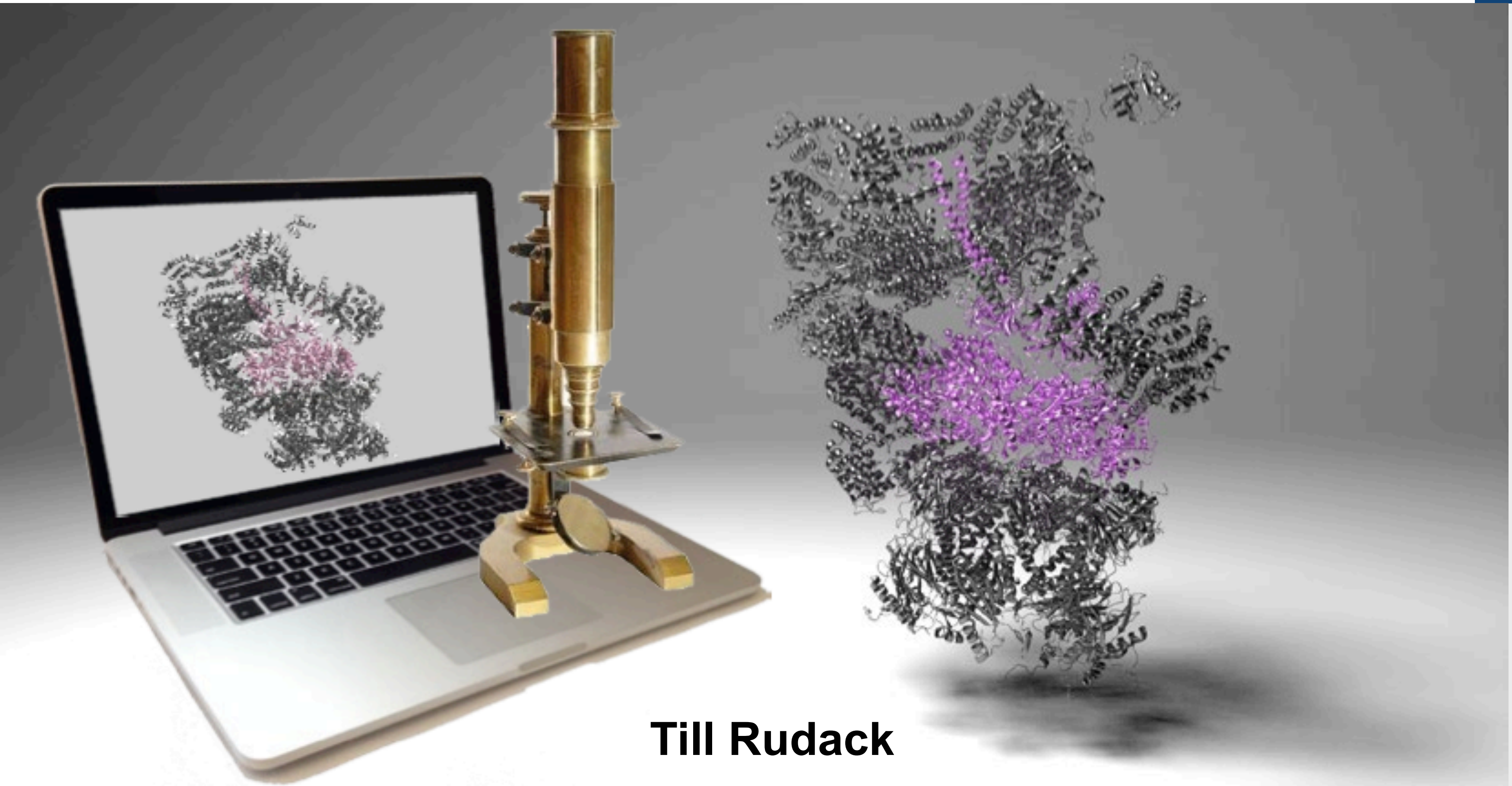
Integrative Modeling
Examples from Modern Research



Till Rudack

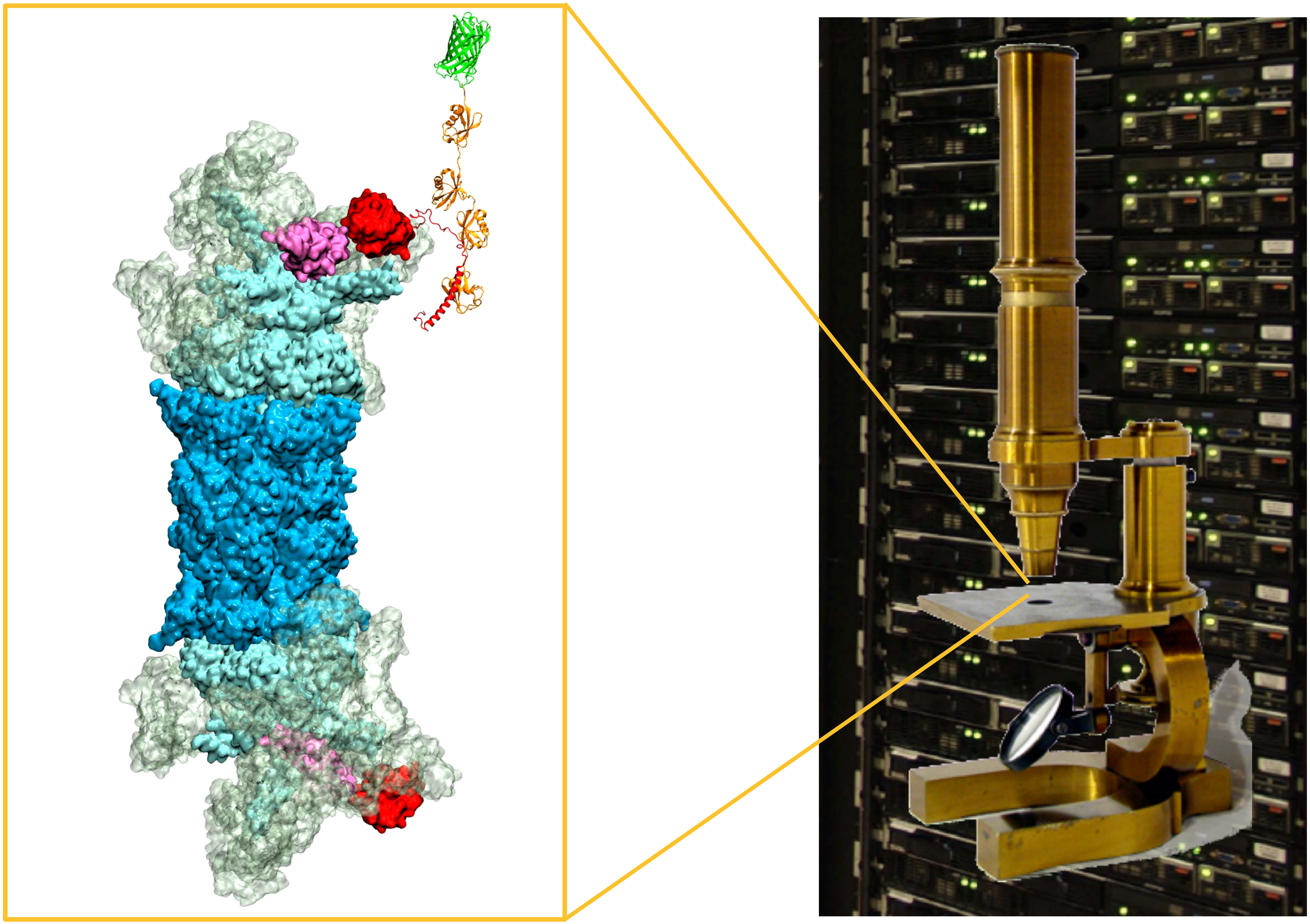
Klaus Schulten Group - Theoretical and Computational Biophysics Group

NIH Center for Macromolecular Modeling and Bioinformatics

University of Illinois at Urbana-Champaign

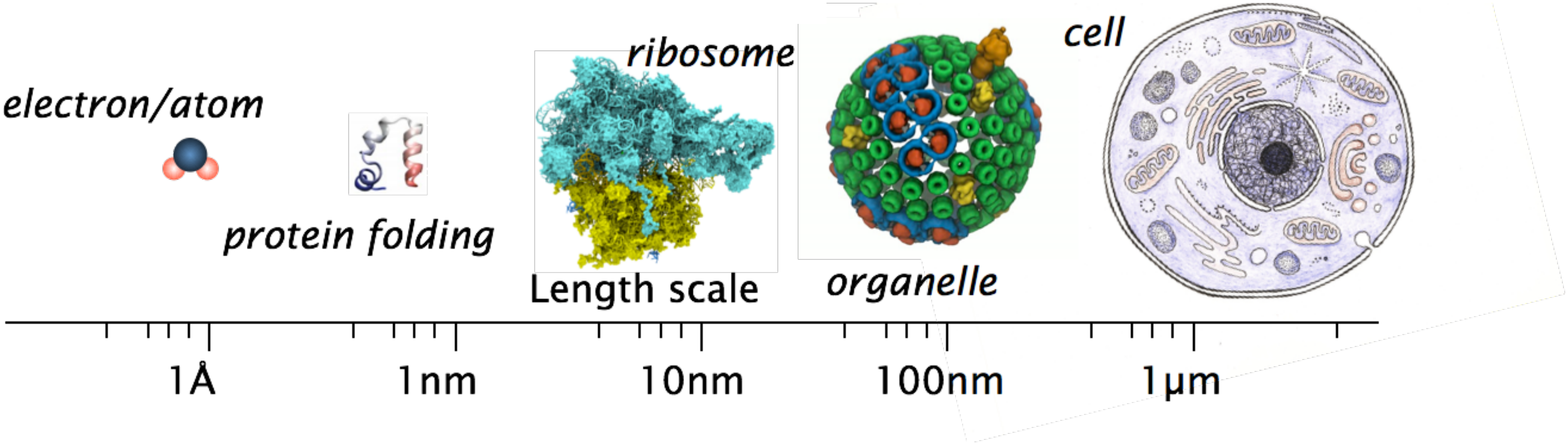
10/21/16

What can we discover with the Computational Microscope?

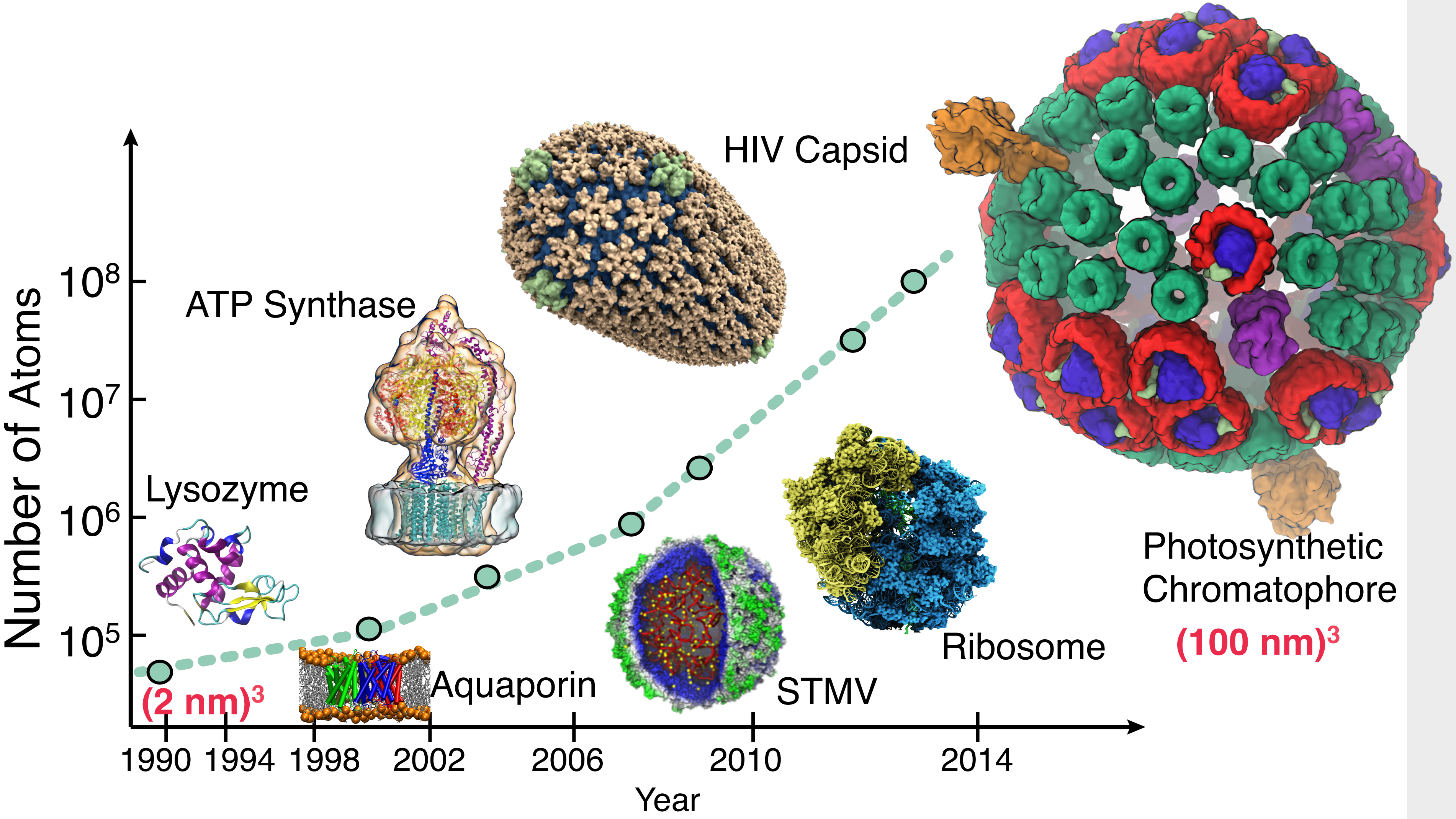


The range of the Computational Microscope

… Views Living Systems from Electron to Cell



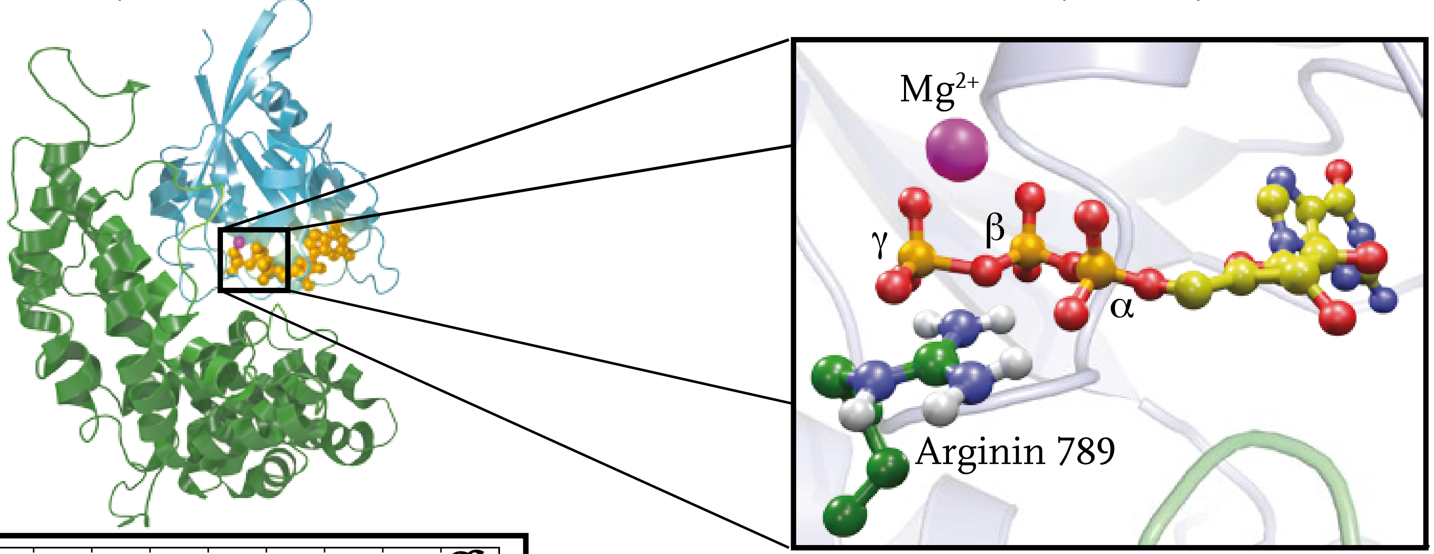
Size Matters!

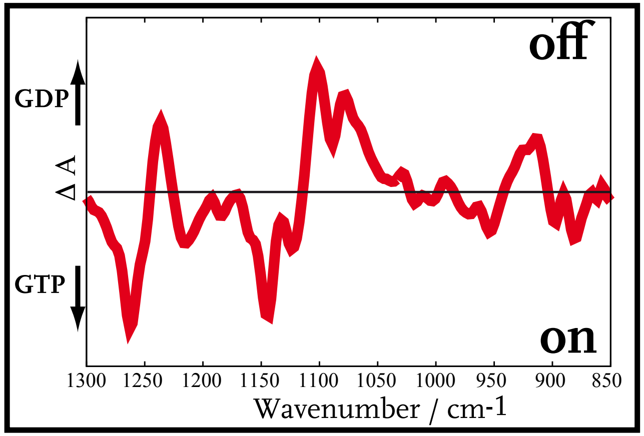


Detail Matters too!

Experiment
X-ray Crystallography
(1.6-2.5 Å)

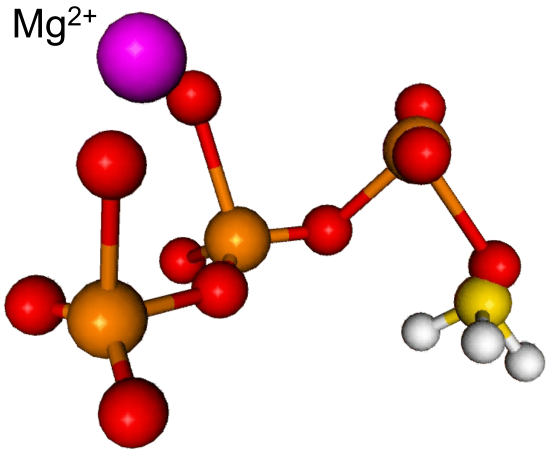
Theory
MD Simulations
(0,01 Å)





FTIR-Spectroskopy ($\Delta\nu = 1.0$ cm^{-1}; 10^{-4} Å)

⟷



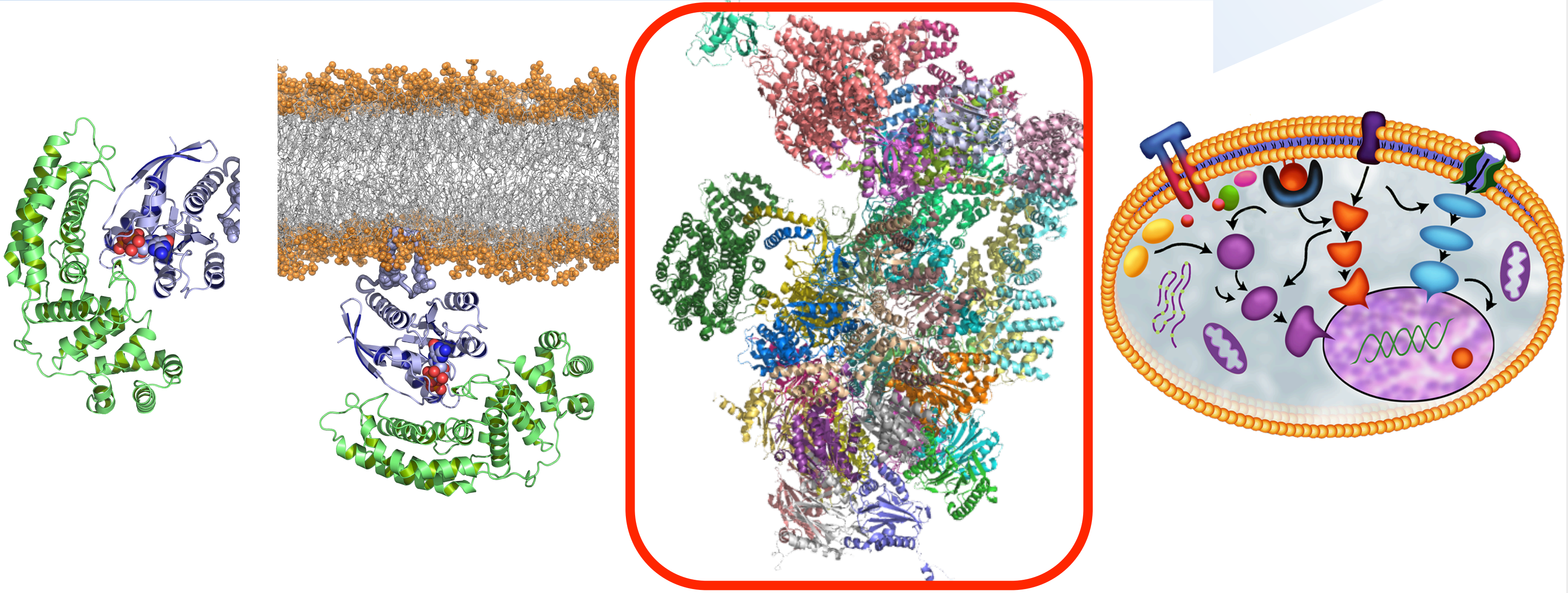
Application of MD simulations: Ras at Membrane

Protein In Solvent

Protein + Membrane

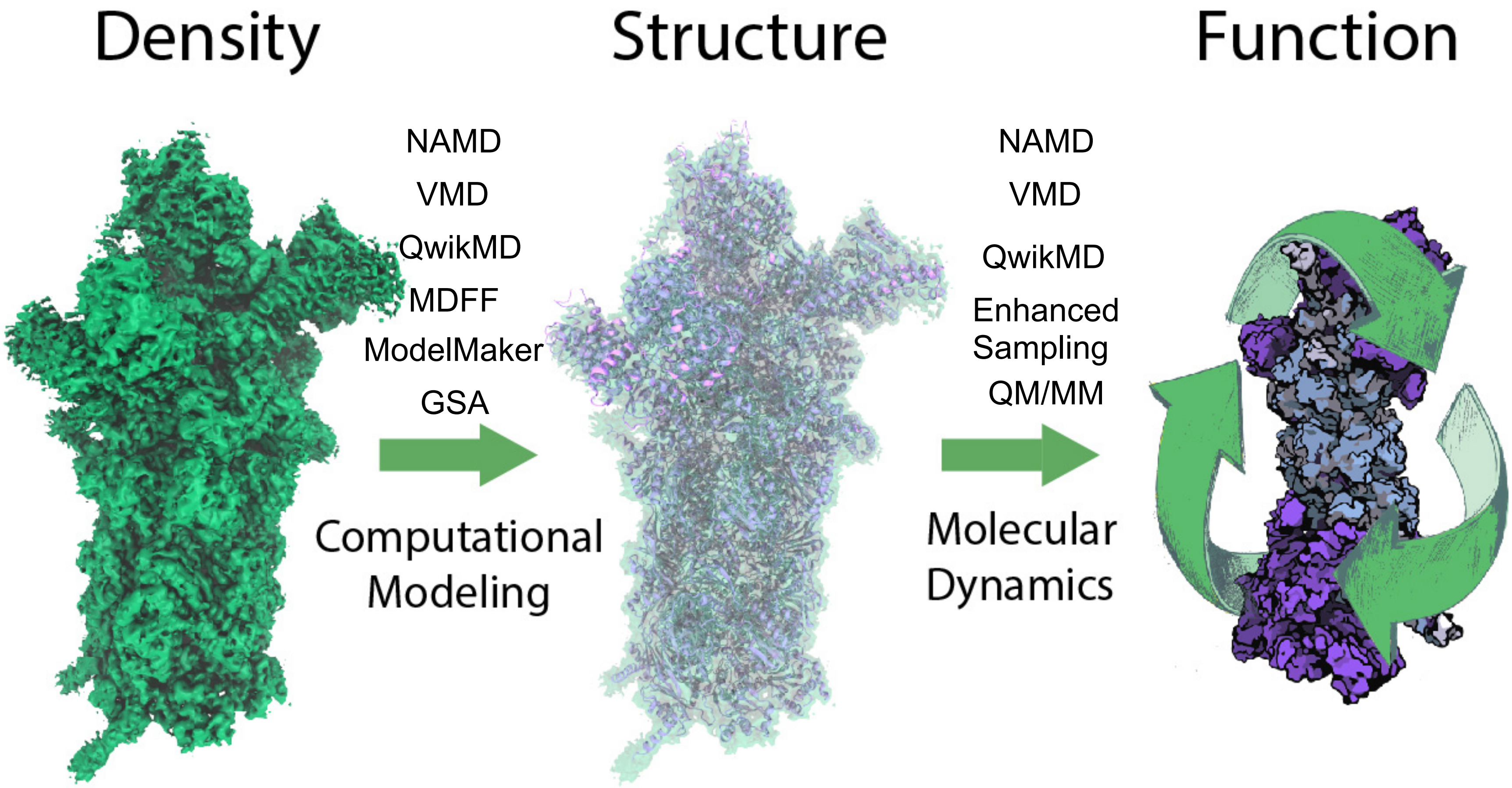
Multi-Protein Complex

Cell



Molecular dynamics simulations connect **function** and **dynamics** to **structural data** from diverse **experimental sources** to investigate critical cellular processes occurring at the **sub-Ångstrom** level up to the **macromolecular** level.

The Key Strategy for Discoveries



Density

Structure

Function

NAMD
VMD
QwikMD
MDFF
ModelMaker
GSA

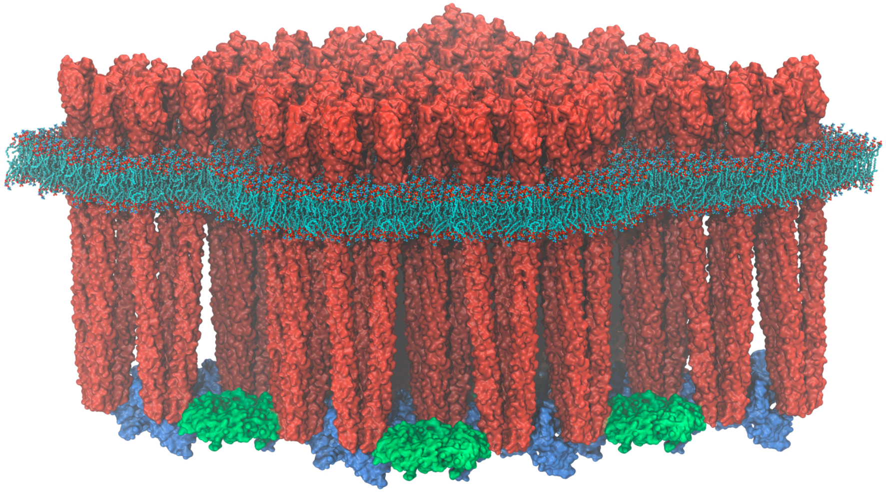
Computational Modeling

NAMD
VMD
QwikMD
Enhanced Sampling
QM/MM

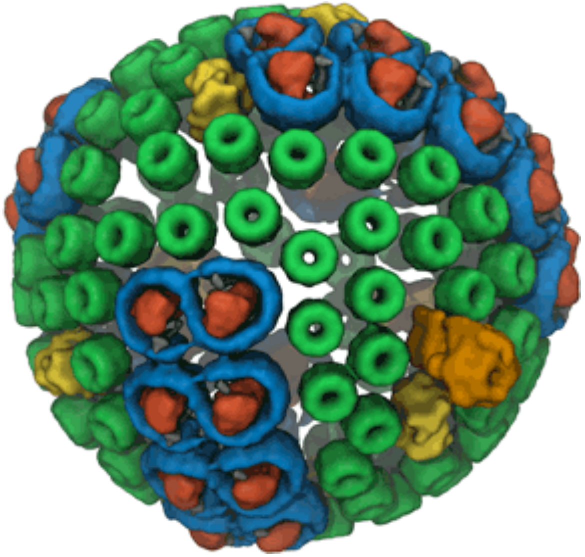
Molecular Dynamics

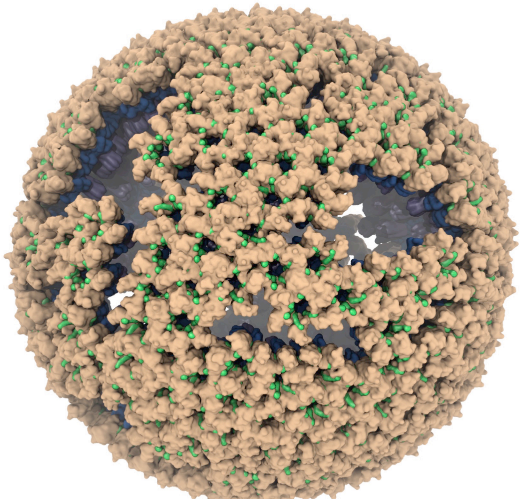
A Sampling of TCBG‘s MDFF Projects

Chemosensory Array

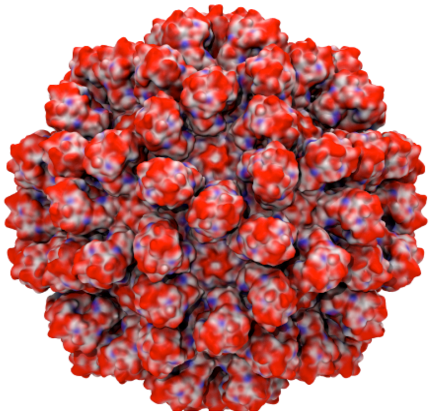


Chromatophore





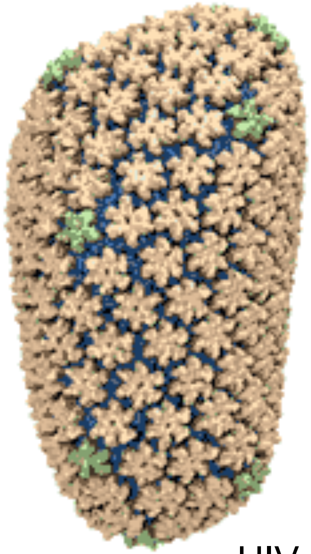
Rous Sarcoma Virus



Theoretic

Univer

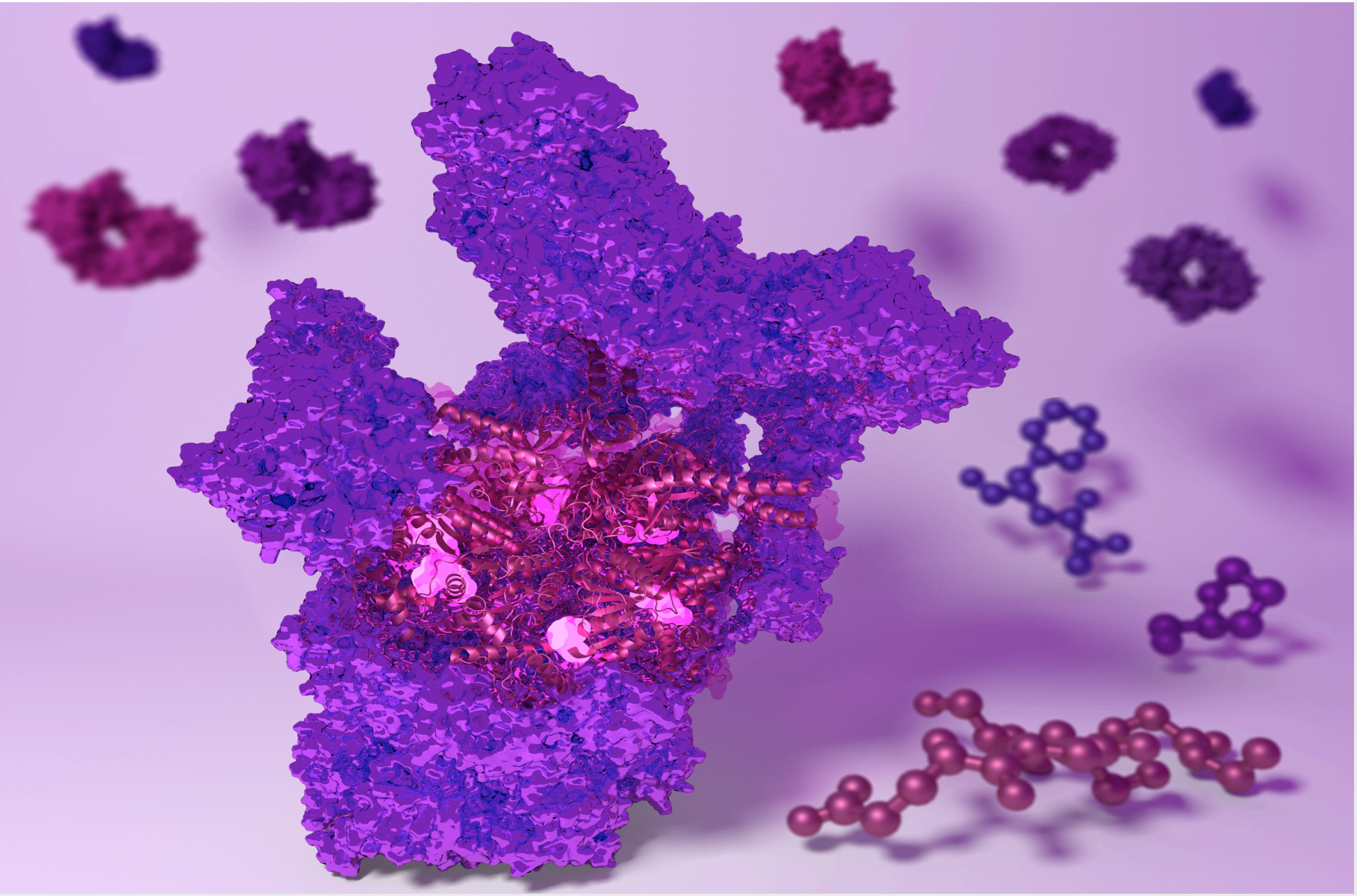
Rabbit Hemorrhagic Disease



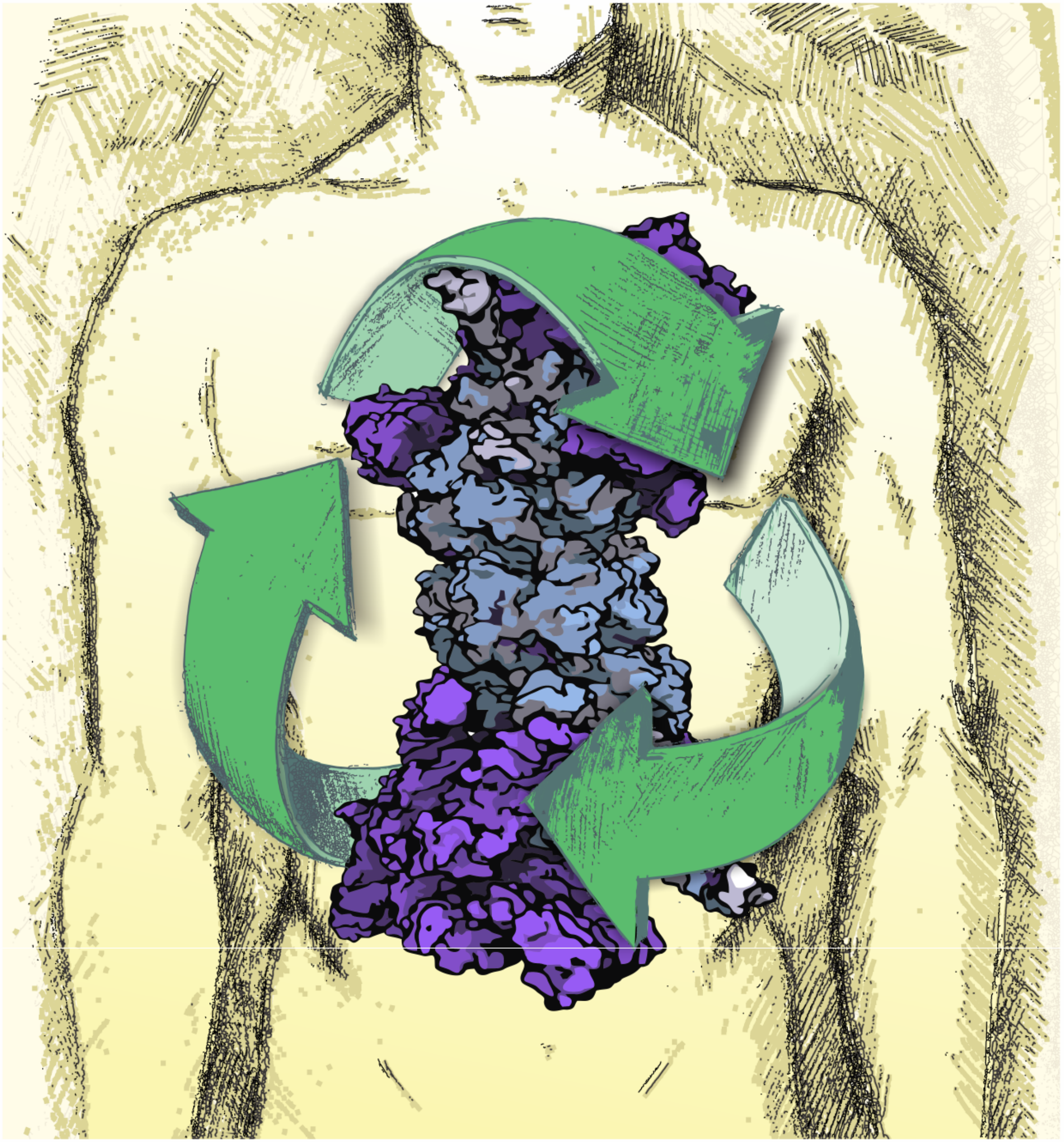
HIV

26S Proteasome

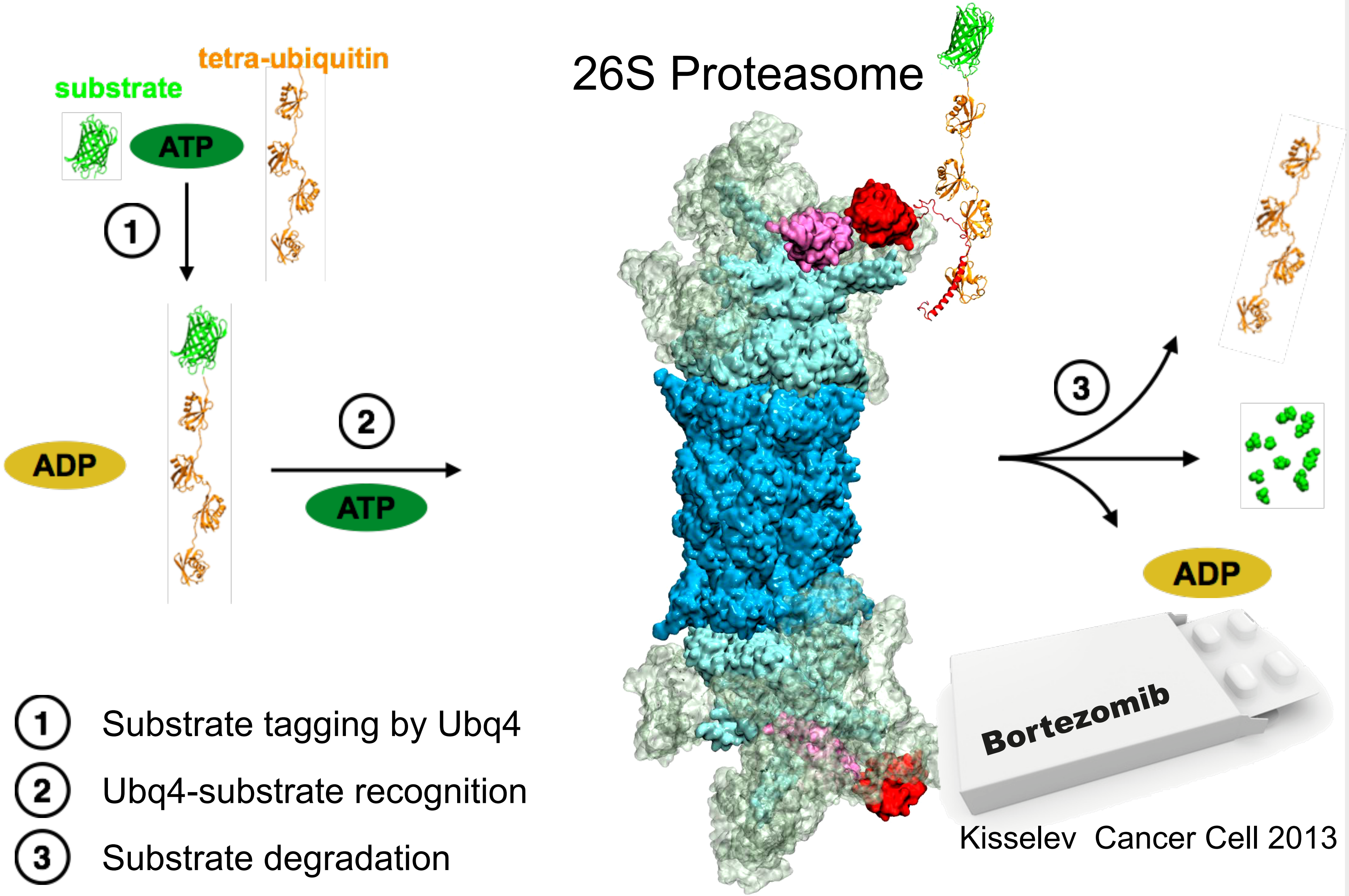
Integrating experimental methods into computational modeling



The Receycling System of the Cell



The ubiquitin proteasome proteolytic pathway



Kisselev Cancer Cell 2013

① Substrate tagging by Ubq4

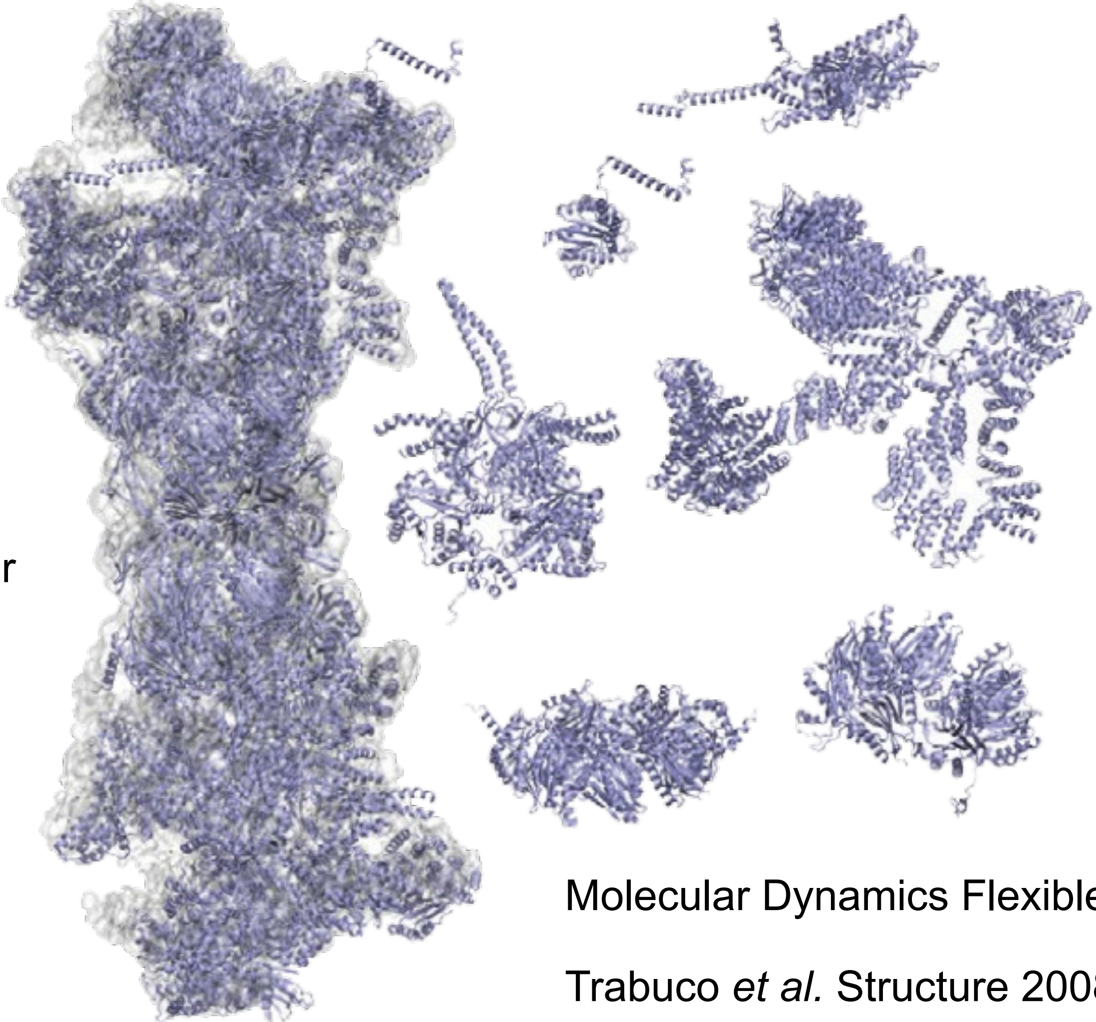
② Ubq4-substrate recognition

③ Substrate degradation

Near-atomic model of the 26S proteasome

Cryo-EM density

Subunits from X-ray crystalography, NMR, and homology modeling



max planck institute
of biochemistry

Wolfgang Baumeister
Friedrich Foerster

PDB-ID 4CR2

EMDB-ID 2594

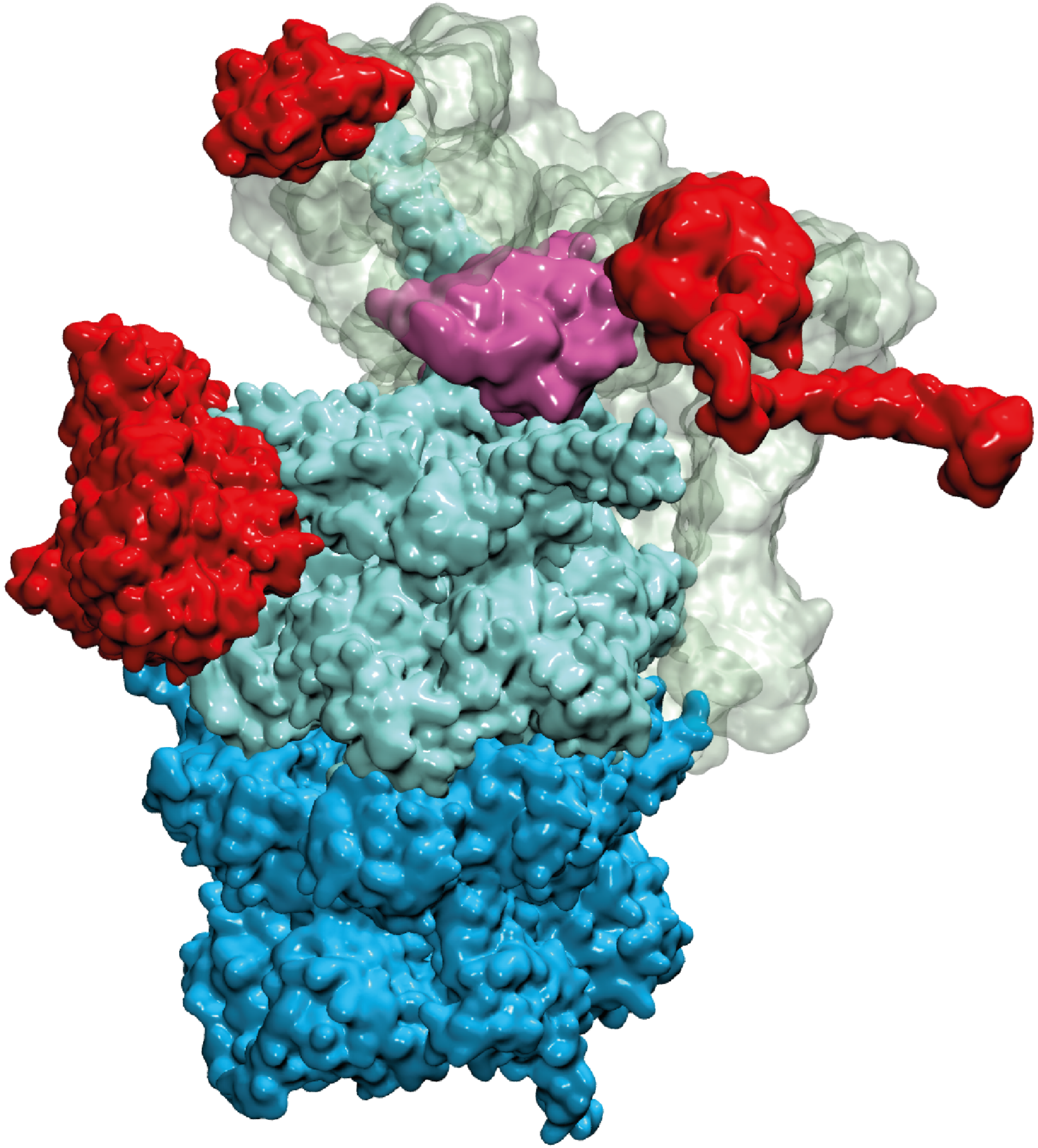
Resolution 7.7 Å

Unverdorben *et al.* PNAS 2014

Molecular Dynamics Flexible Fitting (MDFF):

Trabuco *et al.* Structure 2008

Functional subunits of the 26S proteasome



Ubiquitin
Recognition
(Rpn10, Rpn13, Rpn1)

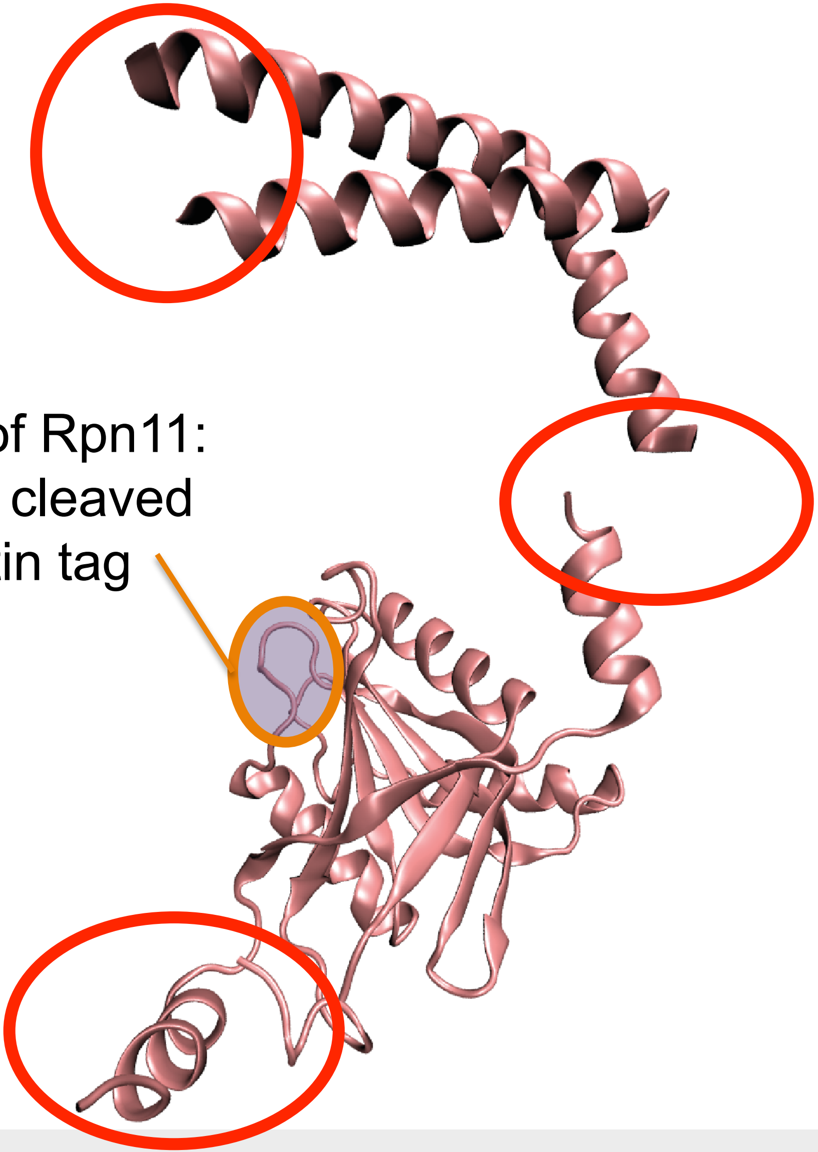
Deubiquitylation
(Rpn11)

Substrate
Unfolding
(ATPase-ring)

Substrate
Degradation
(α-ring, β-ring)

Deubiquitylation subunit: Rpn11

Complete models are a basic prerequisite to **perform MD simulations**



Active site of Rpn11: substrate is cleaved from ubiquitin tag

Deubiquitylation (Rpn11)

Missing segments

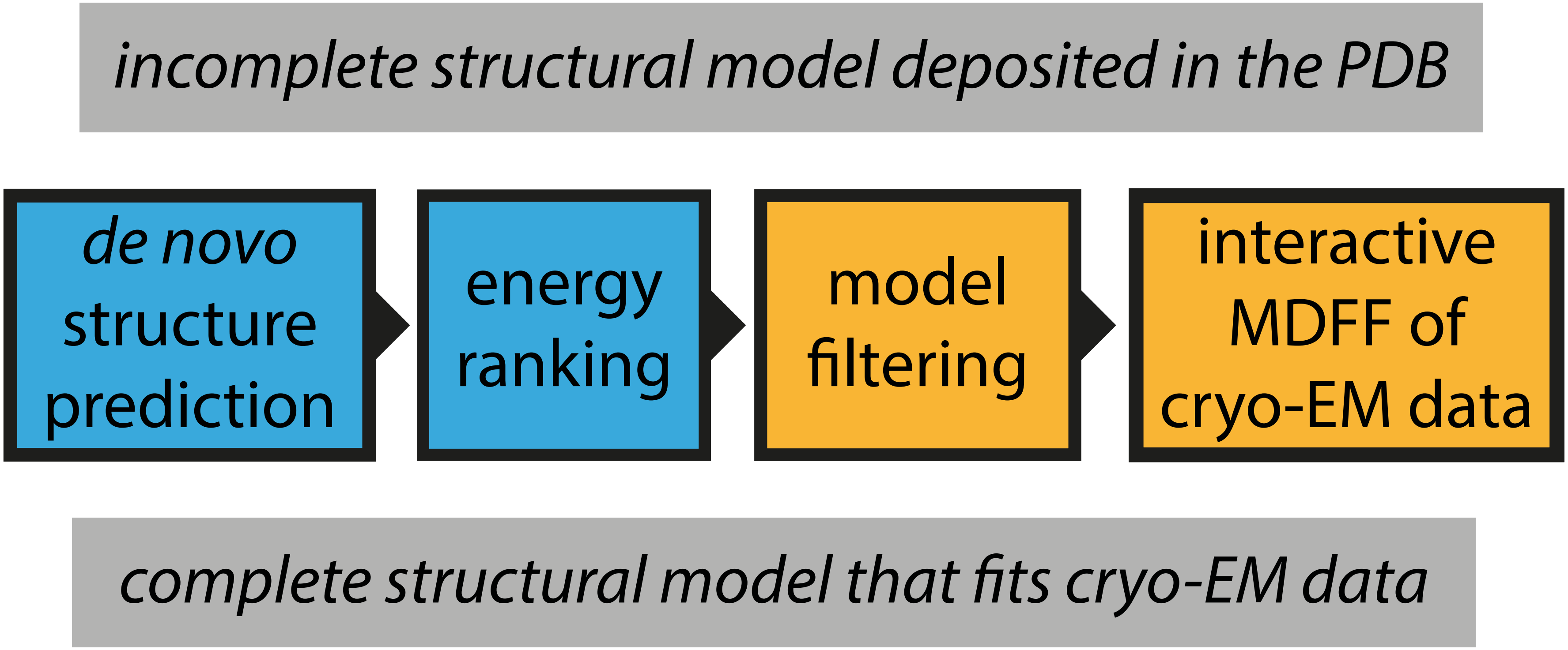
- highly flexible

- ambigous density

Chain V of PDB-ID 4CR2

Unverdorben *et al.* PNAS 2014

Combining Rosetta and MDFF through VMD



incomplete structural model deposited in the PDB

de novo structure prediction

energy ranking

model filtering

interactive MDFF of cryo-EM data

complete structural model that fits cryo-EM data

Rosetta

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

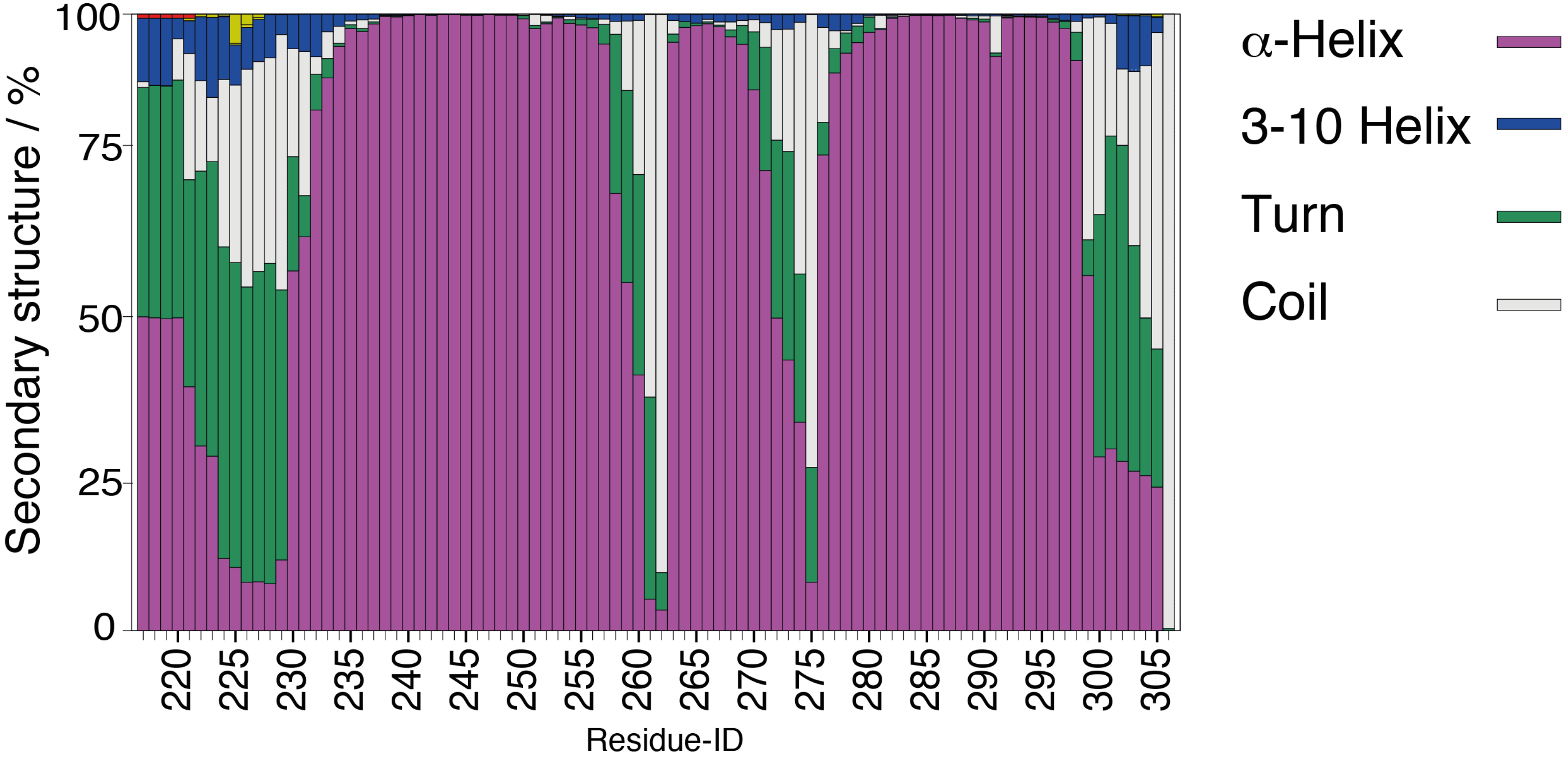
VMD/NAMD

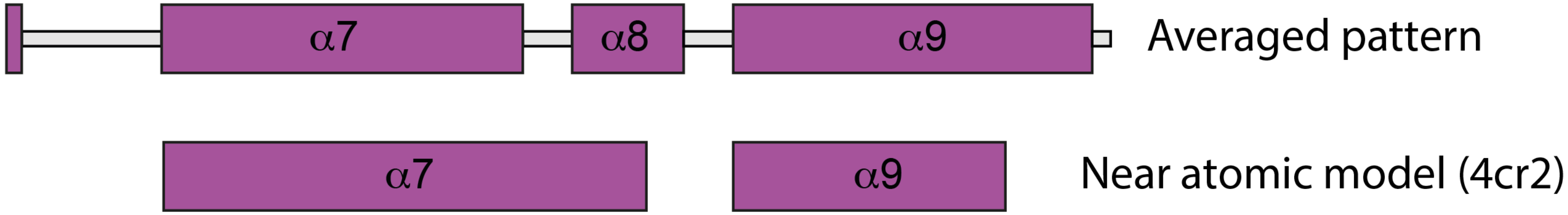
Humphrey *et al.* J. Mol. Graph. 1996
Philips *et al.* J. Comput. Chem. 2005

Integrating user expertise into *de novo* structure prediction

Model filtering by secondary structure

Secondary structure histogram of predicted ensembles of Rpn11's C-terminal tail

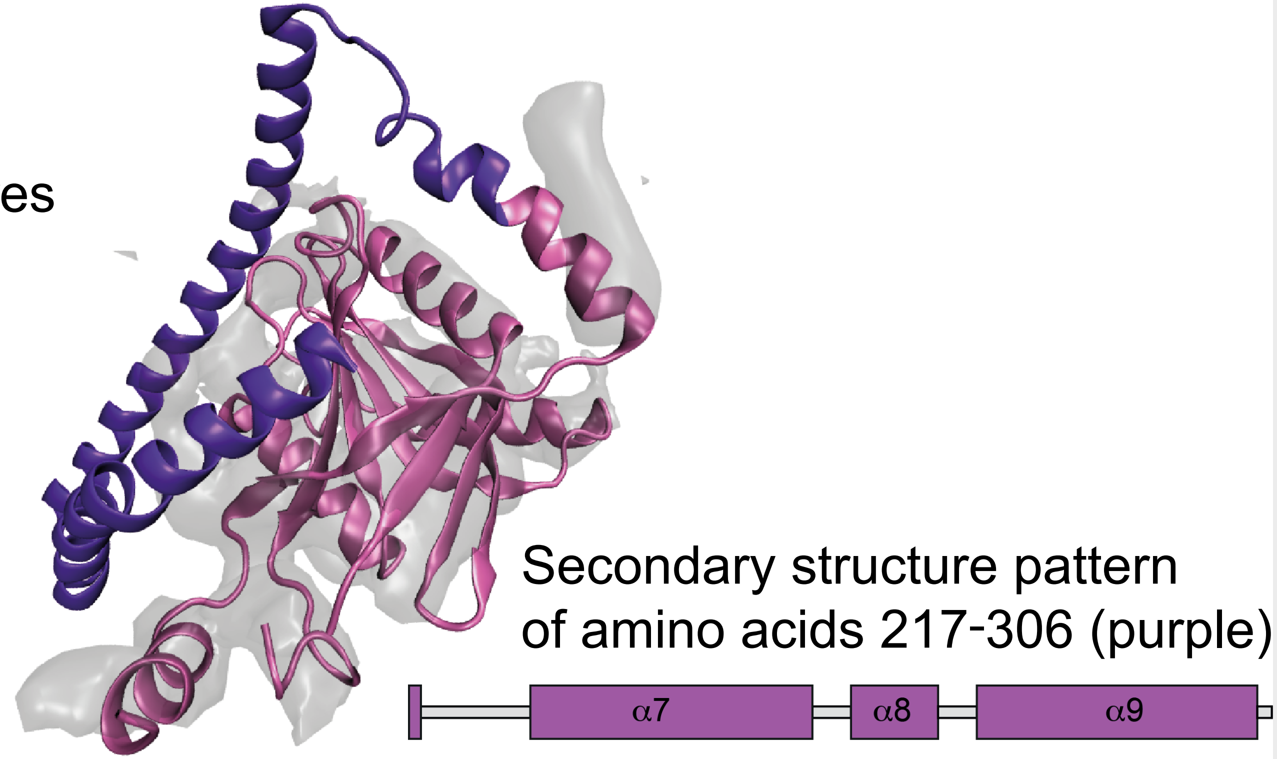




Predicted model

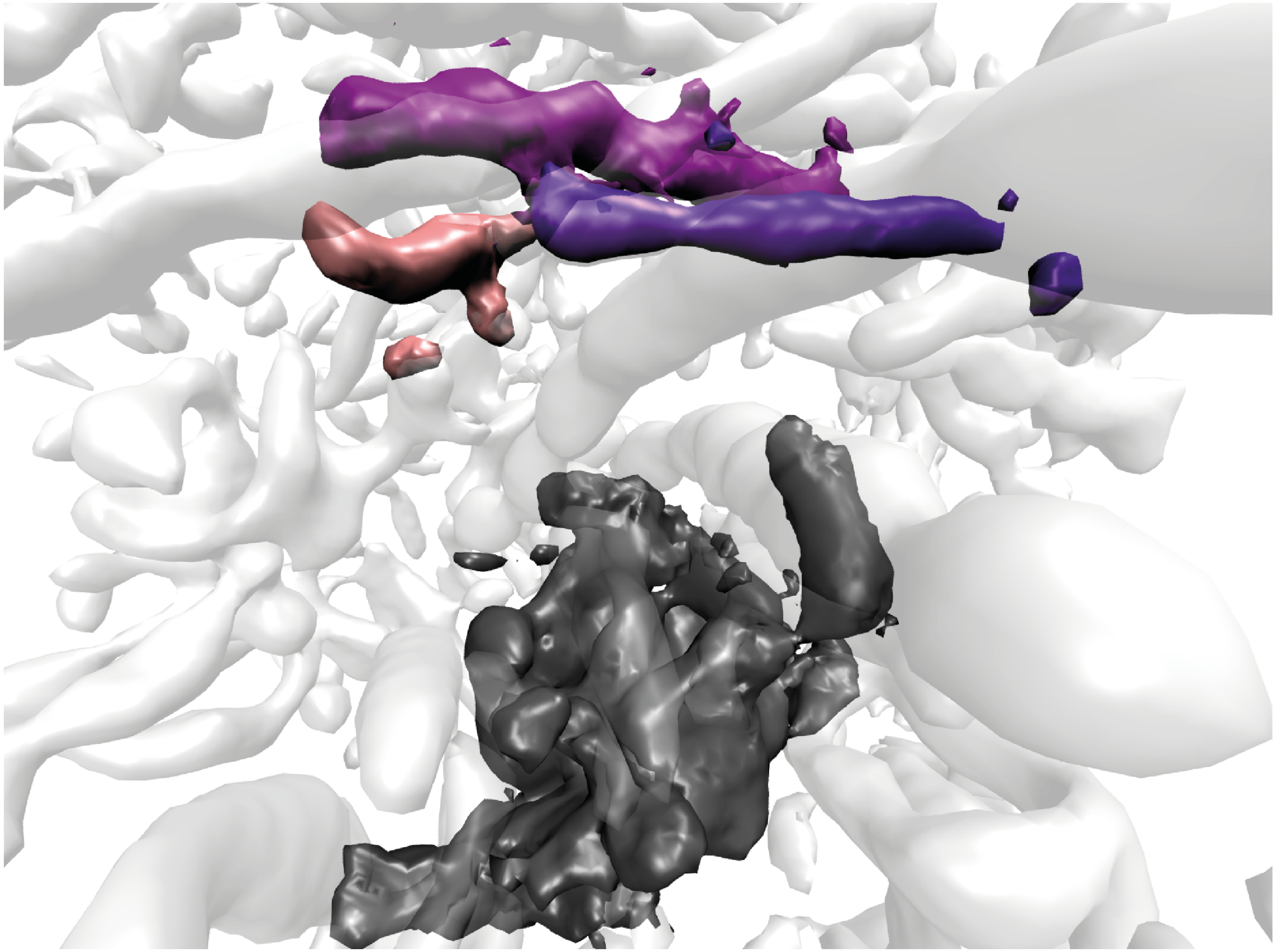
Represenative model of the
predicted averaged secondary structure pattern
for Rpn11‘s C-terminal tail (purple)

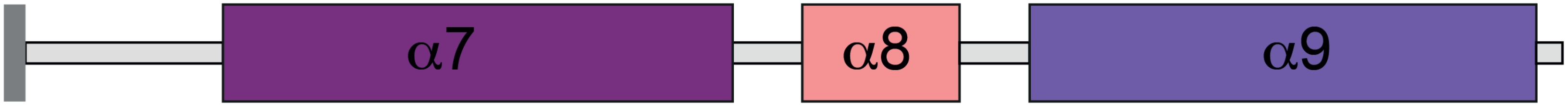
Rosetta tends to
build compact structures



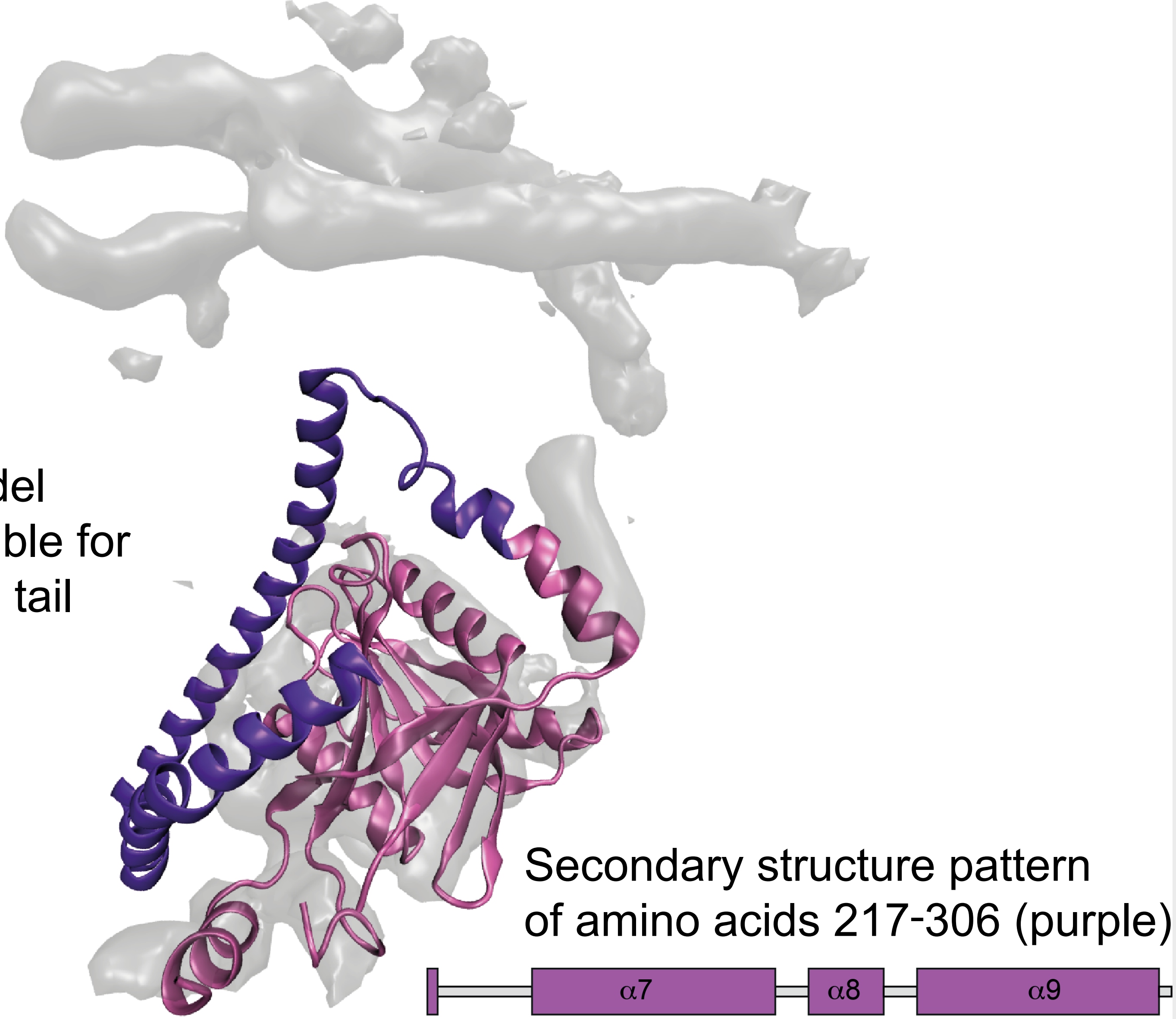
Secondary structure pattern
of amino acids 217-306 (purple)

Visual inspection of cryo-EM density





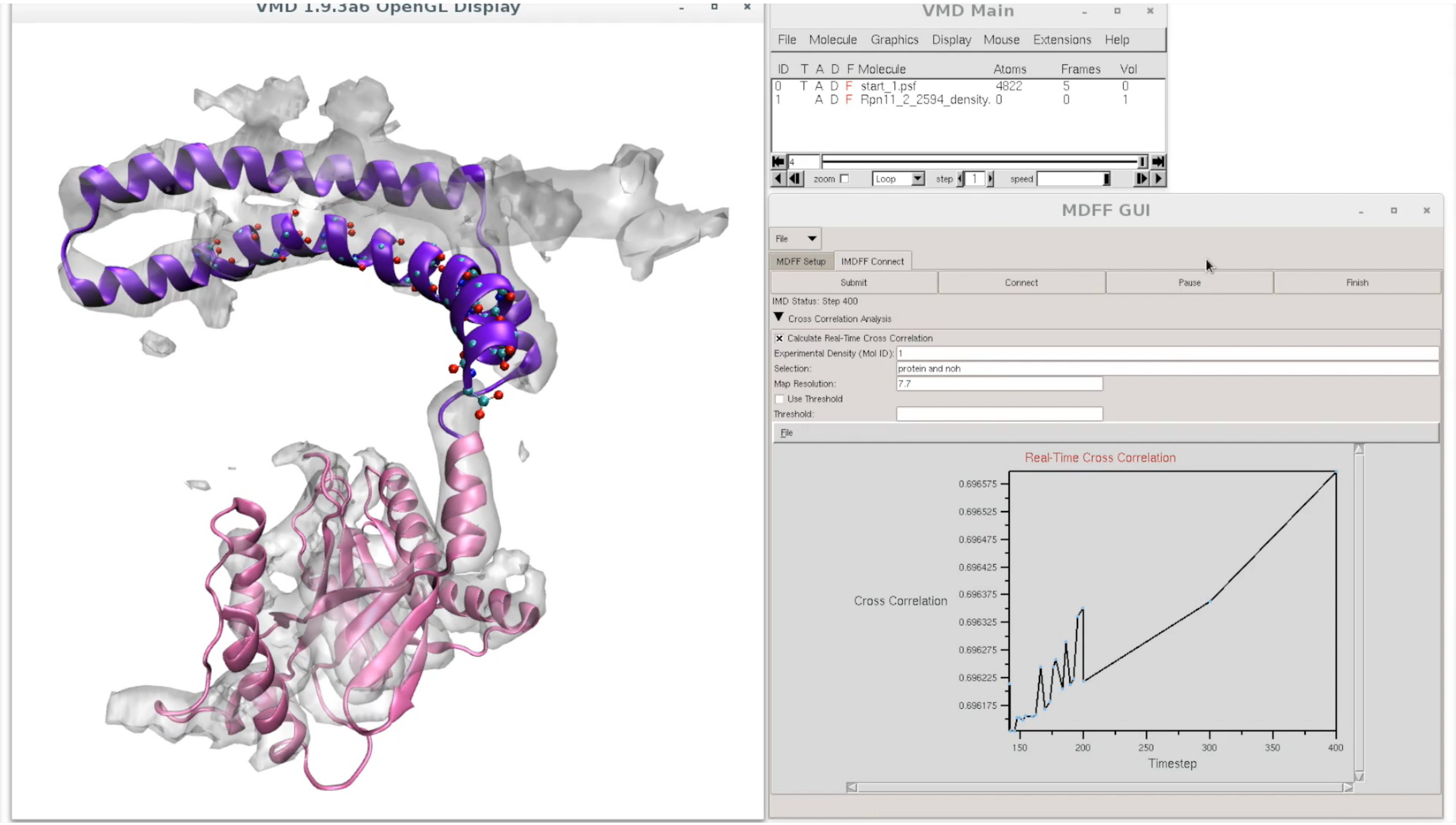
Predicted model to initiate MDFF



Represenative model
of predicted ensemble for
Rpn11‘s C-terminal tail

Secondary structure pattern
of amino acids 217-306 (purple)

Interactive Molecular Dynamics Flexible Fitting



ID	T	A	D	F	Molecule	Atoms	Frames	Vol
0	T	A	D	F	start_1.psf	4822	5	0
1		A	D	F	Rpn11_2_2594_density.	0	0	1

IMD Status: Step 400

☒ Calculate Real-Time Cross Correlation

Experimental Density (Mol ID): 1

Selection: protein and noh

Map Resolution: 7.7

☐ Use Threshold

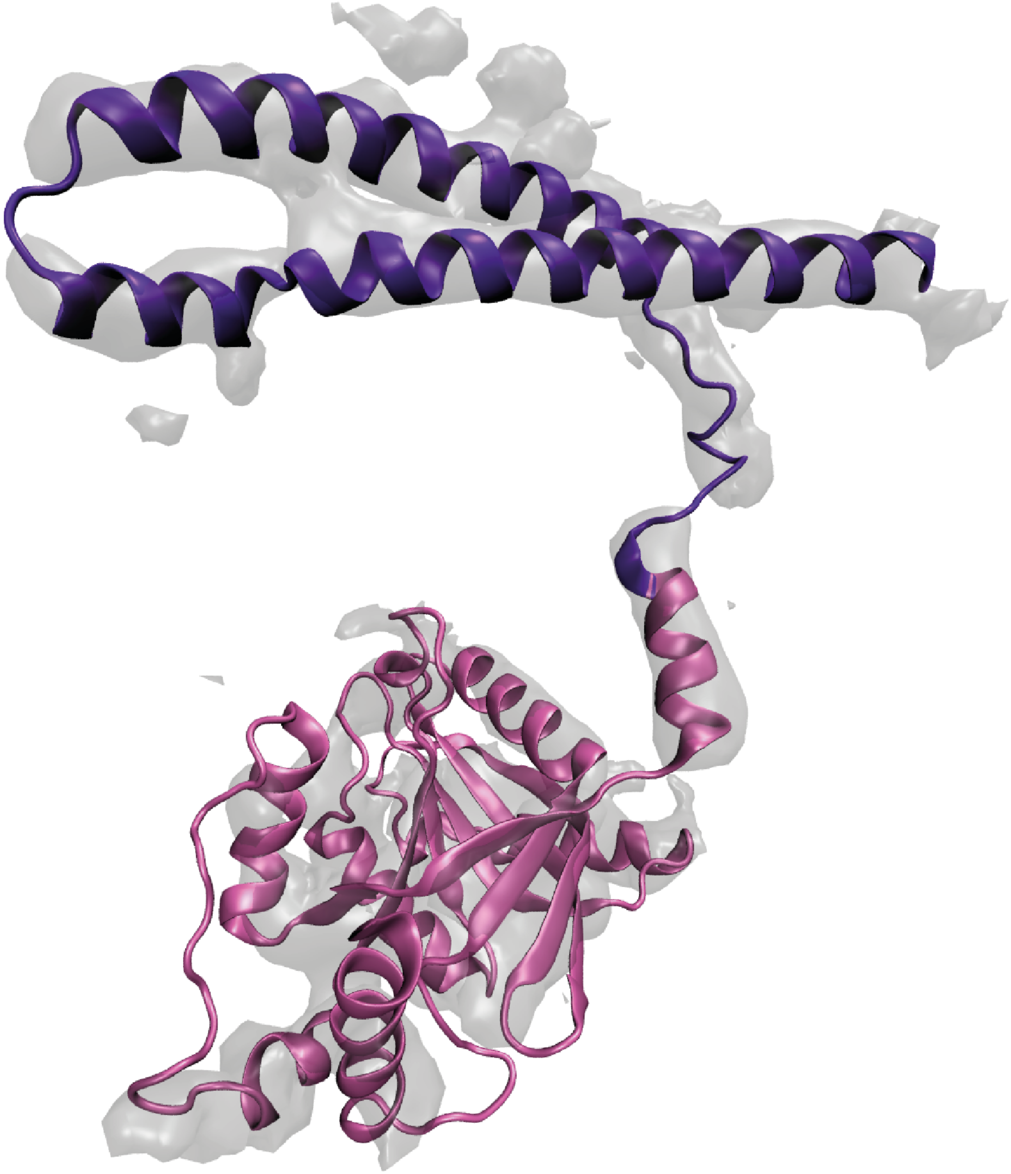
Threshold:

MDFF can be run on Cloud computing for low cost!

MDFF runs can be laucnhed through QwikMD!

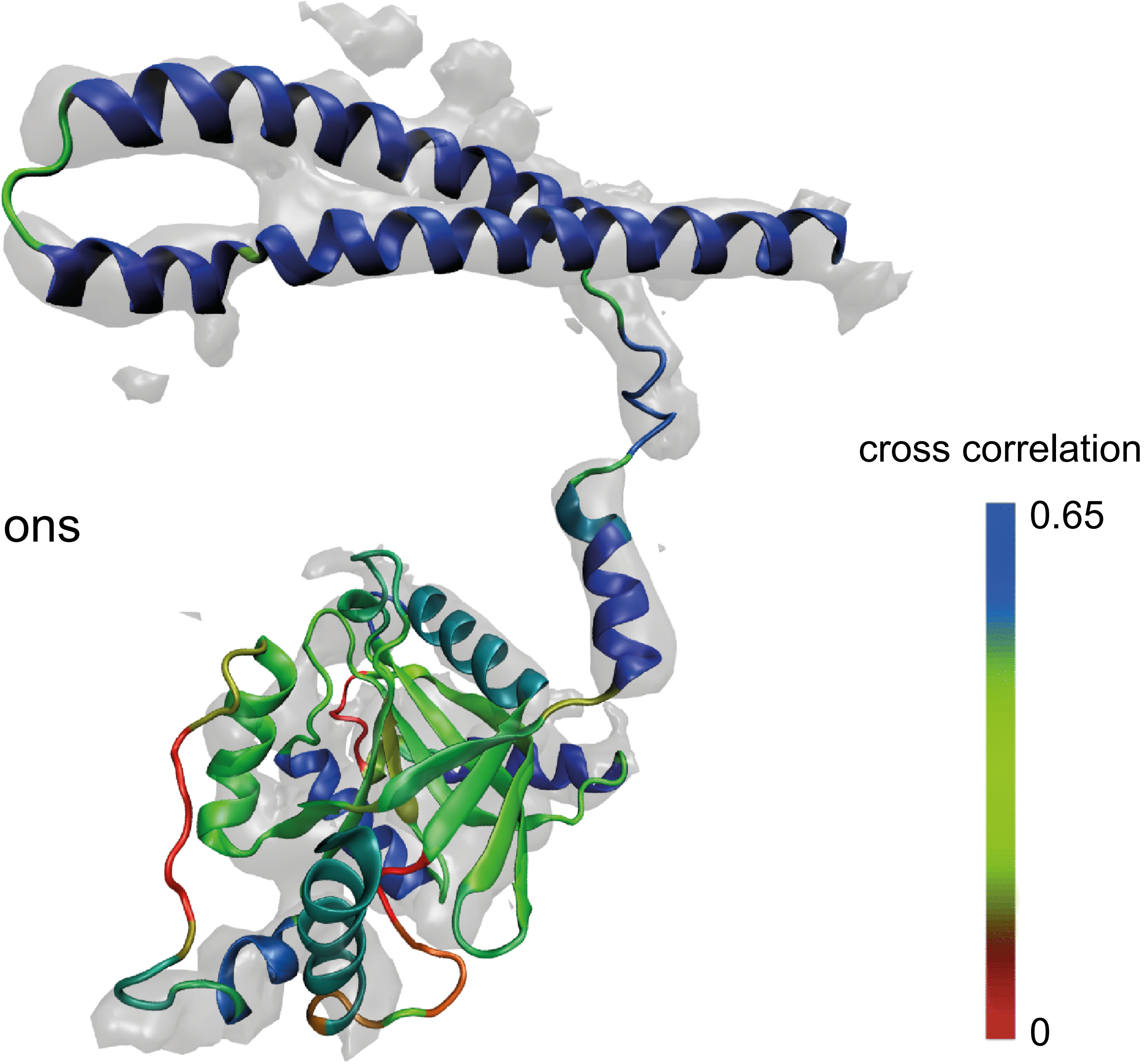
MDFF Tutorial on You Tube and at http://www.ks.uiuc.edu/Research/mdff/

Complete model of Rpn11 fitted to density



Quality check by cross-correlations

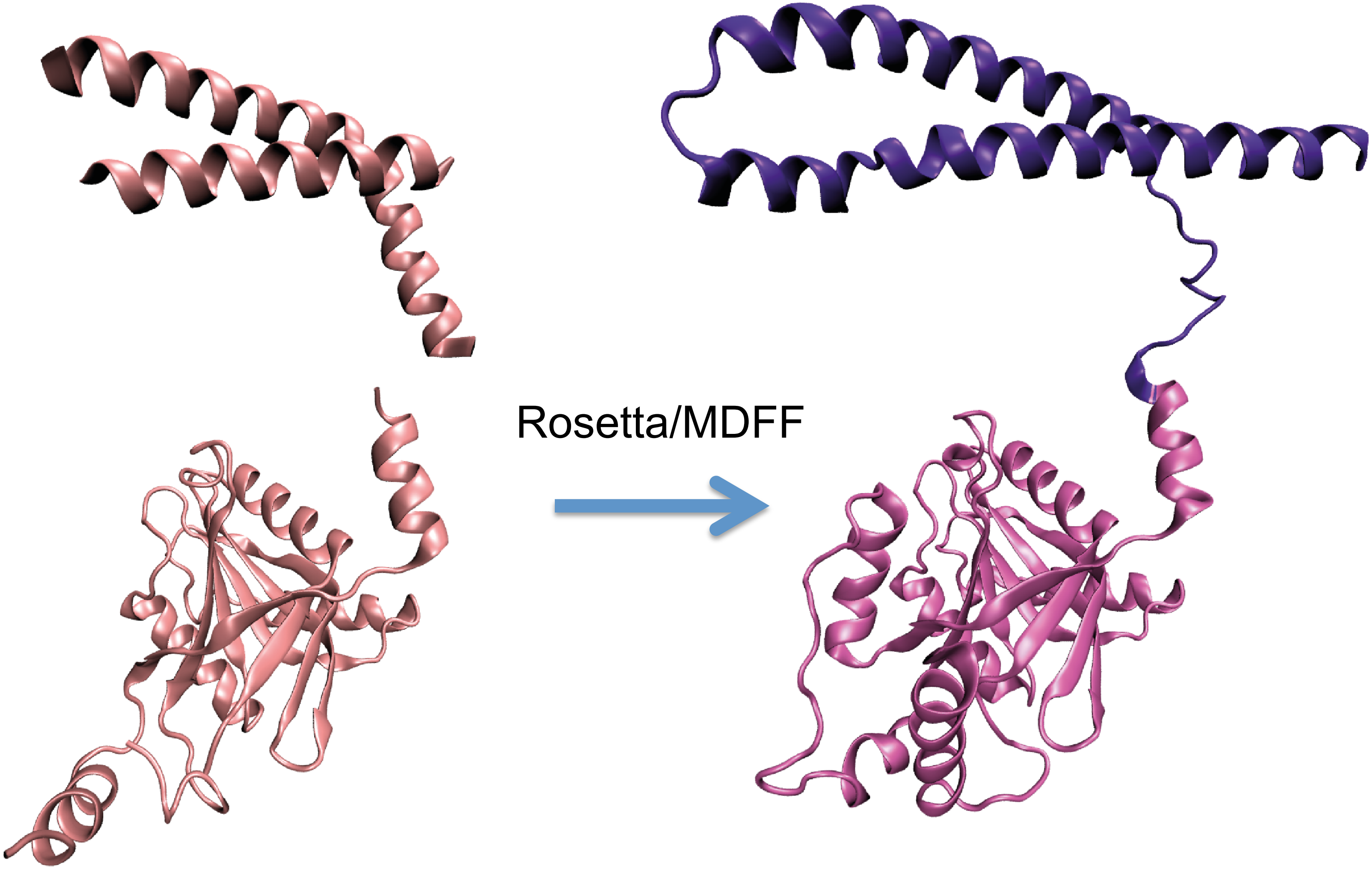
Rpn11 colored by local cross correlations



Incomplete vs. complete model

Incomplete model

Complete model



Rosetta/MDFF

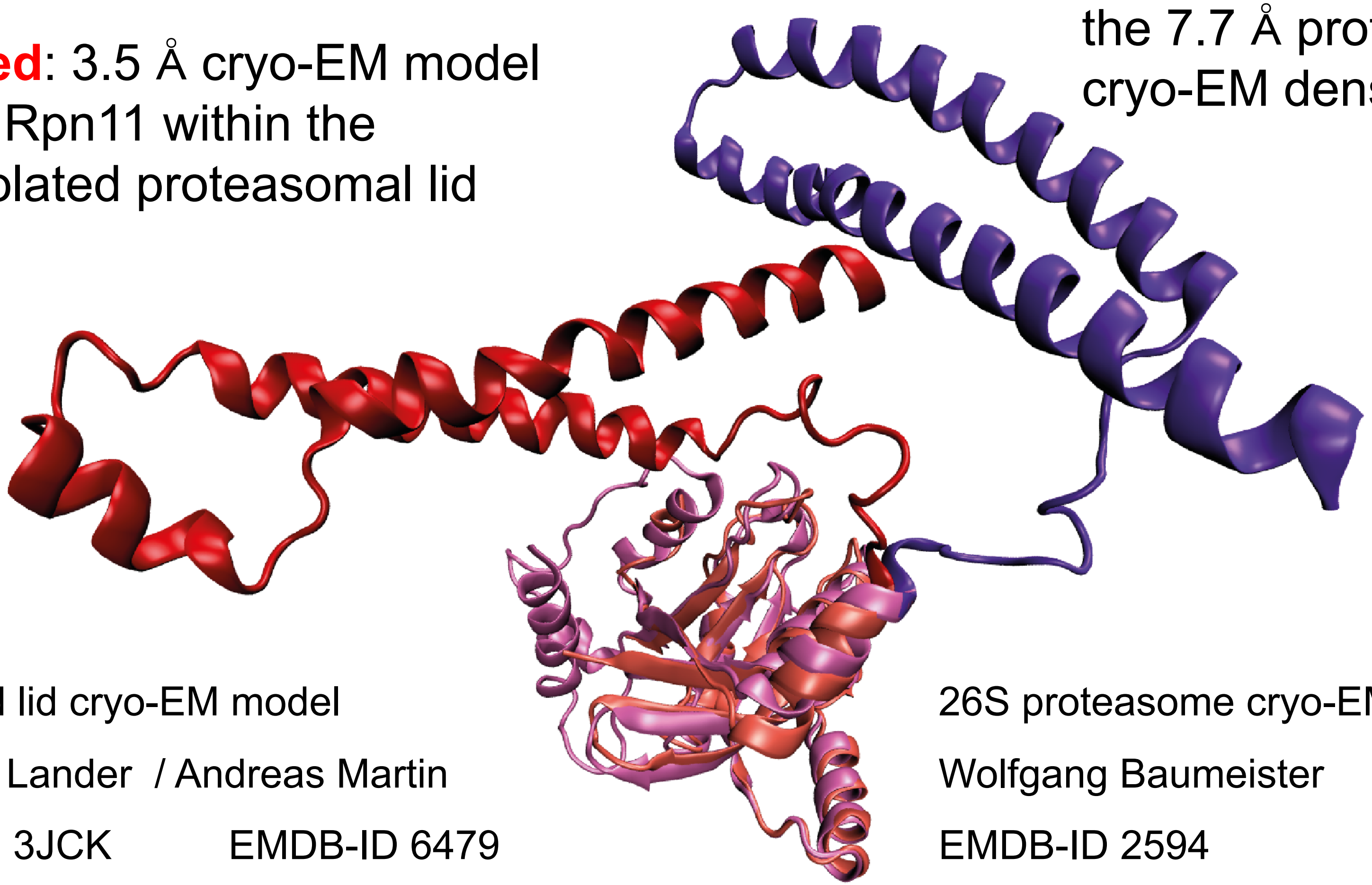
Cross correlation 0.61

Cross correlation 0.63

Low vs. high resolution density model

Purple: completed Rpn11 model within the 7.7 Å proteasomal cryo-EM density

Red: 3.5 Å cryo-EM model of Rpn11 within the isolated proteasomal lid



Isolated lid cryo-EM model

Gabriel Lander / Andreas Martin

PDB-ID 3JCK EMDB-ID 6479

Resolution 3.5 Å

Dambacher *et al.* eLife 2016

26S proteasome cryo-EM density

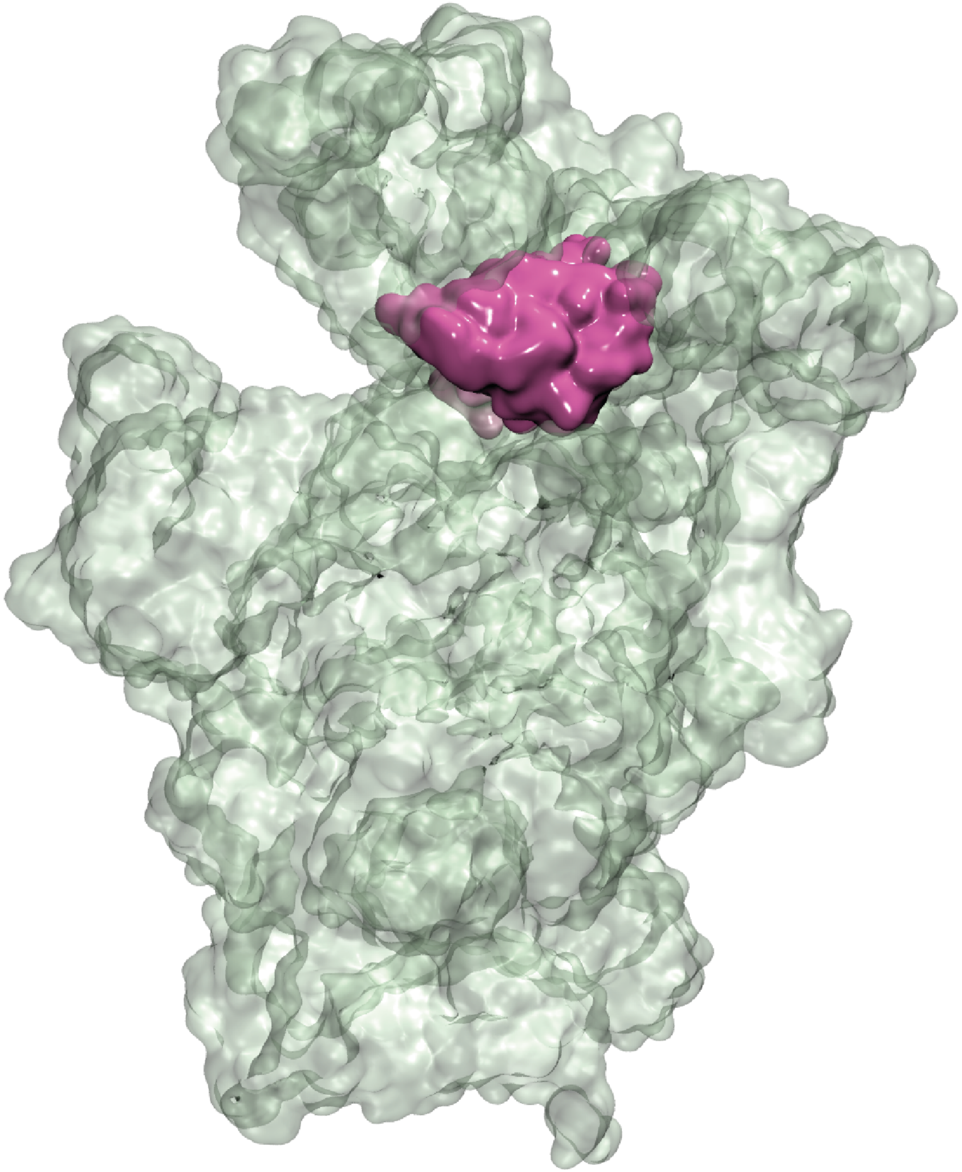
Wolfgang Baumeister

EMDB-ID 2594

Resolution 7.7 Å

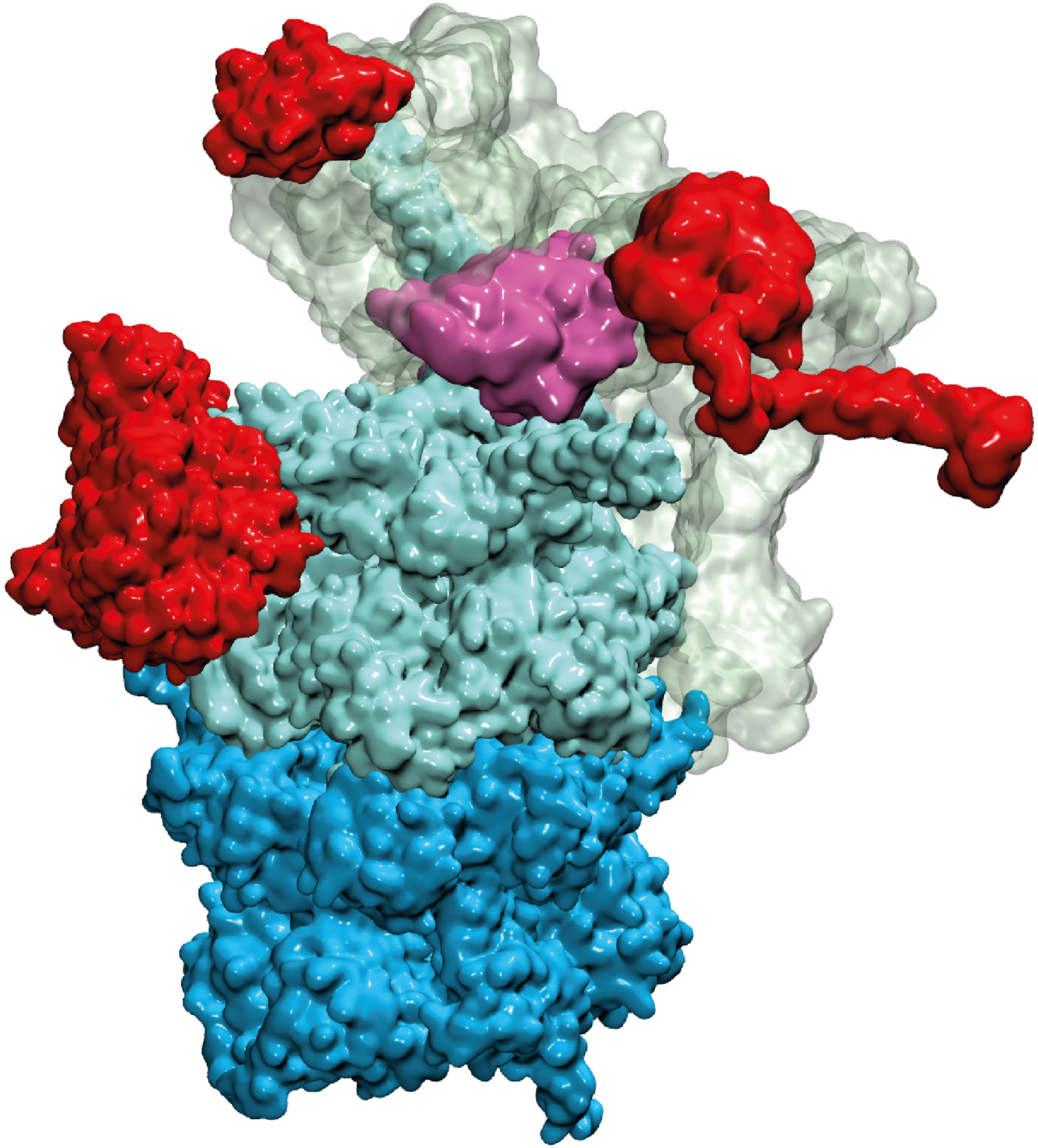
Unverdorben *et al.* PNAS 2014

Low vs. high resolution density model



Deubiquitylation
(Rpn11)

Functional subunits of the 26S proteasome



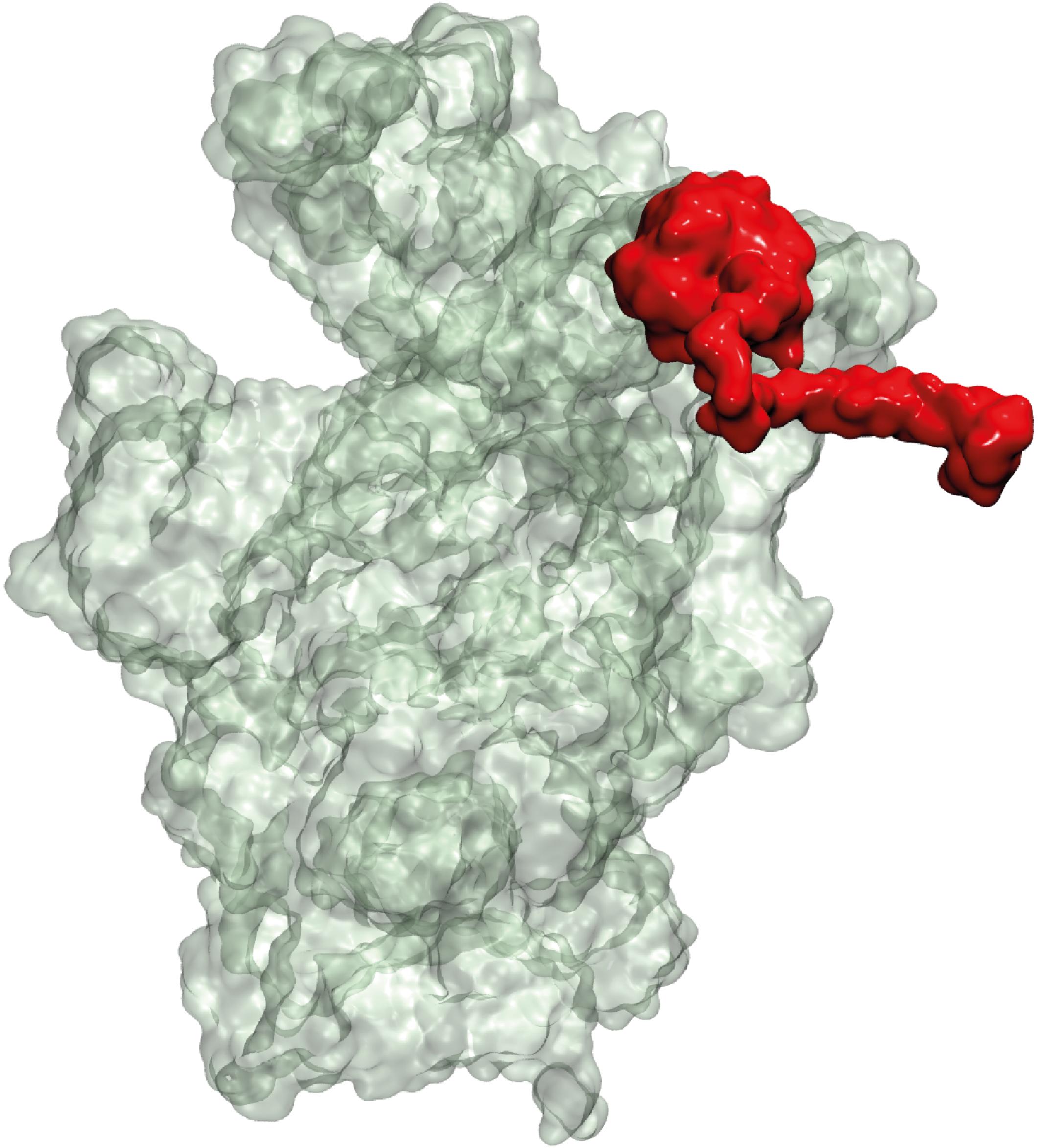
Ubiquitin
Recognition
(Rpn10, Rpn13, Rpn1)

Deubiquitylation
(Rpn11)

Substrate
Unfolding
(ATPase-ring)

Substrate
Degradation
(α-ring, β-ring)

Ubiquitin recognition by Rpn10



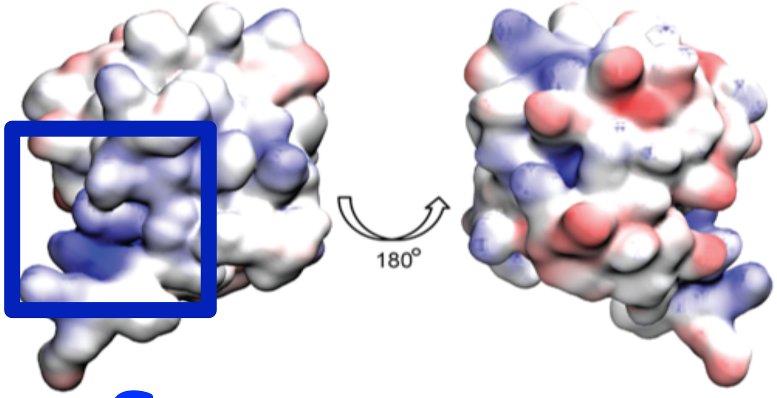
Ubiquitin
Recognition
(Rpn10)

Zhang*, Vucovic*, Rudack*, Han, Schulten 2016 JPC B (in press)

Ubiquitin Recognition

δ-

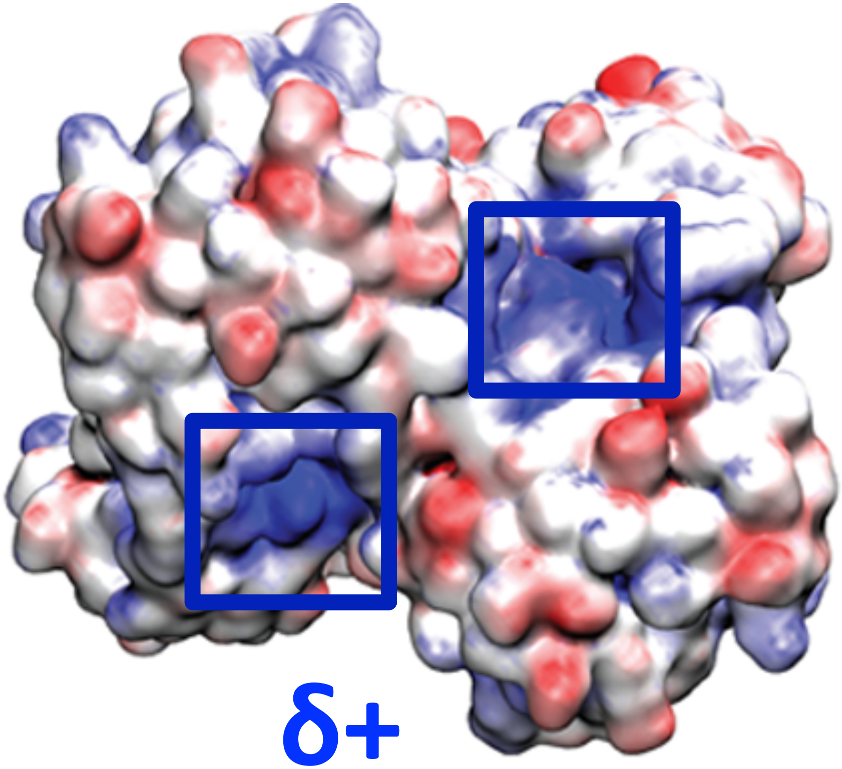
Rpn10
(UIM)



δ+

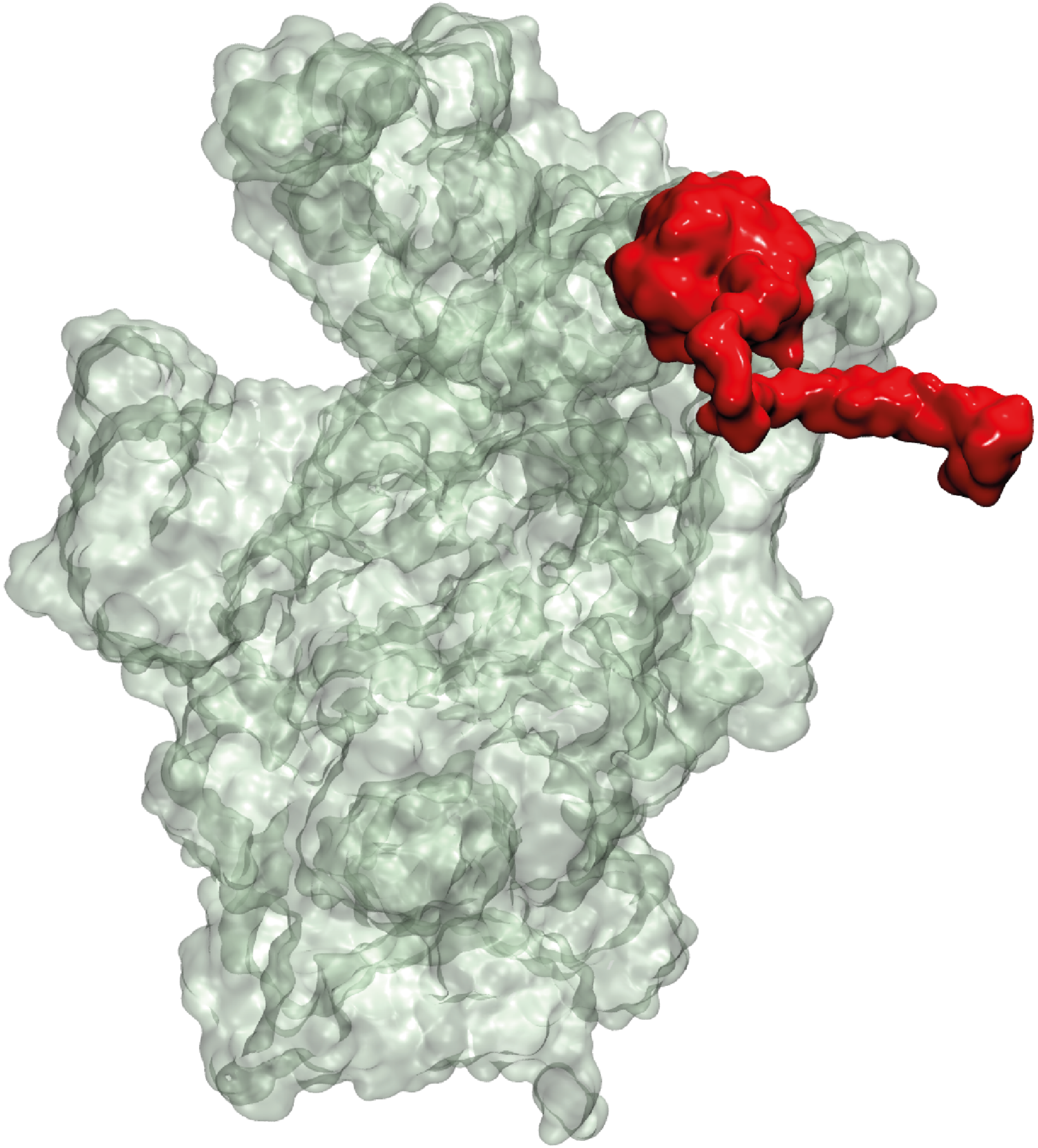
Monoubiquitin

Tetraubiquitin



δ+

Ubiquitin recognition by Rpn10

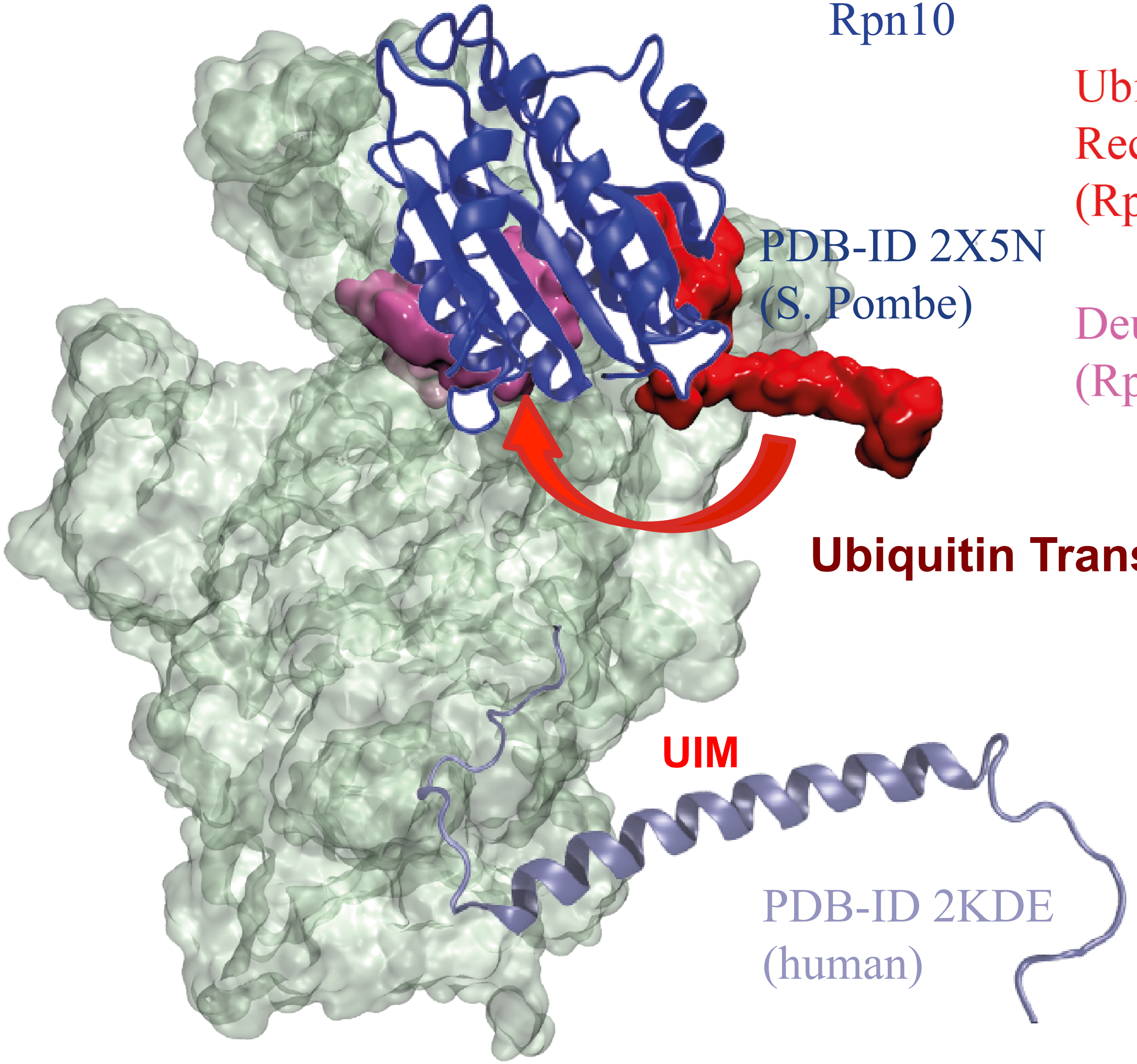


Ubiquitin
Recognition
(Rpn10)

www.ks.uiuc.edu/~trudack

Zhang*, Vucovic*, **Rudack***, Han, Schulten 2016 JPC B (in press)

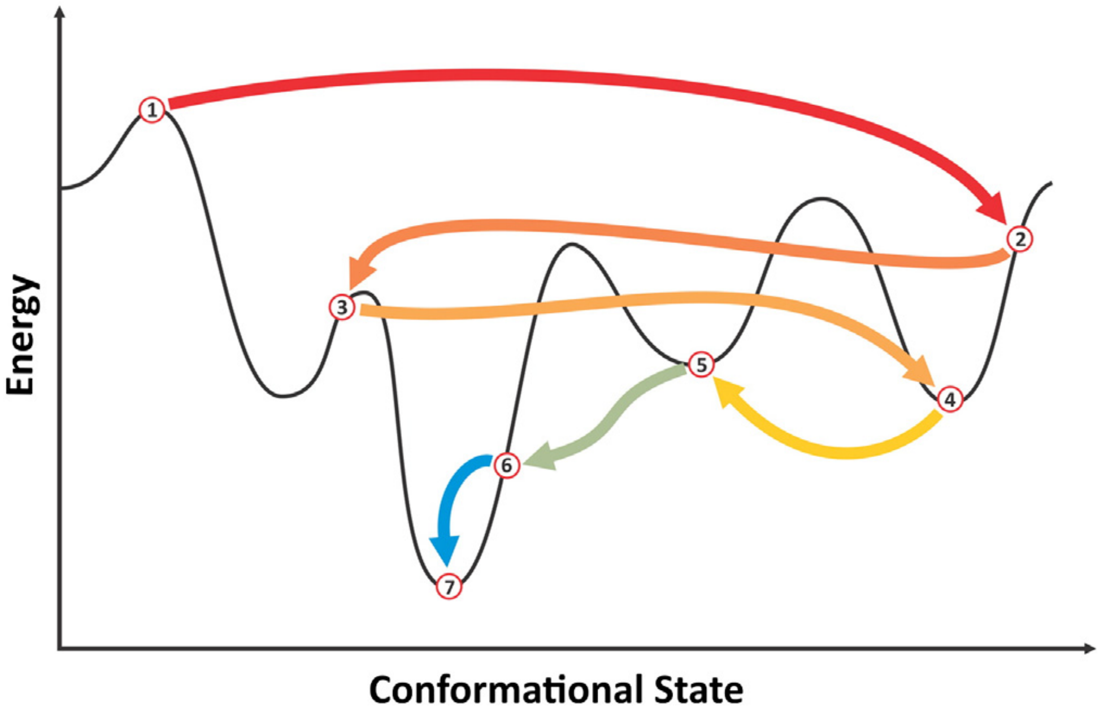
Ubiquitin recognition and deubiquitylation



Generalized Simulated Annealing – GSAFold

GSAFold NAMD Plugin – Allows *ab initio* structure prediction

New implementation of GSA on supercomputers allows the conformational search for large flexible regions.

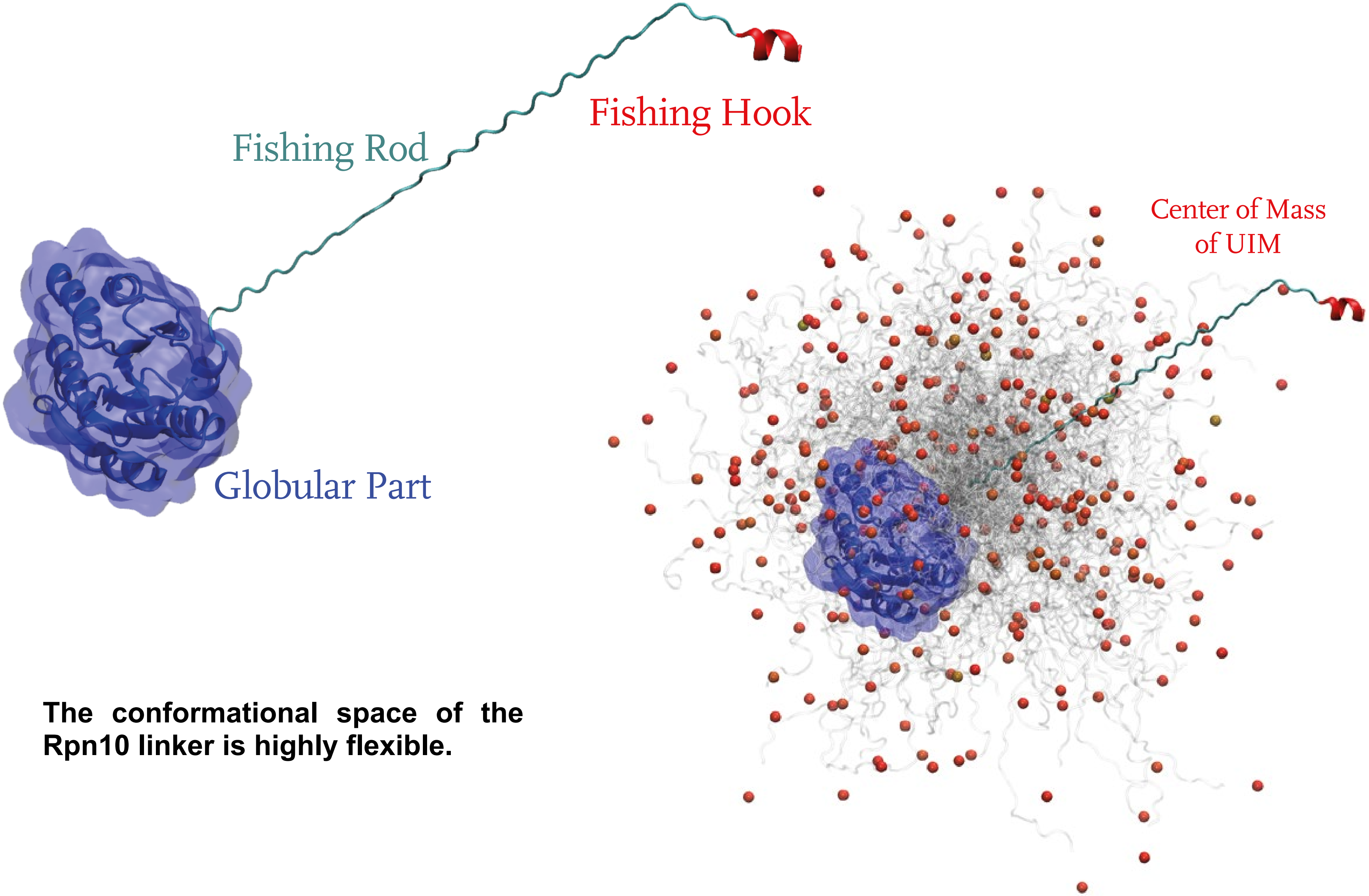


- Amino acid residues connecting Rpn10's UIM with the proteasome are likely to be disordered and stochastic searching algorithms such as GSA can be used to explore their conformational space

- GSAFold coupled to NAMD searches low-energy conformations to be used as starting points for the molecular dynamics studies.

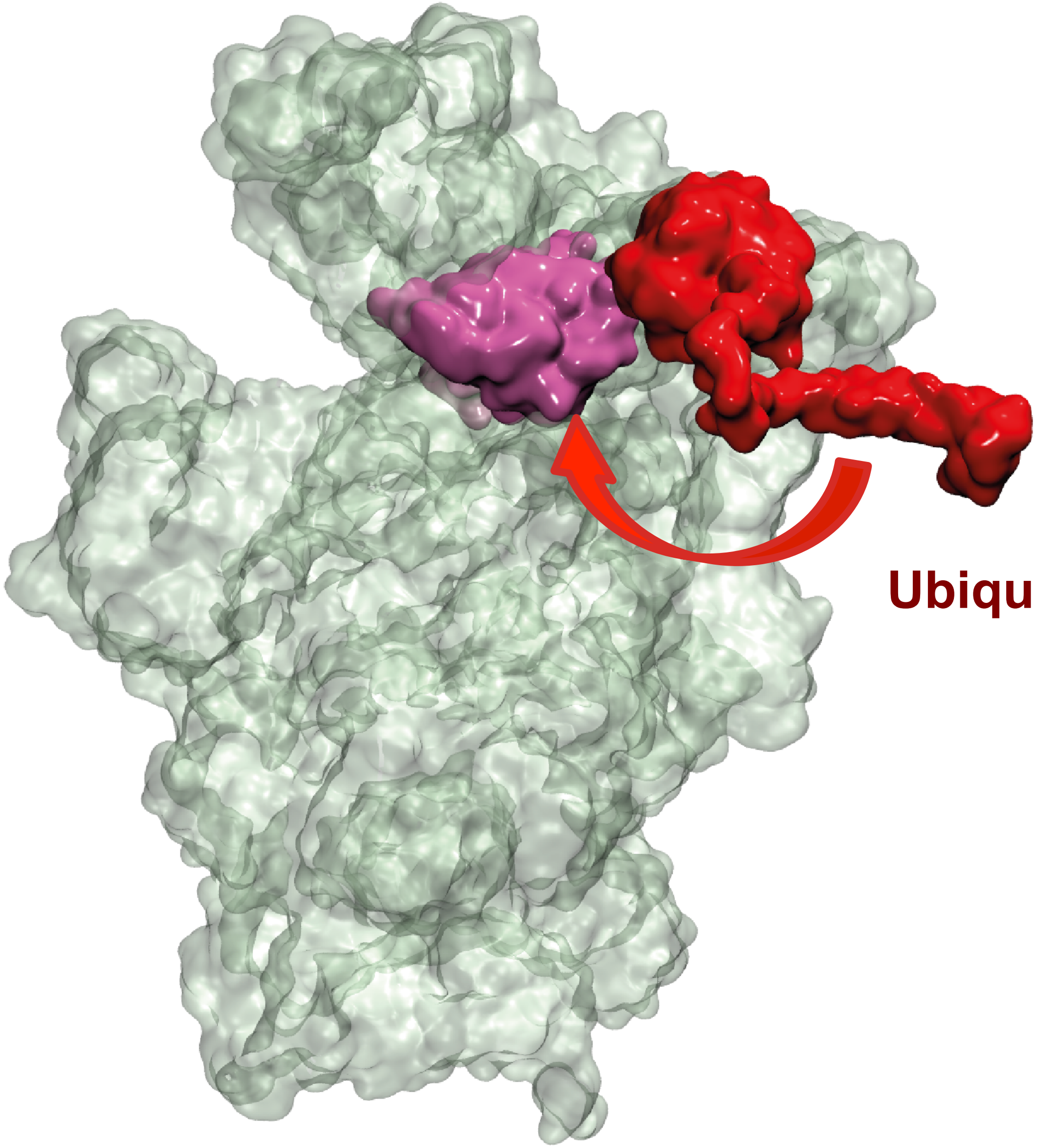
Rafael C. Bernardi Marcelo Melo

Conformation Space of Rpn10 Anchor



The conformational space of the Rpn10 linker is highly flexible.

Ubiquitin Transport to Deubiquitinase Rpn11

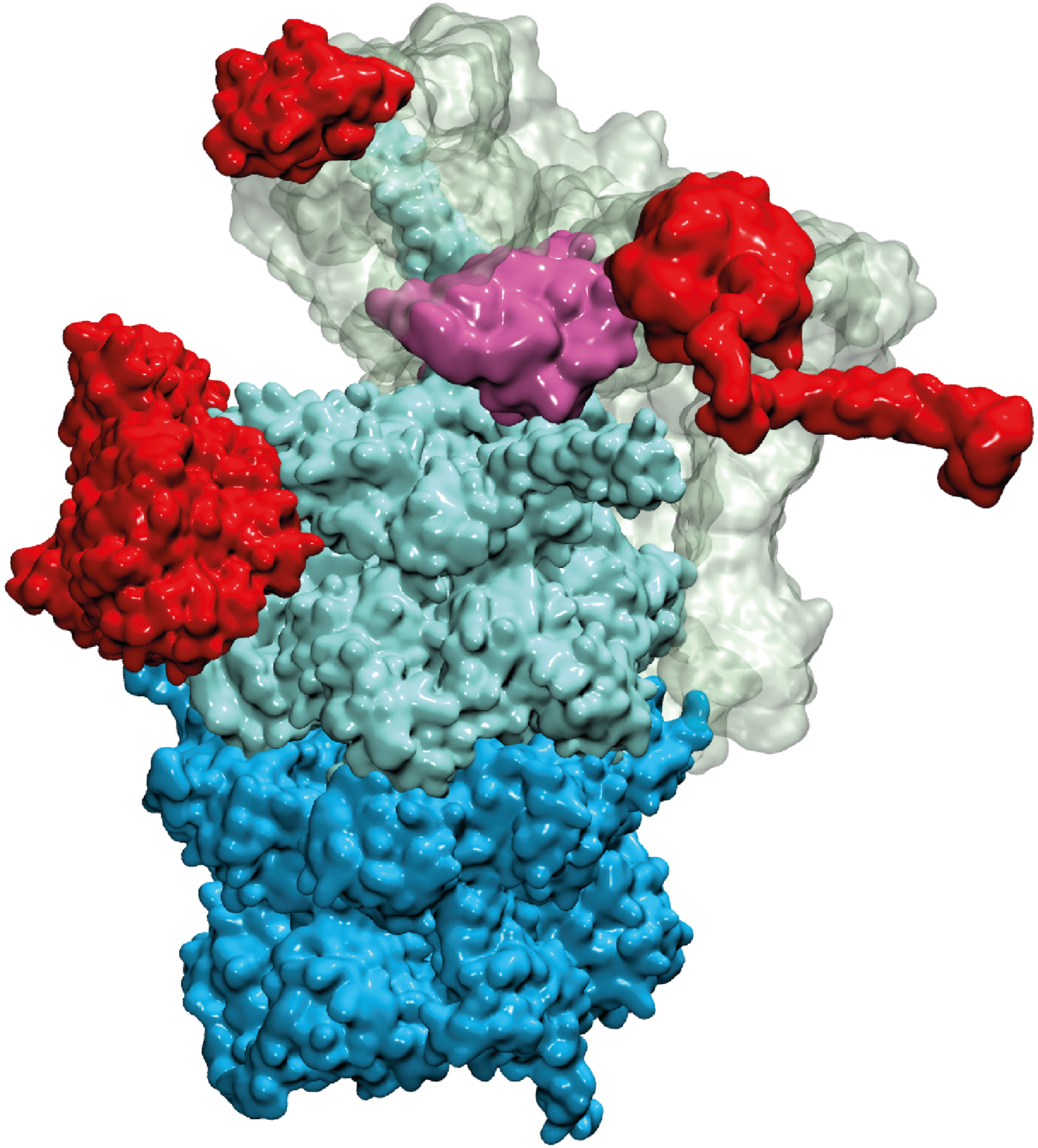


Ubiquitin
Recognition
(Rpn10)

Deubiquitylation
(Rpn11)

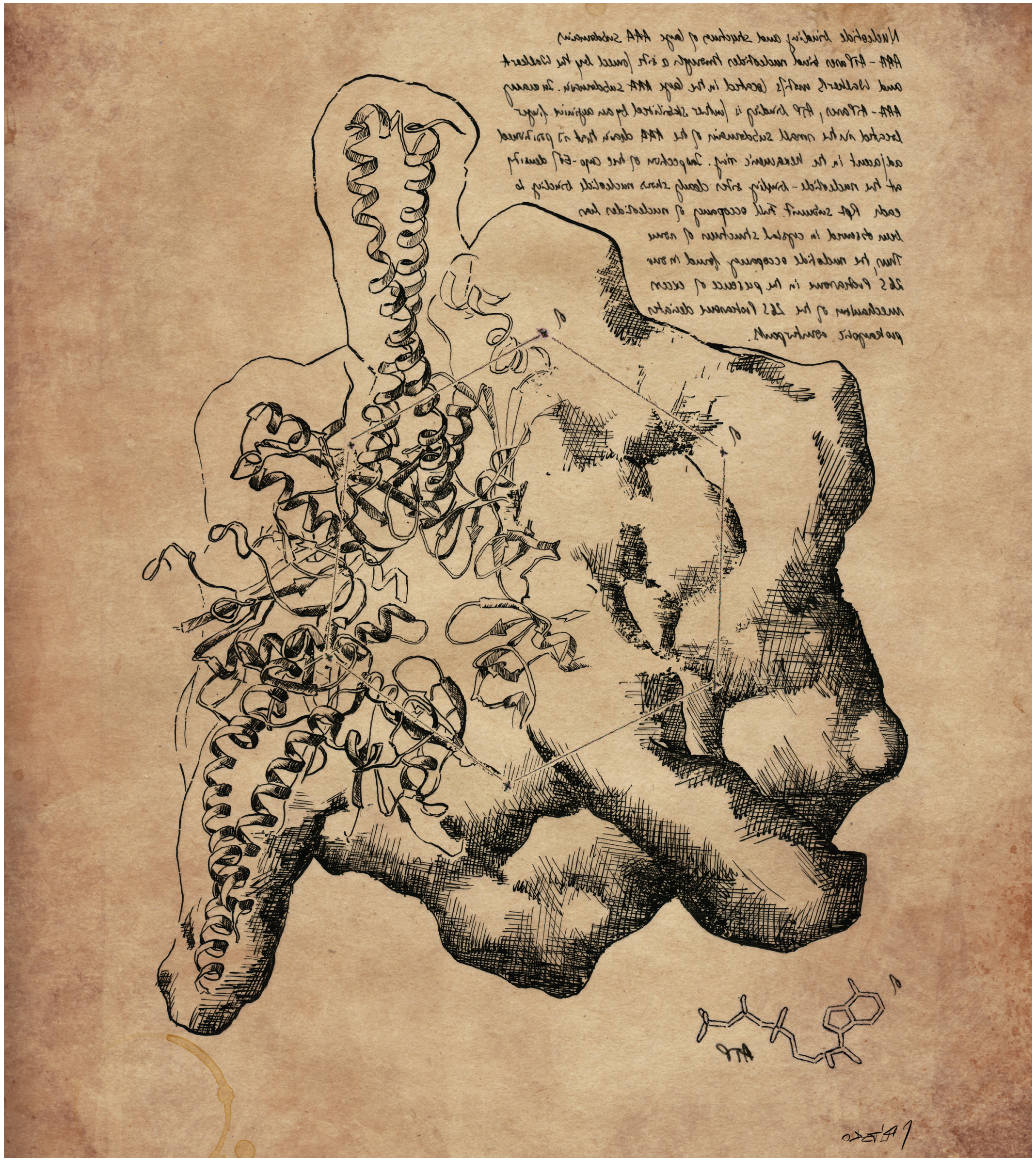
Ubiquitin Transport

Functional subunits of the 26S proteasome

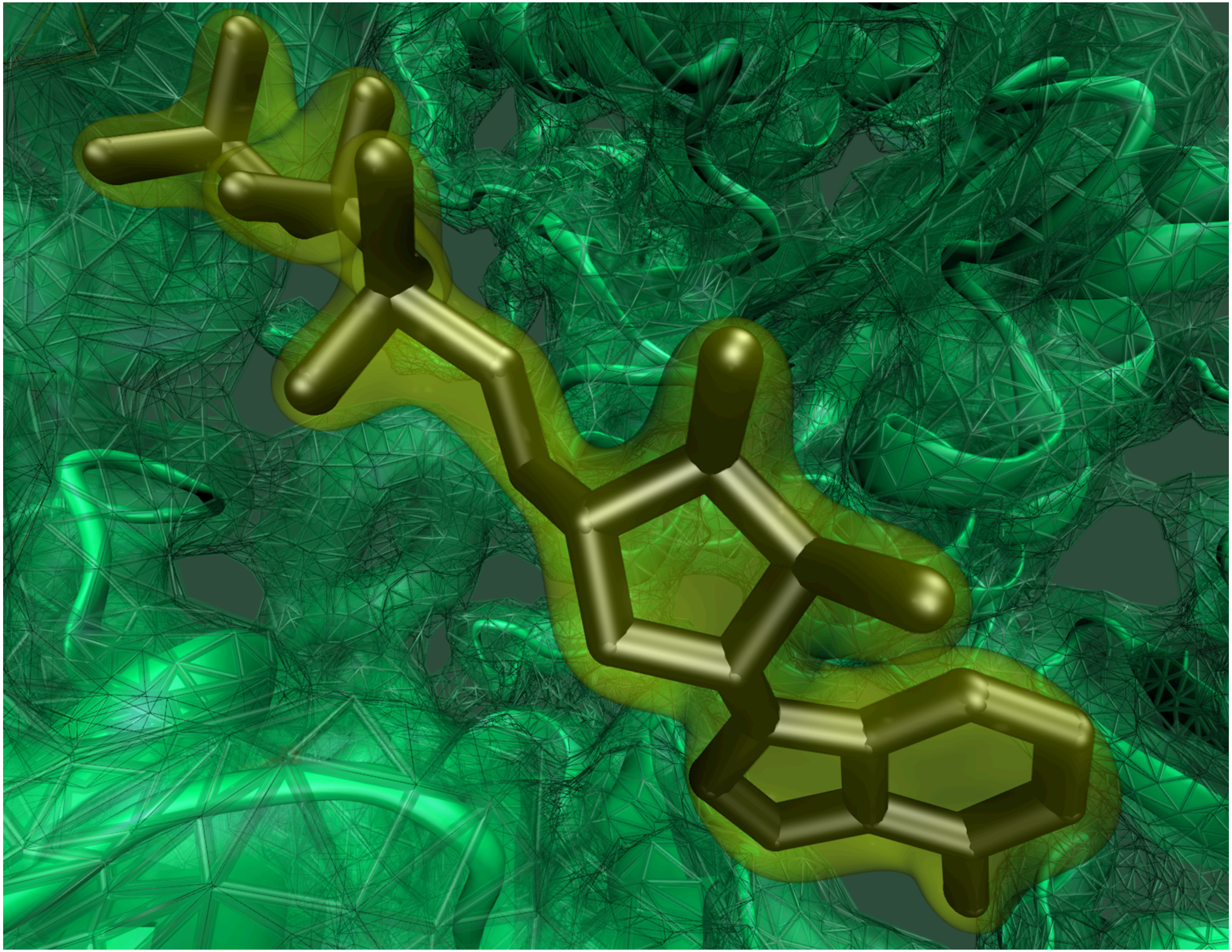


Substrate
Unfolding
(ATPase-ring)

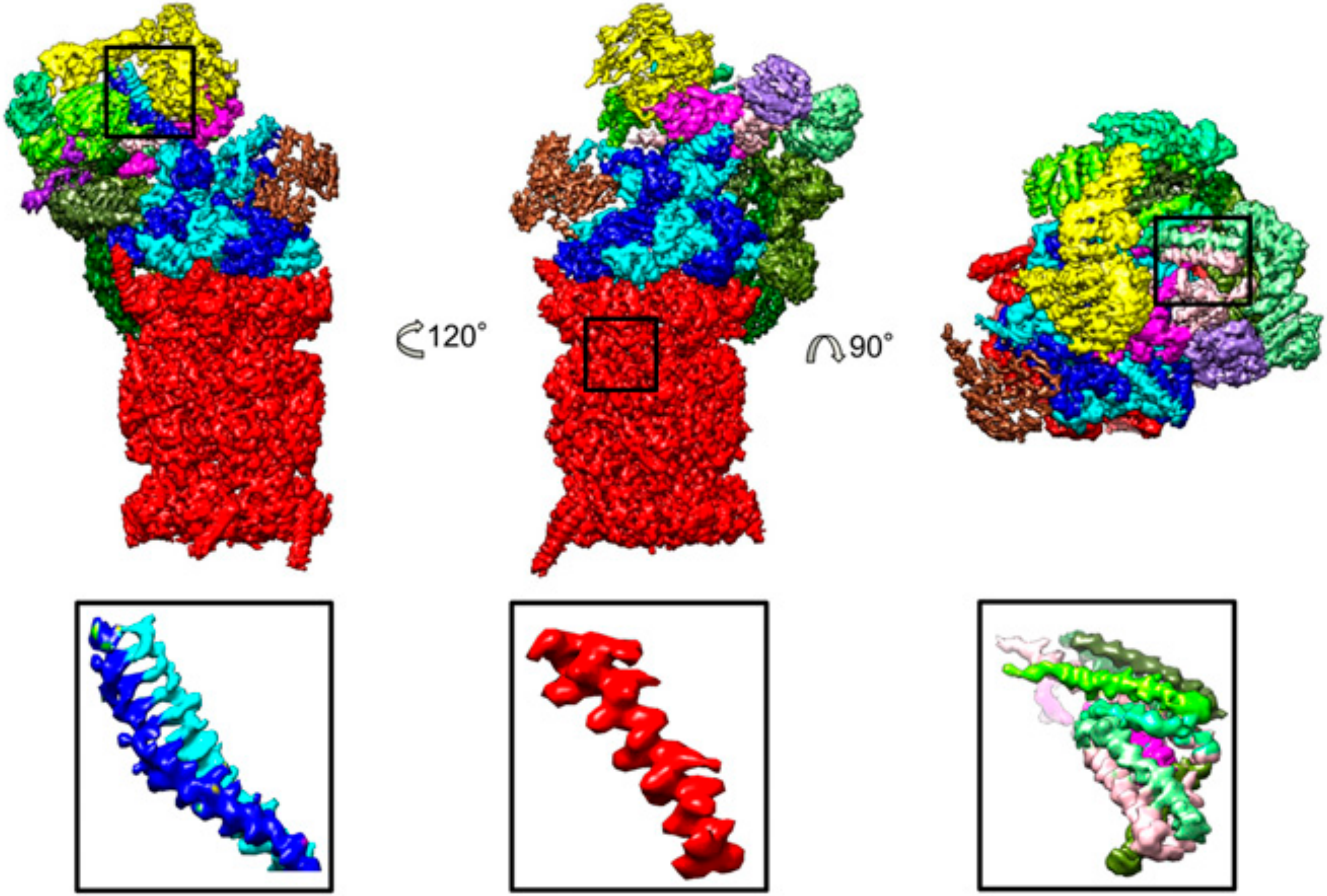
The Motor of the Proteasome



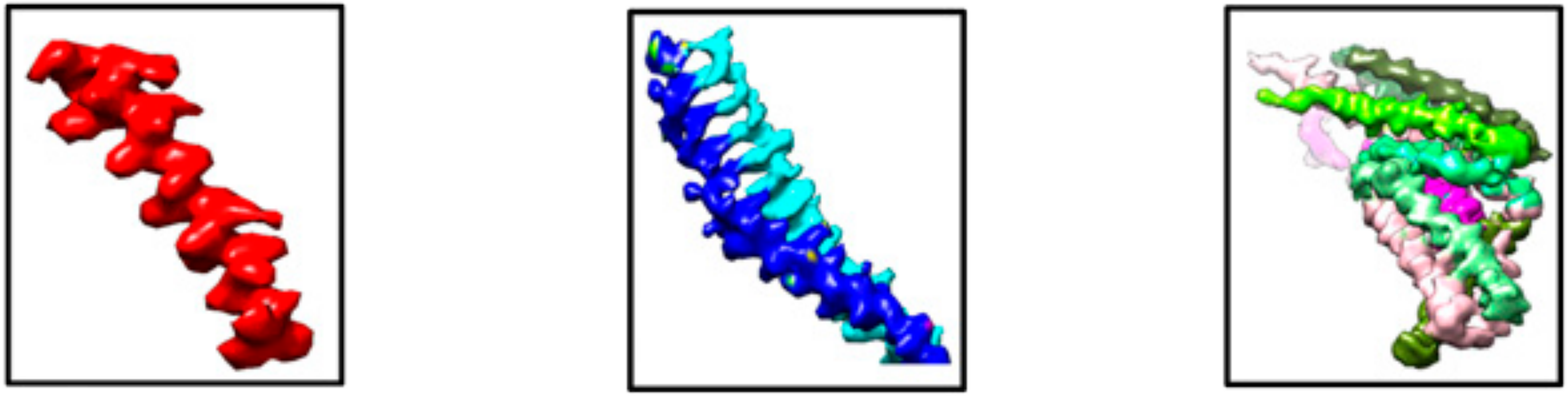
Resolved nucleotides are needed



3.9 Å Resolution Density of the Human 26S Proteasome



High-resolution Real Space Refinement with MDFF



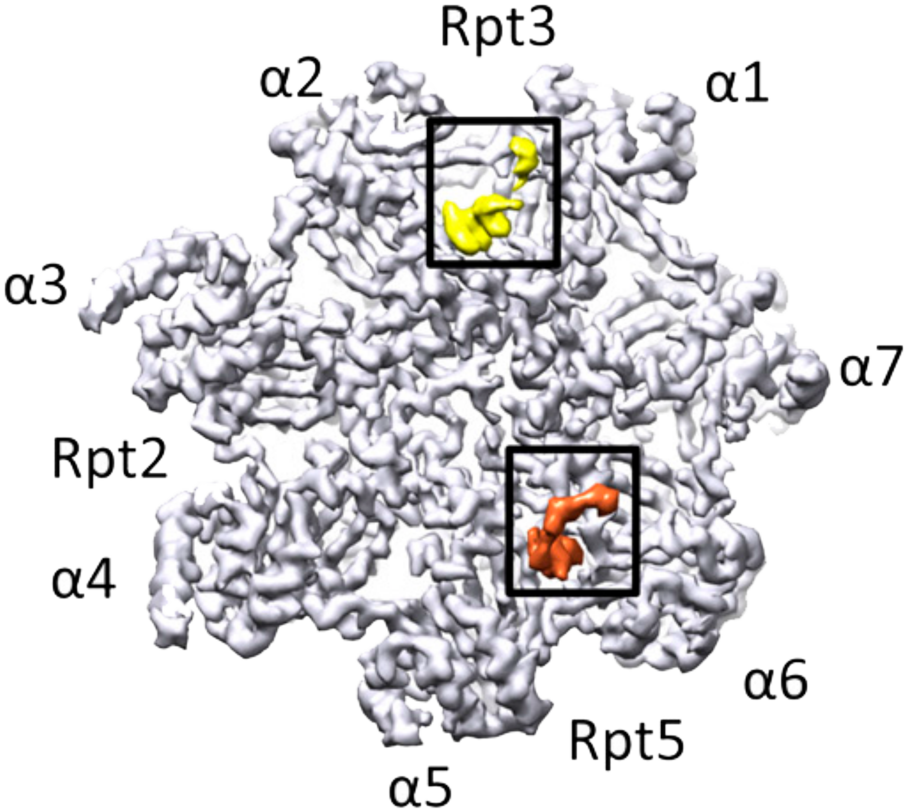
Advantage:
Positions of bulky side chains can be observed from density

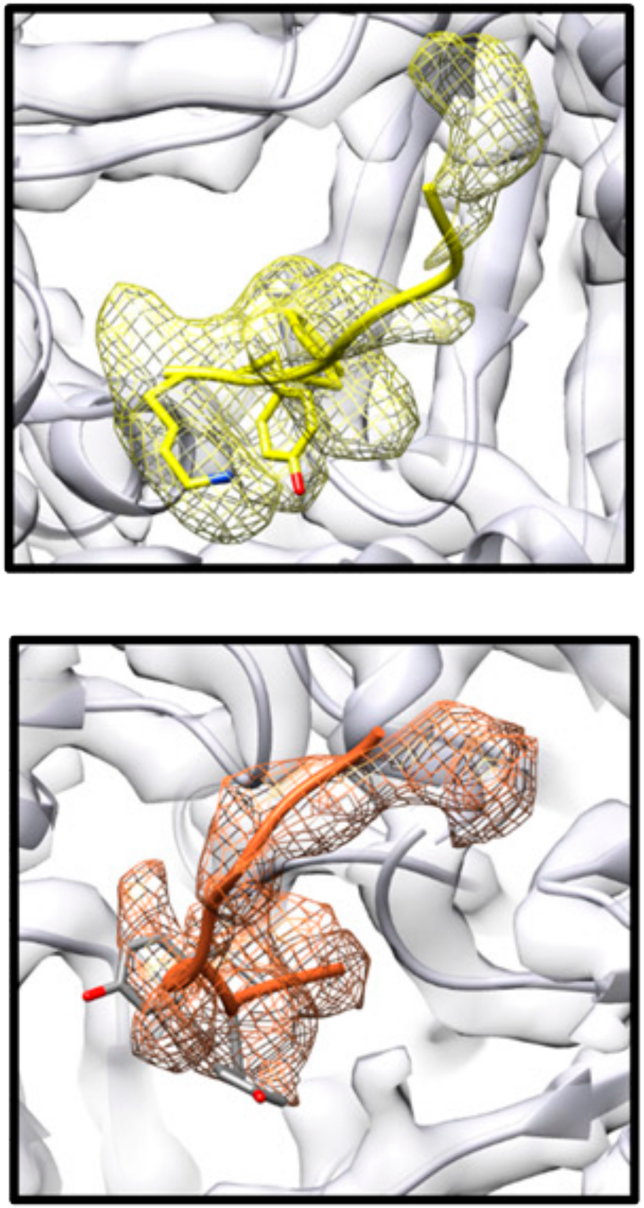
Challenge:
no detailed side chain orientation
X-ray structure refinement tools failed in the range of 4-5 Å resolution

Solution:
combining MDFF with
monte carlo based backbone and side chain rotamer search algorithms
in an iterative manner

Goh, Hadden, Bernardi, Singharoy, McGreevy, Rudack, Cassidy, Schulten,
Annu. Rev. Biophys., 2016 45.1

The ATPase Motor of the 26S Proteasome

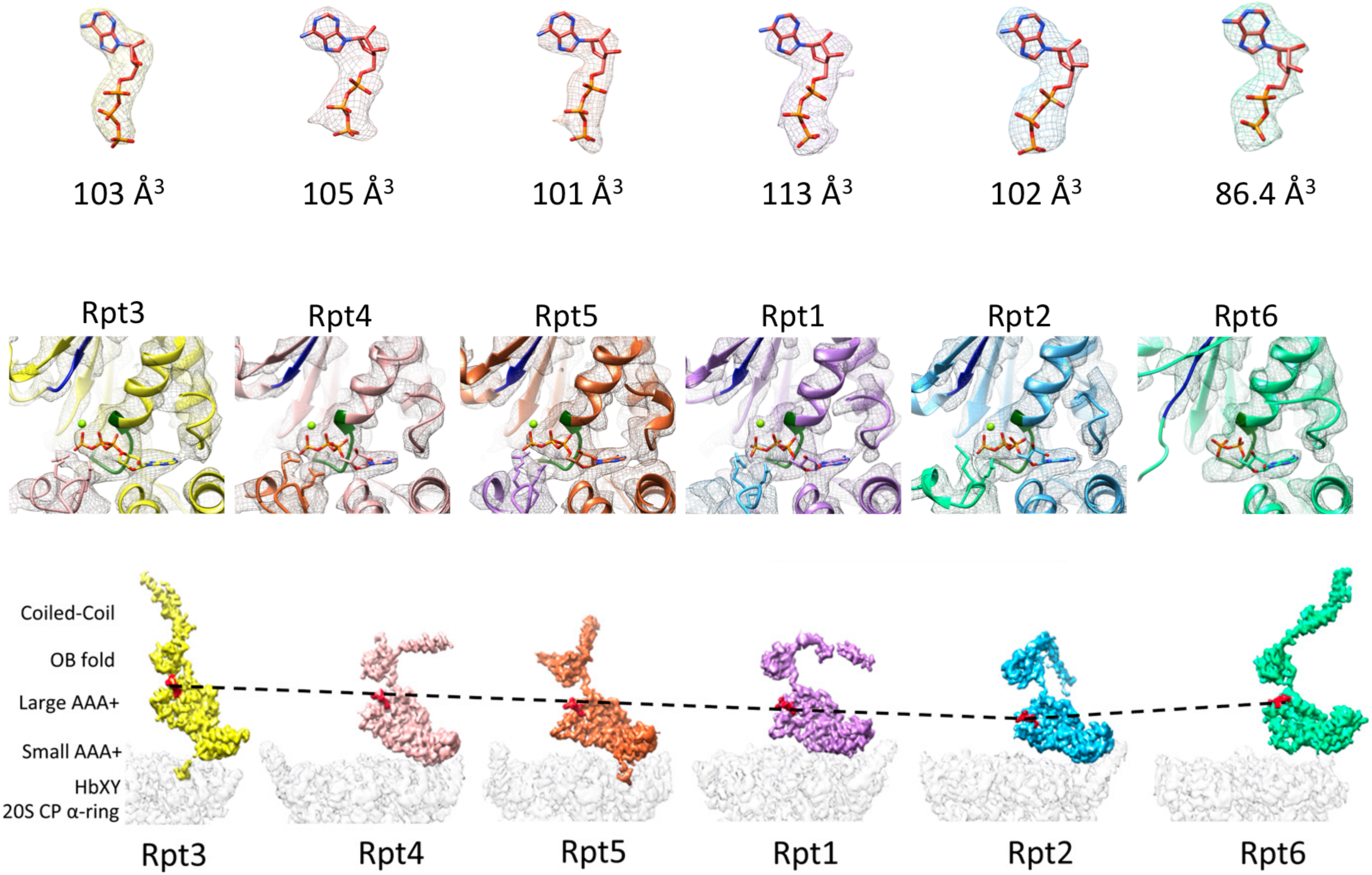




PDB-IDs: 5L4G, 5L4K
EMDB-ID: 4002

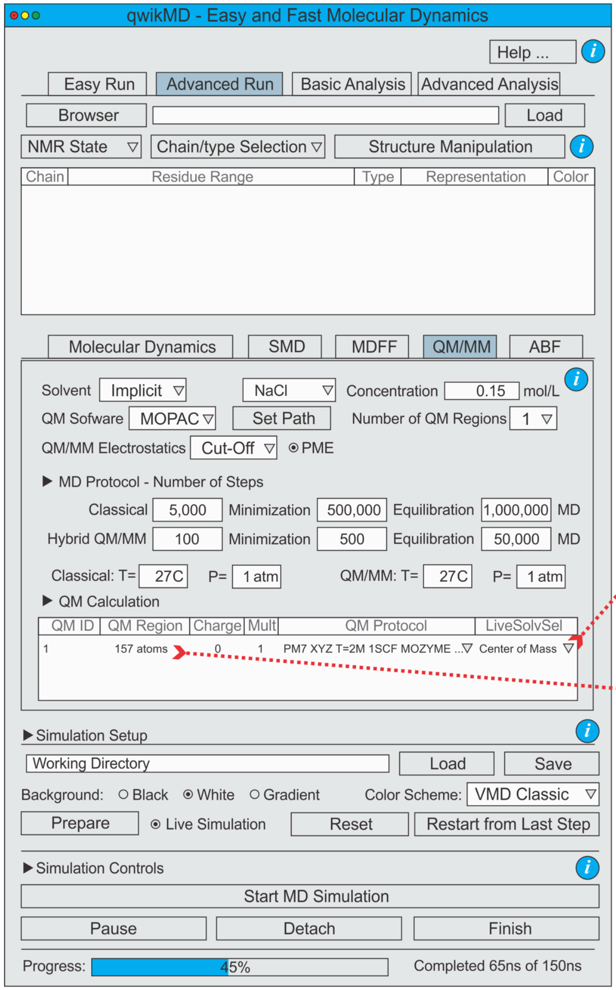
Schweitzer A, Aufderheide A, Rudack T, et al. “The structure of the 26S proteasome at a resolution of 3.9 Å.” PNAS 2016 in press.

The Motor Action of protein unfolding



NAMD QM/MM interface

The atomic structure enable detailed investigations of the unfolding process by path sampling techniques. Chemical reaction in the active sites can be studied through QM/MM simulations.



qwikMD - Easy and Fast Molecular Dynamics

Solvent: Implicit; NaCl; Concentration 0.15 mol/L

QM Sofware: MOPAC; Set Path; Number of QM Regions: 1

QM/MM Electrostatics: Cut-Off; ⊙ PME

▶ MD Protocol - Number of Steps

Classical	5,000	Minimization	500,000	Equilibration	1,000,000	MD
Hybrid QM/MM	100	Minimization	500	Equilibration	50,000	MD

Classical: T= 27C P= 1 atm QM/MM: T= 27C P= 1 atm

▶ QM Calculation

QM ID	QM Region	Charge	Mult	QM Protocol	LiveSolvSel
1	157 atoms	0	1	PM7 XYZ T=2M 1SCF MOZYME ...	Center of Mass

▶ Simulation Setup

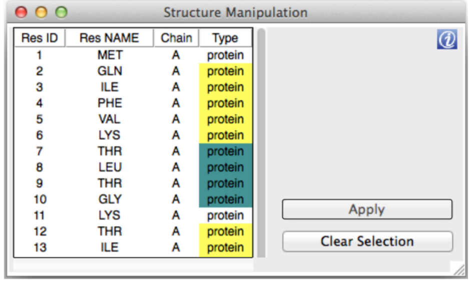
Background: ○ Black ⊙ White ○ Gradient Color Scheme: VMD Classic

⊙ Live Simulation

▶ Simulation Controls

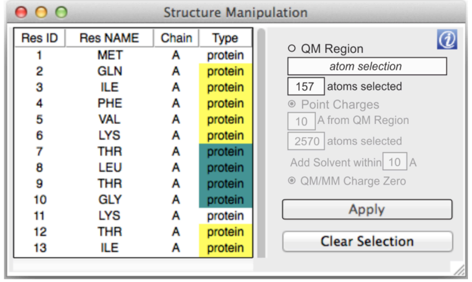
Progress: 45% Completed 65ns of 150ns

NAMD QM/MM interface with MOPAC and ORCA will be released in the second semester of 2016



Structure Manipulation

Res ID	Res NAME	Chain	Type
1	MET	A	protein
2	GLN	A	protein
3	ILE	A	protein
4	PHE	A	protein
5	VAL	A	protein
6	LYS	A	protein
7	THR	A	protein
8	LEU	A	protein
9	THR	A	protein
10	GLY	A	protein
11	LYS	A	protein
12	THR	A	protein
13	ILE	A	protein



Structure Manipulation

Res ID	Res NAME	Chain	Type
1	MET	A	protein
2	GLN	A	protein
3	ILE	A	protein
4	PHE	A	protein
5	VAL	A	protein
6	LYS	A	protein
7	THR	A	protein
8	LEU	A	protein
9	THR	A	protein
10	GLY	A	protein
11	LYS	A	protein
12	THR	A	protein
13	ILE	A	protein

○ QM Region

atom selection

157 atoms selected

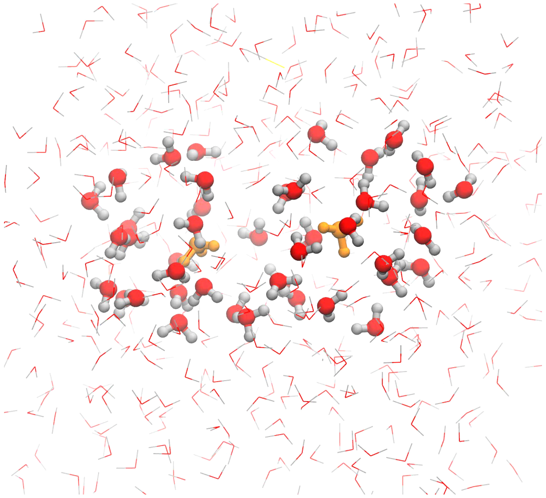
⊙ Point Charges

10 A from QM Region

2570 atoms selected

Add Solvent within 10 A

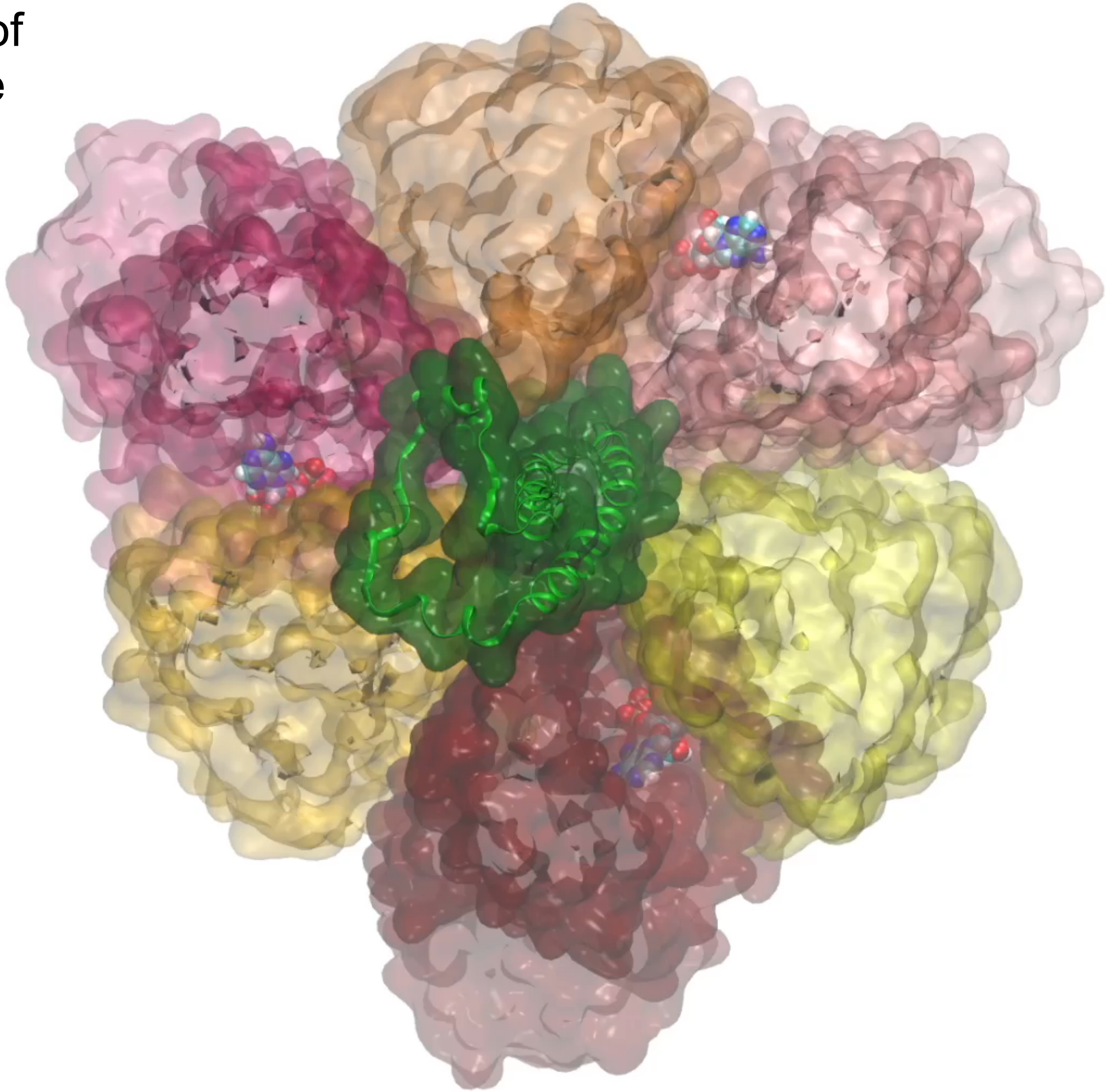
⊙ QM/MM Charge Zero



Next QwikMD release will support QM/MM

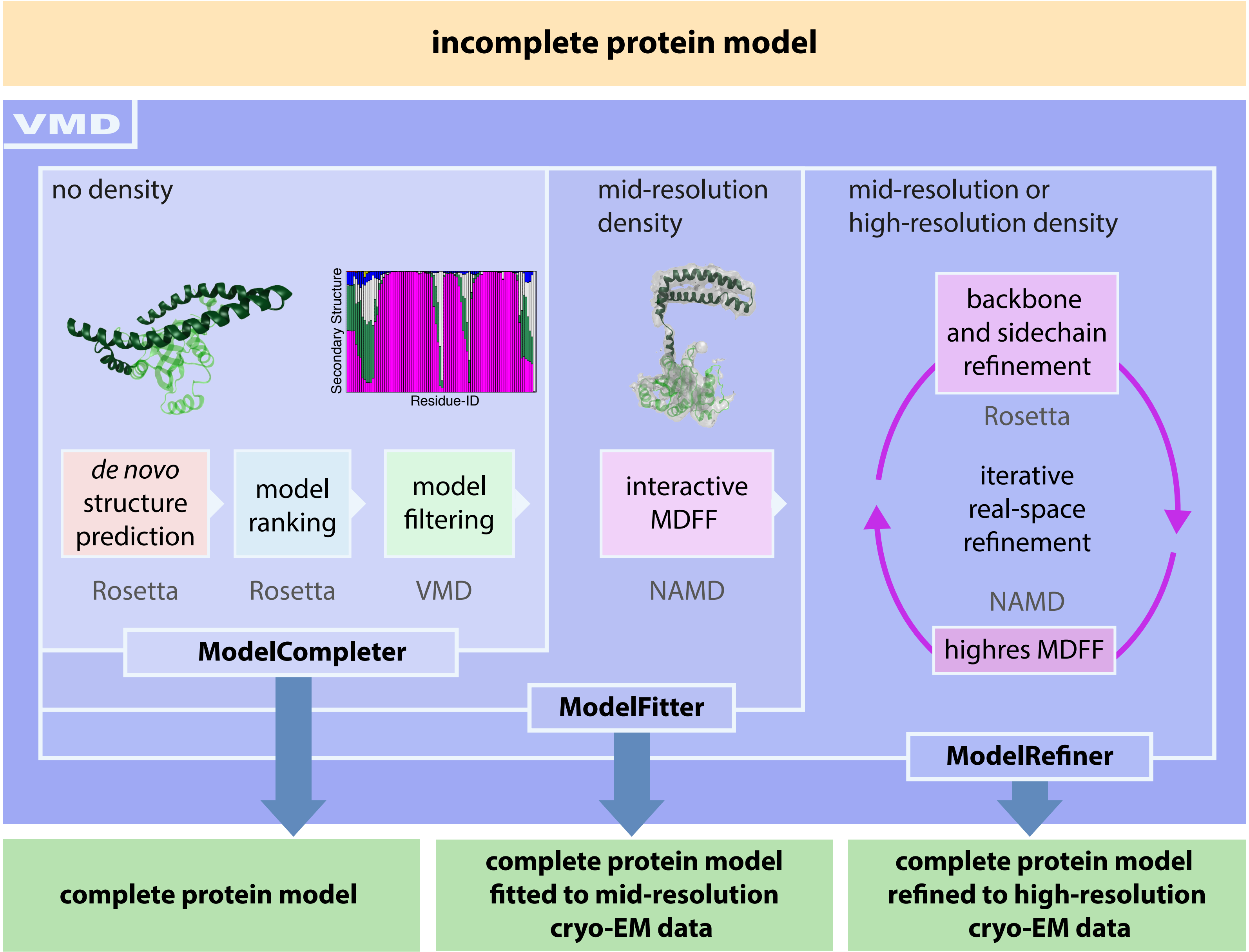
Converting Chemical Energy into Motor Action

Motor Action of
ATP Synthase



Abhi Singharoy

ModelMaker



incomplete protein model

no density

de novo structure prediction

model ranking

model filtering

Rosetta

Rosetta

VMD

ModelCompleter

mid-resolution density

interactive MDFF

NAMD

ModelFitter

mid-resolution or high-resolution density

backbone and sidechain refinement

Rosetta

iterative real-space refinement

NAMD

highres MDFF

ModelRefiner

complete protein model

complete protein model fitted to mid-resolution cryo-EM data

complete protein model refined to high-resolution cryo-EM data

Bridging Computation and Experiment

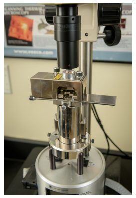
Proteins | **Multi-Protein Complex** | **Proteins + Membrane** | **Cell**

Experimental Data


NMR


X-ray

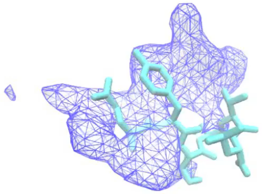

cryo-EM/ET

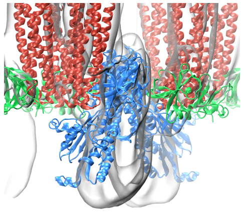

AFM

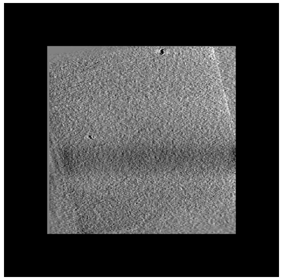

FRET

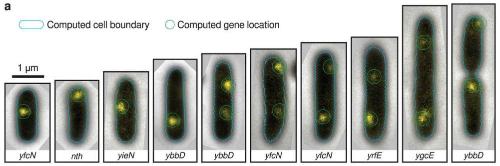
Computational Modeling

Model



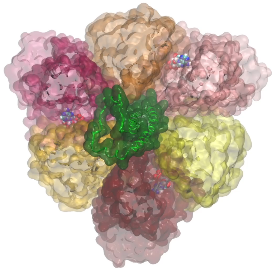


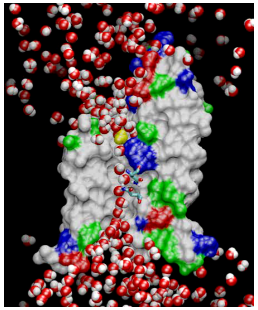




Molecular Dynamics

Function





Visualization Analysis

Experimental Validation

Take Home Message

In order to obtain **biomedical discoveries** different **experimental** and **computational** methods need to be **integrated.**

Automation is important but **user expertise** is equally important**.**

Acknowledgments

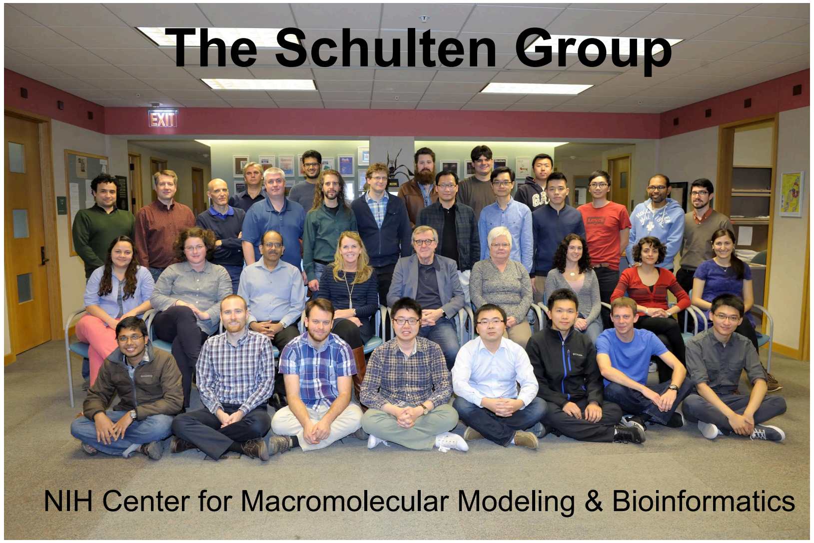
Alexander von Humboldt
Stiftung/Foundation

NATIONAL INSTITUTES OF HEALTH

NSF

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

max planck institute
of biochemistry



The Schulten Group

NIH Center for Macromolecular Modeling & Bioinformatics

Center for the Physics
of Living Cells

BLUE WATERS
SUSTAINED PETASCALE COMPUTING

Theory

Klaus Schulten
Ryan McGreevy

Experiment

Wolfgang Baumeister
Friedrich Förster
Antje Aufderheide

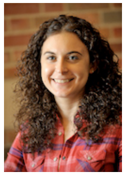
ModelMaker

Ryan McGreevy

GSA

Rafael Bernardi
Marcelo Mello

Workshop



Jodi Hadden

RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

TUM
Technische Universität München

Maximilian Scheurer
Marc Siggel
Justin Porter