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
What can we discover with the Computational Microscope?
The range of the Computational Microscope

... Views Living Systems from Electron to Cell

electron/atom
protein folding
ribosome
organelle
cell

Length scale

1 Å 1 nm 10 nm 100 nm 1 μm
Size Matters!

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Atoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>$10^6$</td>
</tr>
<tr>
<td>1994</td>
<td>$10^7$</td>
</tr>
<tr>
<td>1998</td>
<td>$10^8$</td>
</tr>
<tr>
<td>2002</td>
<td>$10^9$</td>
</tr>
<tr>
<td>2006</td>
<td>$10^{10}$</td>
</tr>
<tr>
<td>2010</td>
<td>$10^{11}$</td>
</tr>
<tr>
<td>2014</td>
<td>$10^{12}$</td>
</tr>
</tbody>
</table>

- **Lysozyme** (2 nm)
- **ATP Synthase**
- **HIV Capsid**
- **Photosynthetic Chromatophore** (100 nm)$^3$
- **Ribosome**
- **STMV**
- **Aquaporin**

www.ks.uiuc.edu/~trudack
**Detail Matters too!**

Experiment
X-ray Crystallography
(1.6-2.5 Å)

FTIR-Spectroscopy
(Δν = 1.0 cm⁻¹; 10⁻⁴ Å)

Theory
MD Simulations
(0.01 Å)

Wavenumber / cm⁻¹

GDP
off

GTP
on

Arginín 789

Mg²⁺
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

NAMD
VMD
QwikMD
MDFF
ModelMaker
GSA

Computational Modeling

Structure

NAMD
VMD
QwikMD
Enhanced Sampling
QM/MM

Molecular Dynamics

Function
A Sampling of TCBG’s MDFF Projects

Chemosensory Array

Chromatophore

Rous Sarcoma Virus

Rabbit Hemorrhagic Disease

HIV

26S Proteasome
Integrating experimental methods into computational modeling
The Recycling System of the Cell
The ubiquitin proteasome proteolytic pathway

1. Substrate tagging by Ubq4
2. Ubq4-substrate recognition
3. Substrate degradation

Bortezomib
Kisselev Cancer Cell 2013
Near-atomic model of the 26S proteasome

Cryo-EM density

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

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

max planck institute of biochemistry

www.ks.uiuc.edu/~trudack
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

Chain V of PDB-ID 4CR2

Deubiquitylation (Rpn11)

Missing segments
- highly flexible
- ambiguous density
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

**VMD/NAMMD**

Humphrey *et al.* J. Mol. Graph. 1996

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

- α-Helix
- 3-10 Helix
- Turn
- Coil

Averaged pattern

Near atomic model (4cr2)
Representative 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

Representative model of predicted ensemble for Rpn11’s C-terminal tail

Secondary structure pattern of amino acids 217-306 (purple)
MDFF can be run on Cloud computing for low cost!

MDFF runs can be launched through QwikMD!

[MDFF Tutorial on YouTube 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

cross correlation

0

0.65
Incomplete vs. complete model

Incomplete model

Cross correlation 0.61

Rosetta/MDFF

Complete model

Cross correlation 0.63
Low vs. high resolution density model

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

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

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

Structure predicted for low resolution matches structure of high resolution Secondary structure pattern obtained by Rosetta/MDFF employing a 7.7 Å density Secondary structure pattern of a structure modeled into a 3.5 Å density

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

δ-  Rpn10 (UIM)

δ+  Monoubiquitin

δ+  Tetraubiquitin

www.ks.uiuc.edu/~trudack
Ubiquitin recognition by Rpn10

Ubiquitin Recognition (Rpn10)
Ubiquitin recognition and deubiquitylation

- PDB-ID 2X5N (S. Pombe) - Ubiquitin Recognition (Rpn10)
- PDB-ID 2KDE (human) - Deubiquitylation (Rpn11)
- Ubiquitin Transport
- UIM
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
The conformational space of the Rpn10 linker is highly flexible.
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
The Motor Action of protein unfolding
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.

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

Next QwikMD release will support QM/MM
Converting Chemical Energy into Motor Action

Motor Action of ATP Synthase

Abhi Singharoy
ModelMaker

VMD

- incomplete protein model

- no density
  - de novo structure prediction (Rosetta)
  - model ranking (Rosetta)
  - model filtering (VMD)

- mid-resolution density
  - interactive MDFF (NAMD)

- mid-resolution or high-resolution density
  - backbone and sidechain refinement (Rosetta)
  - iterative real-space refinement (NAMD)
  - highres MDFF

- complete protein model
  - model fitter (ModelFitter)
  - model completer (ModelCompleter)

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

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

www.ks.uiuc.edu/~trudack
Bridging Computation and Experiment

Proteins | Multi-Protein Complex | Proteins + Membrane | Cell
---|---|---|---
NMR | X-ray | cryo-EM/ET | AFM | FRET

Experimental Data

Computational Modeling

Model

Molecular Dynamics

Function

Visualization Analysis

Experimental Validation

---

Experimental Data

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

The Schulten Group

NIH Center for Macromolecular Modeling & Bioinformatics

Theory
Klaus Schulten
Ryan McGreevy

Experiment
Wolfgang Baumeister
Friedrich Förster
Antje Aufderheide

GSA
Rafael Bernardi
Marcelo Mello

ModelMaker
Ryan McGreevy

Workshop
Maximilian Scheurer
Marc Siggel
Justin Porter

Jodi Hadden