Integrative Determination of Macromolecular Structures and Networks

Andrej Sali
http://salilab.org/

Department of Bioengineering and Therapeutic Sciences
Department of Pharmaceutical Chemistry
California Institute for Quantitative Biosciences
University of California, San Francisco
Integrative Determination of Macromolecular Structures and Networks

Andrej Sali
http://salilab.org/

Department of Bioengineering and Therapeutic Sciences
Department of Pharmaceutical Chemistry
California Institute for Quantitative Biosciences
University of California, San Francisco
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

J. D. Watson
F. H. C. Crick

NATURE

April 25, 1953

No. 4356
X-ray diffraction

Composition
Stoichiometry
Chemical complementarity
To understand and modulate cellular processes, we need their models. These models are best generated by considering all available information.
Contents

1. Integrative structure modeling

2. Integrative structure modeling of 26S proteasome
Structural biology:
Maximize accuracy, resolution, completeness, and efficiency of the structural coverage of macromolecular assemblies

Motivation: Models will allow us to understand how machines work, how they evolved, how they can be controlled, modified, and perhaps even designed.

There may be thousands of biologically relevant macromolecular complexes whose structures are yet to be characterized, involved in a few hundred core biological processes.
**Integrative Structural Biology**
for maximizing accuracy, resolution, completeness, and efficiency of structure determination

Use structural information from any source: measurement, first principles, rules; resolution: low or high resolution to obtain the set of all models that are consistent with it.

A description of integrative structure determination


While it may be hard to live with generalization, it is inconceivable to live without it. Peter Gay, Schnitzler’s Century (2002).
Integrative models from our lab

- Ribosomes, Frank, Akey
- PCSK9-Fab, Cheng, Agard, Pons
- Actin, Chiu
- TRiC/CCC, Frydman, Chiu
- RyR channel, Serysheva, Chiu
- Hsp90 landscape, Agard
- Nuclear Pore Complex, Rout, Chait
- Nup84 complex, Rout, Chait
- Nup84 hub, Rout, Chait
- Nup82 complex, Rout, Chait
- Nup133, Rout, Chait
- SEA complex, Rout, Chait, Dokudovskaya
- PDE6, Chu
- Spindle PoleBody, Davis, Muller
- Microtubule nucleation, Agard
- 26 Proteasome, Baumeister
- 40S-eIF1-eIF3, Aebersold, Ban
- PhoQ His kinase, DeGrado
- TFIIH, Ranish
- Prion aggregation, Prusiner
Integrative models from our lab

Ribosomes, Frank, Akey
PCSK9-Fab, Cheng, Agard, Pons
Actin, Chiu
TRiC/CCC, Frydman, Chiu
RyR channel, Serysheva, Chiu
Hsp90 landscape, Agard
Substrate folding by Hsp90, Agard

Nuclear Pore Complex, Rout, Chait
Nuclear Pore Complex transport, Rout, Chait, Aitchison, Chook, Liphardt, Cowburn
Nup84 complex, Rout, Chait
Nup84 hub, Rout, Chait
Nup82 complex, Rout, Chait
Nup133, Rout, Chait
SEA complex, Rout, Chait, Dokudovskaya
PDE6, Chu

Spindle PoleBody, Davis, Muller
Microtubule nucleation, Agard
26 Proteasome, Baumeister
PhoQ His kinase, DeGrado
TFIIH, Ranish
40S-elf1-elf3, Aebersold, Ban
Prion aggregation, Prusiner
Integrative Modeling Platform (IMP)
http://integrativemodeling.org

R. Pellarin, M. Bonomi, B. Raveh, S. Calhoun, C. Greenberg, G.Dong, S.J. Kim, I. Chemmama

Open source, versions, documentation, wiki, examples, mailing lists, unit testing, bug tracking, ...

Simplicity
Chimera
Domain-specific
restrainer
PMI

Flexibility

IMP C++/Python library

Representation:
- Atomic
- Rigid bodies
- Coarse-grained
- Multi-scale
- Symmetry / periodicity
- Multi-state systems

Scoring:
- Density maps
- EM images
- Proteomics
- FRET
- Chemical and Cys cross-linking
- Homology-derived restraints
- SAXS
- H/D Exchange
- Native mass spectrometry
- Genetic interactions
- Statistical potentials
- Molecular mechanics forcefields
- Bayesian scoring
- Library of functional forms (ambiguity, ...)

Sampling:
- Simplex
- Conjugate Gradients
- Monte Carlo
- Brownian Dynamics
- Molecular Dynamics
- Replica Exchange
- Divide-and-conquer
- enumeration

Analysis:
- Clustering
- Chimera
- Pymol
- PDB files
- Density maps
Integration across computational resources

Goal: Maximize accuracy, resolution, completeness, and efficiency of the structural coverage of macromolecules

---

Experiment

Hypothesis

Model
Outcome of the First Hybrid / Integrative Methods Task Force Workshop


We describe the proceedings and conclusions from the first Integrative Methods Task Force Workshop that was held at the European Bioinformatics Institute in Hinxton, UK, on October 6 and 7, 2014. At the workshop, experts in the various experimental fields that are contributing to these integrative studies, experts in integrative modeling, and experts in data archiving addressed a series of central questions. What data should be archived? How should integrative models be represented? How should the data and integrative models be validated? How should the data and models be archived? What information should accompany the publication of integrative models?
Pushing the envelope of structural biology by integration of all available information

• Size

• Static systems in single and multiple states

• Dynamic systems

• Bulk and single molecule views

• Impure samples

• Overlapping with other domains such as systems biology
Contents

1. Integrative structure modeling

2. Integrative structure modeling of 26S proteasome
The 26S proteasome acts at the end of the ubiquitin proteasome pathway

Bohn S. and Förster F. *Handbook of Proteolytic Enzymes*, 2012
The 26S proteasome architecture

Bohn S. and Förster F. *Handbook of Proteolytic Enzymes*, 2012
How to determine the molecular architecture of the complete 26S proteasome?

The 26S proteasome has been refractive to single methods for many years, presumably because of conformational and compositional heterogeneity:

• dissociation of the 19S particle into heterogeneous subcomplexes during purification and concentration,

• presence of proteasome interacting proteins,

• conformational variability of some 19S subunits.
Mapping the Phase Space of Models for Transport through the NPC

Gathering information and translation into spatial restraints

- **X-ray, NMR, homology modeling**
  - Component atomic models

- **Electron microscopy**
  - Overall shape, component positions

- **Chemical cross-linking**
  - Protein-protein contacts

- **Proteomics**
  - Protein-protein proximities
RP components and their representation

AAA-ATPase hexamer ring

homology model based on PAN structure (Bohn et al, PNAS, 2010).

PC-repeat containing subunits

 PCI containing subunits

 MPN containing subunits

Ubiquitin receptors
RP components and their representation

AAA-ATPase hexamer ring

homology model based on PAN structure (Bohn et al, PNAS, 2010).

PC-repeat containing subunits

Rpn1
Rpn2

PCI containing subunits

Rpn3
Rpn5
Rpn6
Rpn7
Rpn9
Rpn12

MPN containing subunits

Rpn8
Rpn11

Ubiquitin receptors

Rpn10
Rpn13

precision, efficiency, availability

Atomic
Fixed coarse
Flexible coarse
Hybrid
RP components and their representation

AAA-ATPase hexamer ring

PC-repeat containing subunits
- Rpn1
- Rpn2

PCI containing subunits
- Rpn3
- Rpn5
- Rpn6
- Rpn7
- Rpn9
- Rpn12

MPN containing subunits
- Rpn8
- Rpn11

Ubiquitin receptors
- Rpn10
- Rpn13

Restraints: Geometric complementarity
- Excluded volume

Restraints: Excluded volume

Restraints: Chain connectivity
- Radius of gyration
- Excluded volume

Restraints: Fixed coarse

Restraints: Flexible coarse

Restraints: Hybrid

X-ray, NMR, homology modeling

AAA-ATPase hexamer ring

homology model based on PAN structure (Bohn et al, PNAS, 2010).

precision, efficiency, availability
Cryo-EM map of the *S. pombe* 26S proteasome

Particles: 375,000
Symmetry: C2
Increment: 0.5°
FSC @ 0.5: 8.4 Å
FSC @ 0.3: 7.1 Å

F. Foerster, S. Bohn, W. Baumeister
Cryo-EM map of the *S. pombe* 26S proteasome

Particles: 375,000
Symmetry: C2
Increment: 0.5°
FSC @ 0.5: 8.4 Å
FSC @ 0.3: 7.1 Å

Restraints: Cross-correlation between a model and the map

F. Foerster, S. Bohn, W. Baumeister
Cryo-EM of knockout mutants localizes Rpn10 and Rpn13

Cryo-EM of knockout mutants localizes Rpn10 and Rpn13

Restraints: Positions of Rpn10 and Rpn13 are fixed while sampling other subunits.

Fitting of *D. melanogaster* Rpn6 X-ray structure into the cryo-EM map localizes Rpn6

Fitting of *D. melanogaster* Rpn6 X-ray structure into the cryo-EM map localizes Rpn6

Structure - map cross-correlation

Fitting of *D. melanogaster* Rpn6 X-ray structure into the cryo-EM map localizes Rpn6

Fitting of *D. melanogaster* Rpn6 X-ray structure into the cryo-EM map localizes Rpn6


Restrains: Position of Rpn6 is fixed while sampling other subunits.
Fitting of *D. melanogaster* Rpn6 X-ray structure into the cryo-EM map localizes Rpn6


Restraints: Position of Rpn6 is fixed while sampling other subunits.

Similarly, for the AAA-ATPase Rtp1-6 heteromeric ring (*Bohn et al., PNAS*, 2010).
Cross-linking / mass spectrometry data


Disuccinimidyl suberate (DSS)
Cross-linking / mass spectrometry data


Inter-molecular cross-linking of exposed Lys residues:

- 12 Rpt-Rpn residue-specific crosslinks (*S.p.*)
- 3 Rpn-Rpn residue-specific crosslinks (*S.p.*)

Disuccinimidyl suberate (DSS)
Cross-linking / mass spectrometry data


Inter-molecular cross-linking of exposed Lys residues:
- 12 Rpt-Rpn residue-specific crosslinks (*S.p.*)
- 3 Rpn-Rpn residue-specific crosslinks (*S.p.*)

Restrains: upper distance bounds on cross-linked atoms or beads.
Sampling good-scoring 19S structures
Sampling good-scoring 19S structures

Discretization

discretization of the map into 238 anchor points
Sampling good-scoring 19S structures

discretization of the map into 238 anchor points

localization of coarse subunit models, subject to proteomics data

**enumeration** of all configurations with at most 5 violations
Sampling good-scoring 19S structures

**Discretization**
- Discretization of the map into 238 anchor points

**Localization**
- Localization of coarse subunit models, subject to proteomics data
- Enumeration of all configurations with at most 5 violations

**Fitting**
- Local rigid body fitting of alternative atomic subunit models
- Selection of best subunit models by fitting quality
Sampling good-scoring 19S structures

Discretization
- discretization of the map into 238 anchor points

Localization
- localization of coarse subunit models, subject to proteomics data
- enumeration of all configurations with at most 5 violations

Fitting
- local rigid body fitting of alternative atomic subunit models
- selection of best subunit models by fitting quality

Refinement
- atomic model refinement subject to cross-linking and position restraints

Elizabeth Villa
Sampling good-scoring 19S structures

**Discretization**
- Discretization of the map into 238 anchor points

**Localization**
- Localization of coarse subunit models, subject to proteomics data
  - **Enumeration** of all configurations with at most 5 violations

**Fitting**
- Local rigid body fitting of alternative atomic subunit models
  - Selection of best subunit models by fitting quality

**Refinement**
- Atomic model refinement subject to cross-linking and position restraints

Elizabeth Villa
Ensemble of ~0.5 million best-scoring models
Ensemble of ~0.5 million best-scoring models
Ensemble of ~0.5 million best-scoring models
Ensemble of ~0.5 million best-scoring models

(a) Number of violated restraints vs. number of models

(b) Centroid RMSD [Å] vs. model correlation matrix

(c) Correlation across all models

(d) Cluster 1, 2, and 3 models for Rpn1 to Rpn12
Molecular architecture of the 26S proteasome


Julio Ortiz
Molecular architecture of the 26S proteasome


Julio Ortiz
Interpretation of the 26S structure evolution, function, modulation

- **A**: PC-repeat containing proteins, Ubiquitin receptors, MPN containing proteins, PCI containing proteins.
- **B**: Rpn1, Rpn2, Rpn13, Rpn10, Rpn9, Rpn5, Rpn6, Rpn7, Rpn7, Rpn3, Rpn12.
- **C**: Rpn1, Rpn2, Rpn13, Rpn10, Rpn9, Rpn5, Rpn6, Rpn9.
- **D**: Rpn12, Rpn3, Rpn7, Rpn5, Rpn6.
1. Determination of assembly *structures* and mapping of *networks* benefit greatly from the inclusion of all available information.

2. Developers and users of open source *Integrative Modeling Platform (IMP)* are most welcome.

3. Molecular architecture and function of the *26S proteasome*. 

**Summary**
Acknowledgements

Integrative modeling
Seung Joong Kim
Peter Cimermancic
Barak Raveh
Ben Webb
Charles Greenberg
Dina Schneidman
Sara Calhoun
Daniel Saltzberg
Shruthi Viswanath
Ilan Chemmama
Seth Axen
Elina Tjioe
Daniel Russel
Max Bonomi
Riccardo Pellarin
Frank Alber
Bret Peterson
Maya Topf
Mike Rout (RU)
Brian Chait (RU)
David Agard (UCSF)
Tom Ferrin (UCSF)
Trisha Davis (Univ of Wash)
Mark Winey (U Colorado)
Ivan Rayment (U Wisconsin)
Wah Chiu (Baylor)
Nevan Krogan (UCSF)
Robert Stroud (UCSF)
Roger Kornberg (Stanford)
Stanley Prusiner (UCSF)
Jeff Ranish (ISB)
Haim Wolfson (TAU)
Nenad Ban (ETH)
...

26S
Karen Lasker

Wolfgang Baumeister
Friedrich Foerster
Stephan Bohn
Elizabeth Villa
Pia Unverdorben
Florian Beck
Ganesh Pathare
Eri Sakata
Stefan Nickell
Andreas Bracher
Julio Ortiz (movie)

Ruedi Aebersold
Thomas Walzthoeni
Alexander Leitner
Martin Beck

Carol Robinson
Florian Stengel

wwPDB Hybrid/Integrative Methods Task Force
Helen Berman
Torsten Schwede
Jill Trewhella
Stephen Burley
Gerard Kleywegt
Frank Dimao
Klaus Schulten
Jens Meiler
Gerhard Hummer

Funding
NIH, NSF