Integrative structure determination of macromolecular assemblies

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
Disseminating structural models

Publishing models in a **printed paper**

Depositing models in a **computer database**

Depositing **input data** in a computer database

Depositing modeling **protocols** for converting data to models

**Enable** others to interact with data and models: test, improve, use data and models
Types of structural models (static and dynamic):

- **information**: X-ray, NMR, EM, and SAXS structures; “theoretical” models; hybrid models
- **representation**: atomic, coarse-grained, multi-scale models

**PDB** is a natural facilitator of establishing conventions, standards, interfaces, assessment criteria, publication criteria, *etc*, thus catalyzing a collaborative community
Contents

1. Integrative (hybrid) structure determination
2. Fitting multiple subunits into an EM map subject to restraints from proteomics
3. Structure of the yeast Nup84 complex
Integrative determination of macromolecular structures 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.
An approach to integrative structure determination

Integrative Modeling Platform (IMP)


- IMP-1.0 available at [http://salilab.org/imp/](http://salilab.org/imp/) (3/10/10)
- Open source, SVN, documentation, wiki, examples, mailing lists, unit testing, bug tracking, ...

Simplicity → Flexibility

- Chimera tools/web apps
- Domain-specific applications
- IMP C++/Python library
- Restrainer

Model

- Angle restraint
- Conjugate gradients
- Distance score
- Nonbonded list
- IO
- Cross correlation
- Connectivity restraint
- Domin
- SAXS score
- Volume restraint
- Monte Carlo
- Particle
- Rigid body
- Harmonic
Configuration of 456 proteins in the Nuclear Pore Complex with M. Rout & B. Chait

Protein Shape

Protein Stoichiometry

Affinity Purification
Overlay Assay
75 composites 7 contacts

Protein-protein Proximities

Protein Localization

Bioluminescence Imaging

Electron Microscopy

Symmetry

Ultracentrifugation

30 S-values 1 S-value

Bioinformatics and Membrane Fractionation

30 protein sequences

Quantitative Immunoblotting
30 relative abundances

Determination by experiment versus prediction by modeling

Integrative structure determination

NMR spectroscopy

EM microscopy

X-ray crystallography

Monday, November 7, 11
Contents

1. Integrative (hybrid) structure determination
2. Fitting multiple subunits into an EM map subject to restraints from proteomics
3. Structure of the yeast Nup84 complex
Assembly architecture from atomic structures of subunits, EM density map of assembly, and proteomics

Protein Data Bank

EM Data Bank

BioGrid, ...

12
Fitting multiple subunits into a density map: Scoring

**Input:**
- Atomic, coarse components
- Low resolution density map of the assembly
- Proteomics data

**Output:**
- Assembly configuration

**Find assembly configurations that satisfy:**

- Shape complementarity
- Quality-of-fit
- Envelope protrusion
- Connectivity

Optimization / sampling

**Divide-and-Conquer** (DOMINO)

1. **Represent** the scoring function as a graph.

\[
\begin{align*}
F(y_1, ..., y_8) &= \alpha_2(y_2) + \alpha_6(y_6) + \alpha_7(y_7) \\
&+ \beta_{1,2}(y_1, y_2) + \beta_{1,3}(y_1, y_3) + \beta_{1,4}(y_1, y_4) + \beta_{1,5}(y_1, y_5) \\
&+ \beta_{2,7}(y_2, y_7) + \beta_{2,8}(y_2, y_8) + \beta_{3,6}(y_3, y_6) + \beta_{3,8}(y_3, y_8) \\
&+ \beta_{4,7}(y_4, y_7) + \beta_{5,7}(y_5, y_7) + \beta_{7,8}(y_7, y_8)
\end{align*}
\]

2. **Decompose** the set of variables into relatively decoupled subsets (a junction tree algorithm).

3. **Optimize** each subset independently by a traditional optimizer, to get the optimal and a number of suboptimal solutions.

4. **Gather** subset solutions into the best possible global solutions (message passing algorithms; eg, belief-propagation).

---


**Proof-of-principle:** Integrative structure determination of human RNAPII

Lasker *et al*, MCP 2010

---

**Data Gathering**
- Rpb1
- Rpb2
- Rpb7
- Rpb8
- Rpb9
- Rpb10
- Rpb11
- Rpb12
- Rpb3
- Rpb4
- Rpb5
- Rpb6

**Data Translation into Spatial Restraints**
- Comparative models
- EM density map
- Proteomics data
- Geometric complementarity
- EM quality of fit
- Pairwise distance from proteomics
- Connectivity from proteomics

**Optimization & Analysis**
- Density map segmentation
- MultiFit optimization of EM quality of fit and geometric complementarity
- Filtering by proteomics data
- Best scoring configuration
- Resolved affinity purifications

---

Cramer *et al*, *Science*, 2000 (X-ray)
Kostek *et al*, *Structure*, 2006 (EM)
Gavin *et al*, *Nature* 2006 (proteomics)
Krogan *et al*, *Nature*, 2006 (proteomics)
Assessment of an integrative model of human RNAPII

I. atomic representation

reference model

human model

a
c
b
d
Rpb6 Rpb1 Rpb3
Rpb4 Rpb2 Rpb7
Rpb8 Rpb9
Rpb10
Rpb11
Rpb11
Rpb12
Rpb5

II. coarse-grained representation

human model

reference model

e
g
f
h

reference model - human subunit models fit on the corresponding subunits in the crystallographic yeast RNAPII structure
Additional configurational restraints

1. Affinity purification with domain deletion constructs
   Orienting subunits by identification of interacting domains
   J. Phillips; with J. Fernandez, M. Rout:

2. 2D EM class averages
   Filtering models by matching their optimal projections to images
   J. Velazquez, D. Schneidman

3. Assembly subcomplex stoichiometry by native mass spectrometry
   Ambiguous network of protein proximities
   D. Russel, J. Phillips; with A. Politis, C. Robinson:

4. Small Angle X-ray Scattering (SAXS)
   Filtering models by their shape
   D. Schneidman, S.-J. Kim
Contents

1. Integrative (hybrid) structure determination
2. Fitting multiple subunits into an EM map subject to restraints from proteomics
3. Structure of the yeast Nup84 complex
Towards a higher resolution structure of the NPC

Characterize structures of the individual subunits, then fit them into the current low-resolution structure, aided by additional experimental information.

The Nup84 complex in the NPC

- 7-protein complex
- Forms the two outer rings of the NPC
- Present in 16 copies in the NPC
- Proteins share a common ancestor with vesicle coating complexes
Nup84 complex: Representation

Nup84

- 3JRO, 3IKO

Nup133

- 1XKS

Nup145c

- 3IKO, 3BG0

Nup120

- 3F7F, 3HXR

Nup85

- 3EWE, 3F3F

Sec13

- 2PM6, 3F3F, 3JRO

Seh1

- 3F3F, 3EWE

S. cerevisae Nups

Human Nups

Brohawn, Schwartz 2008
Nagy et al. 2009
Berke et al. 2004
Boehmer et al. 2008
Sampathkumar et al.
Nagy et al. 2009
Hsia et al. 2007
Seo et al. 2009
Leska et al. 2009
Brohawn et al. 2008
Debler et al. 2008
Goldberg et al. 2007
Debler et al. 2008
Brownhawn, Schwartz 2008
Debler et al. 2008
Brohawn et al. 2008
Nup84 complex: Data

Subunit positions & orientations
Affinity purifications with domain truncations
J. Fernandez, J. Franke, B. Chait, M. Rout

Small angle X-ray scattering
S.J. Kim, A. Martel, H. Tsuruta, NYSGXRC, J. Tainer

Subunit conformations

Negative stain EM particle averages at ~3nm resolution
R. Diaz, D. Stokes, J. Velazquez

High-throughput crystallography
NYSGXRC, P. Sampathkumar, M. Sauder, S. Burley

Yeast Nup133
Yeast Nup145
Nup84 complex: Optimization

Random starting configuration -> MC/CG Optimization -> Model

- Restraints
- Fitting to 2D Electron Microscopy Maps
- Affinity Purification Domain Mapping
Nup84 complex: Ensemble of good scoring solutions

- 10,000 good scoring structures
- All restraints are satisfied (2D-EM, domain deletion, ...)
- Domain-domain orientations are resolved uniquely.
- Full ensemble precision is ~1 nm
Assessing the well-scoring models

1. Existence of a good-scoring model.
2. Precision of the ensemble of good-scoring models.
3. Check model against unused data (cross-validation).
4. Known precision / accuracy for “similar” cases.
5. Non-random patterns in the model.

Modeling facilitates assessing the data as well as models in terms of precision and accuracy.
Assessment: Agreement with heterodimeric crystallographic structures

- Nup85-Seh1, closest ensemble structure: 3ewe
- Nup84-Nup145c, closest ensemble structure: 3iko
- Nup145c-Sec13, closest ensemble structure: 3bg0
Towards a near-atomic structure of the NPC

Nup84 complex

NPC

16 x
1. Assembly structure determination benefits greatly from the inclusion of all available information, including heterogeneous data sources.

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

3. General and efficient assembly of subunit models based on domain deletion pullouts, 2D EM projections, 3D EM maps, SAXS profiles, and native MS.

4. Near atomic model of the Nup84 complex.
Acknowledgments

QB3 @ UCSF

Keren Lasker (DOMINO)  
Jeremy Phillips (NPC)  
Seung Joong Kim (NPC)  
Daniel Russel (IMP)  
Javier Velazquez (2D EM)  
Ben Webb (IMP)  
Massimiliano Bonomi (SPB)  
Charles Greenberg (EM)  
Riccardo Pellarin (proteomics)  
Elina Tjoe (IMP)  
Dina Schneidman (SAXS)  
Peter Cimermancic  
Natalia Khuri  

Former members:

Frank Alber (USC)  
Friederich Förster (MPI)  
Damien Devos (EMBL)  
Maya Topf (Birkbeck College)  
Narayanan Eswar (Du Pont)  
Marc Marti-Renom (Valencia)  
Mike Kim (Google)  
Dmitry Korkin (UM, Columbia)  
Fred Davis (HHMI)  
M. Madhusudhan (Singapore)  
D. Eramian (UCSF)  
Min-Yi Shen (Applied Biosys)  
Bret Peterson (Google)  

Rockefeller University

Mike Rout  
Javier Fernandez-Martinez  
Loren Hough  
John LaCava  
Jody Franke  
Jaclyn Novatt  

Brian Chait  
Matthew Sekedat  
Rosemary Williams  

John Aitchison (ISB)  
David Stokes (NYSBC)  
Chris Akey (BU)  
Robert Stroud (UCSF)  
Stephen Burley (Lilly)  
Steven Almo (AECOM)  
Hiro Tsuruta (Stanford)  
John Tainer (BNL)  

Wolfgang Baumeister (MPI)  
Trisha Davis (Univ of Wash)  
Tom Ferrin (UCSF)  
Haim Wolfson (TAU)  
David Agard (UCSF)  
Wah Chiu (Baylor)  
Joachim Frank (Columbia)  
Nevan Krogan (UCSF)  
Al Burlingame (UCSF)  
Carol Robinson (Cambridge)  

NIH, NSF  
The Sandler Family Foundation  
Human Frontiers Science Program  
IBM, Intel, Hewlett-Packard, NetApps, Pfizer, Structural Genomix Pharmaceuticals, Mike Homer, Ron Conway