Conformational States of Macromolecular Assemblies Explored by Integrative Structure Calculation

Konstantinos Thalassinos,1 Arun Prasad Pandurangan,2 Min Xu,3 Frank Alber,3,* and Maya Topf2,*

1Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, UK
2Institute of Structural and Molecular Biology, Birkbeck College, University of London, London WC1E 7HX, UK
3Molecular and Computational Biology, University of Southern California, 1050 Childs Way, Los Angeles, CA 90089, USA
*Correspondence: alber@usc.edu (F.A.), m.topf@cryst.bbk.ac.uk (M.T.)
http://dx.doi.org/10.1016/j.str.2013.08.006
Why Study Protein Complexes?

- Oligomerization state
- Number of bound ligands
- Subunit stoichiometry
- Dynamics

Clare, D.K., et al. ATP-triggered conformational changes delineate... Cell 2012.
How can we use multiple techniques to probe the composition and dynamics of molecular complexes?
Mass Spectrometry

- Measures mass-to-charge ($m/z$) ratio of ionized species
- Ion source induces formation of gas-phase peptide ions
- Peptides are separated based on $m/z$ ratios
- Detector reads the number of ions at each $m/z$ ratio
- Protein characteristics can be extrapolated by mapping peptide fragments to original protein
Probing Dynamics with Mass Spectrometry

Chemical Crosslinking + MS (XL-MS)

• Examine non-covalent interactions between proteins or within a protein based on proximity

• Provides distance constraints for structural models

• Solution-based: natural fold, multiple conformations

• Challenge: high number of possible crosslinked peptides

Holding, A. XL-MS: Protein cross-linking coupled with mass spectrometry. Methods. 2015
Ion-mobility MS (IM-MS)

- Separate different conformations of the same protein-protein complex
- The rotationally averaged collision cross-section (CCS; shape) of complex affects ion mobility

Lanucara, et al. The power of ion-mobility... Nature Chemistry. 2014
Hydrogen-deuterium exchange coupled to MS (HDX-MS)

• Detect N-H, O-H, S-H exchange with deuterium in D$_2$O by MS

• Provides information about structural flexibility due to protein folding/unfolding, conformational changes, hydrogen bonding, or solvent exposure

• Recent improvements in instrumentation, sample preparation, and data analysis

Figure 1A
Cryo-electron tomography (Cryo-ET)

3-D images of the sample are from a reconstructed series of subtomograms collected at different tilt angles.

Cryo-electron tomography techniques

- *In situ* structure determination of complexes
- High noise level, crowded cellular compartments
- Subtomogram image processing is computationally intensive

Cryo-electron microscopy (Cryo-EM)

Reconstruct 3-D structures from a series of images of the protein of interest taken at different angles

Low resolution (5-25Å) EM density maps

- Fit atomic models from X-ray crystallography, NMR, structure prediction, etc into EM density maps

- Multiple factors: model accuracy, electron density resolution, scoring functions, number of components, etc

- Rigid fitting: 6 degrees of freedom to fit atomic model to density map

- Flexible fitting: Deform atomic model using molecular mechanics force field and forces that match electron density map

Figure 1C, D
Example: Integrative modeling of the 26S proteasome holocomplex

• Data: cryo-EM density map, 12 residue-specific cross-links, interactions from proteomic studies, atomic structures of individual subunits

• Convert data into spatial restraints

• Localize subunits (spatial restraints; message passing algorithm) -> fit subunits (MultiFit) -> flexible fitting refinement (molecular dynamics flexible fitting, MDFF)

• Evaluate models based on restraints, completeness, similarity between models, non-integrated structural data

“Multiple comparative models were built for each subunit, fitted into the cryo-EM map, and the best-scoring configurations were subjected to flexible fitting.”
Putting it all together

• Use a scoring function to integrate MS data and distance constraints

• XL-MS spatial constraints have been combined with 3D density fitting and 2D class-average images to improve structure determination

• Data can also be used to validate/support model structures (IM-MS)

• Resolution of data determines how well it can be integrated
QUESTIONS?
Determining the architectures of macromolecular assemblies

Frank Alber\textsuperscript{1*}, Svetlana Dokudovskaya\textsuperscript{2*†}, Liesbeth M. Veenhoff\textsuperscript{2*†}, Wenzhu Zhang\textsuperscript{3}, Julia Kipper\textsuperscript{2†}, Damien Devos\textsuperscript{1†}, Adisetyantari Suprapto\textsuperscript{2†}, Orit Karni-Schmidt\textsuperscript{2†}, Rosemary Williams\textsuperscript{2}, Brian T. Chait\textsuperscript{3}, Michael P. Rout\textsuperscript{2} & Andrej Sali\textsuperscript{1}
Nuclear Pore Complex

Why do we care?
- Allows the transport of molecules in and out of the nucleus of a cell
- **2000** Nuclear Pore Complexes in a single vertebrate cell
- Structure? $\rightarrow$ **Function!**

Why is solving the structure such a challenge?
- **30** distinct proteins, total of **456** proteins
- **X-ray Crystallography** and **NMR** could not handle such a large complex
- **Lower-resolution** methods do not give us atomic information
Presentation Outline:

1. Data Generation
2. Integrating the Data
   • Optimization
   • “Ensemble Analysis”
3. Validation
4. Strengths and Weaknesses of Study
Data Generation

1. Components List: Previously determined

Structure of the Nuclear Pore Complex

Data Generation

2. Shape and Size: Ultracentrifugation

High speed spinning:
Data Generation

3. Amount of each protein (stoichiometry): Immunoblotting
Data Generation

4. Low resolution position of proteins: Immuno-Electron Microscopy
Data Generation

5. Overall Shape: Cryo-Electron Microscopy

Cryo-EM bird’s eye view

Cartoon Reconstructed Side View
Data Generation

6. Protein Connectivity: Affinity Purification-Mass Spectrometry (AP-MS)

Tagged Protein in Solution → Immunoprecipitation → Gel Visualization → Identification by Mass-Spectrometry
Data Generation

6. Protein Connectivity: Affinity Purification-Mass Spectrometry (AP-MS)
Integrating the Data: Optimization

1. Components List
2. Shape and Size of Proteins
3. Amount of each protein
4. Localization of each type of Protein
5. Overall Shape
6. Protein Connectivity

\[ p(C/I) = \prod_f p_f(C/I_f) \]

Structural information from the Experimental Data
Cartesian Coordinates of components
Integrating the Data: Ensemble Analysis

- Goal: 1000 possible structures \(<----\) 1 highly probably native structure

Protein Positioning + Protein Contacts = Protein Configuration
Validation

- **Cross-validated:** leaving 10% of the data out still converged on the same global maximum
- **Consistent...**with unused data including, but not limited to:
  - electron microscopy images
Strengths
• Experimental data difficult to generate
• Questionable transferability of methods
• Plethora of experimental data
• Novel use of data types and integration methods
• Really well written

Weaknesses
Beyond this paper...
• Evaluation biological implication of the solved structure
• Application on a different macromolecular complex
• Make it high-throughput? 😞

“Indeed, it is hard to conceive of any combination of errors that could have biased our structure towards a single solution…”
Questions?
Molecular architecture of the yeast Mediator complex

Philip J Robinson\(^1\), Michael J Trnka\(^2\), Riccardo Pellarin\(^3,4\), Charles H Greenberg\(^3\), David A Bushnell\(^1\), Ralph Davis\(^1\), Alma L Burlingame\(^2\), Andrej Sali\(^3\), Roger D Kornberg\(^1\)*

\(^1\)Department of Structural Biology, Stanford University School of Medicine, Stanford, United States; \(^2\)Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, United States; \(^3\)Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, United States; \(^4\)Structural Bioinformatics Unit, Institut Pasteur, CNRS UMR 3528, Paris, France
Mediator

- Massive protein complex conserved in eukaryotes
- Essential for RNA pol II transcription and regulation of transcription
Mediator

- Massive protein complex conserved in eukaryotes
- Essential for RNA pol II transcription and regulation of transcription

What was previously known about it?

- Three modules: **Head**, **Middle**, and **Tail**
- Structural rearrangement occurs upon interaction with RNA pol II
  - In the RNA polymerase holoenzyme, RNA pol II is surrounded by the Head and Middle modules
Mediator
- Massive protein complex conserved in eukaryotes
- Essential for RNA pol II transcription and regulation of transcription

What was previously known about it?
- Three modules: Head, Middle, and Tail
- Structural rearrangement occurs upon interaction with RNA pol II
- In the RNA polymerase holoenzyme, RNA pol II is surrounded by the Head and Middle modules
Goal of paper:
Determine mediator’s’s structure
Integrative Structure Determination

Four stages of Integrative Structure Determination:
1. Gathering data
2. Representing and translating the data into spatial restraints
3. Sampling the conformational space and identifying good scoring solutions
4. Analyzing and assessing the ensemble of solutions

Integrative modeling platform (IMP) software used for integrative structure determination
Integrative Structure Determination

Data:
XS-MS dataset, X-ray crystal structures, homology models, cryo-EM density maps
Integrative Structure Determination

Data:
XS-MS dataset, X-ray crystal structures, homology models, cryo-EM density maps

Spatial restraints:
Constructed a set of 21 Mediator subunit model representations
Selected model that best “docked” into cryo-EM maps
Integrative Structure Determination

Data:
XS-MS dataset, X-ray crystal structures, homology models, cryo-EM density maps

Spatial restraints:
Constructed a set of 21 Mediator subunit model representations
Selected model that best “docked” into cryo-EM maps

Sampling the conformational space:
165,523 Mediator model configurations produced
Selected the 500 best and grouped them into four clusters (based on RMSD)
XL-MS Results

Note: XL-MS was performed on Mediator when Mediator was in the form of a complex with pol II (holoenzyme)
The N-Termini of Med14 and Med17 are components of the Mediator Middle module.
Inter-modular crosslinks not considered further:
- Inconsistent with cryo-EM density map of free Mediator.
- Modules move upon binding RNA pol II.
Integrative Structure Determination: Cluster 1

Eventually narrowed down to one cluster:
- Cluster 3 had Mediator with orientation inverted (which is inconsistent with prior experiments).
- **Cluster 1** also showed the best cross-link satisfaction statistics and had the best average score.
Mediator Model
Mediator Model: Verification

• Verification based on previous data from yeast two-hybrid assays and immunoprecipitation assays.

• Three disagreements:
  • Med3-Med21
  • Med1-Med7
  • Med1-Med5
Mediator Model: **Med17** and **Med14**

- **Med17**: Temperature-sensitive mutations of Med17 have been discovered that abolish all RNA pol II transcription at the restrictive temperature.
- The N-terminus of the Head subunit Med17 acts as an intermolecular bridge by forming an extensive cross-linking network within the Middle module.
Mediator Model: **Med17** and **Med14**
Mediator Model: Med17 and Med14

- Med14 subunit originally identified as a repressor protein in yeast
- Med14 makes extensive contacts with proteins from all three modules, and is the only Mediator subunit that does so
Mediator Model: Med17 and Med14
Mediator & RNA Pol II: Core Initiation Complex

• Core initiation complex: comprising the Mediator Head module, a minimal Middle module, pol II, a nucleic acid scaffold, and the general transcription factors TBP, TFIIB and TFIIF

• Cross-links between Head module and RNA pol II consistent with holoenzyme cryo-EM data.

• Cross-links between Tail and Middle modules to RNA pol II not consistent with holoenzyme cryo-EM data.

• This suggests that with the tail present and the absence of the nucleic acid scaffold and general factors, the Mediator-polymerase holoenzyme has a different configuration.
Mediator & RNA Pol II: Core Initiation Complex

- Core initiation complex: comprising the Mediator Head module, a minimal Middle module, pol II, a nucleic acid scaffold, and the general transcription factors TBP, TFIIB and...
Final Thoughts and Things to Work On

• What has been done:
  • A complete picture of the mediator complex and its modules has been constructed.
  • It has been shown that, in going from the core initiation complex to the holoenzyme, the mediator complex changes conformation to bring the Tail module in contact with RNA pol II.

• What needs to be worked on:
  • Resolve discrepancies with past immunoprecipitation and yeast two-hybrid results
  • A better description on how this model fits in with our understanding of transcription e.g. showing how the conformational change might turn transcription “on” or ”off”
  • Better characterizing the conformational change: What exactly changes?
  • A major weakness was reconciling data from mediator as part of the holoenzyme to data from mediator by itself. Perhaps more consistency in further analysis would be nice.