Integrated Solutions for Structural Biology: Introducing a Powerful New Workflow for Cross-linking Studies in Biology

Presenter:
Rosa Viner
Complexity of the Proteome; Layers of the Proteome

- 5200 protein families (PFAM) with unknown structure
- Recognition of (supra)molecular complexes as the functional subunit of biology in health and disease
- De-constructing and constructing complexes accelerates our understanding of biological function
- Technologies capture structural information for proteins, DNA/RNA, metabolites, drugs

Protein databank January 2017

- No structural information at all
- With less confident comparative model
- With one confident comparative model
- At least one structure (X-ray/NMR)
What Questions Does Structural Biology Answer?

- 3D Structure
- Stoichiometry
- Composition
- Topology (binding partners, cofactors, messengers, etc.)
- Binding affinity
- Dynamics
- Aggregates
- Biological function
Integrative Structural Biology Workflow Requires Many Tools

True Synergy - Combining Mass Spec and Cryo EM result in more complete, more accurate models of ever larger macromolecular complexes.
2017 Breaking News: Nobel Prize for CryoEM SPA Workflow

Vitrification → Sample prep → Screening → High Resolution Acquisition → High resolution data acquisition → Data Processing

Biochemistry

Screening

High Resolution
Acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction

Vitrification

Sample prep

Screening

High Resolution Acquisition

High resolution data acquisition

Data Processing

Processing: 3D reconstruction
MS+ FEI Cryo TEM Workflows

LC/MS/MS
Sample compositional homogeneity
Subunit Composition
Stoichiometry

Native MS
Electron Microscopy
Subunit topology
Atomic model building

HDX, XL-MS
Identification of non-protein associated molecules

Integrative Structural Biology Workflow Requires Many Tools
Thermo Scientific Native Mass Spectrometry Workflow

Since 2013

CE-MS
WAX-MS
SEC-MS

SEC Separation of Dimer and Monomer

MabPac SEC-1

pH=7

Native - MS

Orbitrap Analysis

Exactive EMR
Q Exactive BioPharma UHMR

m/z up to 80 kDa

Thermo Scientific™
BioPharma Finder™ Software
Thermo Scientific Hydrogen/Deuterium Exchange MS Workflow

Inactive Native Protein

H/D Exchange in D₂O

Quenching

In-line Digest

Orbitrap Analysis

H/D-X PAL™ & Chronos

Thermo Scientific™ BioPharma Finder™ Software

Since 2014
Thermo Scientific Cross-linking Mass Spectrometry Workflow

In vivo cross-linked cells
- Lyse and sonicate
- Digest
- Enrich

Orbitrap Analysis

Thermo Scientific™Proteome Discoverer™ 2.2 XlinkX node

Since 2016
Introduction of Cross-Linking Mass Spectrometry

Distance information between two cross-linked groups

- Sidechain length
- Linker length
- Sidechain length

Distance constraint between Cα: 24.2 Å

Enzymatic digest

Crosslinking

Separation

MS analysis

MSMS analysis

DSS
Disuccinimidy l suberate
MW 368.34
Spacer Arm 11.4 Å

DSSO
Disuccinimidy l sulfoxide
MW 388.35
Spacer Arm 10.3 Å
Cross-Linking Measures  Distance Constraints

Yeast Enolase 1
Pierce Portfolio of Chemical Cross-linking Reagents

1994

DSS
Disuccinimidyl suberate
MW 398.34
Spacer Arm 11.4 Å

BS3-d4
Bis(succinimidyl/1,2,2,7,7 suberate-d4
MW 576.45; Crosslink Mass 142.06
Spacer Arm 11.4 Å

2016

DSSO

SDA
(NHS-Diarnino)
MW 225.20
Spacer Arm 3.9 Å
MS-Cleavable Crosslinkers Facilitate MS Analysis

Disuccinimidyl Sulfoxide (DSSO)

Disuccinimidyl Dibutyric Urea (DSBU)

\[ \Delta m = m_L - m_S = m_{A-L} - m_{A-S} = m_{B-L} - m_{B-S} \]

\[ m_A = m_{A-L} - m_L = m_{A-S} - m_S \]

\[ m_B = m_{A-B} - m_L = m_{B-S} - m_S \]

Kao, A et al, MCP, 2011;
Muller, M et al, Anal.Chemistry, 2010
Crosslinking Mass Spectrometry Workflow: MS Analysis

Disuccinimidyl Dibutyric Urea (DSBU)

MeroX

Available since May 2017!
Crosslinking Mass Spectrometry Workflow: MS Analysis

Disuccinimidyl Sulfoxide (DSSO)

Available since November 2016!
Crosslinking Mass Spectrometry Workflow: MS Analysis

E. coli DSSO unique XL-peptides

Liu F. et al. (2017.) Nat. Commun. doi: 10.1038/ncomms15473
Crosslink reaction efficiency

Roughly 5% of inter protein, cross-linkable lysine pairs actually get crosslinked.

SCX fractionation

Not yet reaching the full depth

Still room for grabbing more detail from the interatomic studies.

Slide courtesy of Richard Scheltema, Utrecht University
Enrichment Improves Crosslinked Peptides Identification

Strong Cation Exchange (SCX)

Digest

Buffer Exchange

Fractionate

Desalt

LC-MS

Coming Soon!

Viner et al, ASMS 2017, ThP069
Enrichment Improves Crosslinked Peptides Identification

SEC Superdex Peptide PC column (GE Health)

Quantitative XL-MS

QCLM: Isotopically Labeled XL

QMX: Isobaric Mass Tags + XL-MS


Q MIX: Isobaric Mass Tagging + XL-MS

Tune 3.0 Available Now!

• Fully integrated crosslinked peptide search engine (XlinkX)
• Crosslinked peptide annotation
• Assignment of inter-, intra-crosslinked peptides and mono-adducts
• Compatible with non-cleavable and MS-cleavable crosslinkers

Thermo Scientific Proteome Discoverer 2.2 XlinkX* Node

Processing Workflow

Consensus Workflow

Liu F. et al. (2017.) Nat. Commun. doi: 10.1038/ncomms15473
XLinkX Node: Search

Processing Workflow

Definition of Crosslinks
- Addition of any linker (K-K)

Database search
- Any PTM defined in PD
- Full complexity databases
- Fragmentation: HCD/CID, ETD, EThcD
- Acquisition: MS2, MS2-MS2, MS2-MS3
- Percolator supported XL FDR control
- TMT quantitation*

*For internal use only.
**Algorithm:**

- Deisotope + TopX filter spectrum
- Determine mz-error all de-isotoped peaks
- Matching peaks of same charge based on difference +/- mz-errors
- Prioritize on higher intensity

Selectivity: ~10% FDR
MS2 CID – Which fragments can we expect?!

Algorithm:
- Deisotope + TopX filter
- We match both sets of fragments
- Target / decoy search – score is a now standard likelihood calculation as developed by Olson et al; however it’s fragment intensity weighted
- The lowest score is taken to prevent situations where 1 peptide is well covered and the other unidentifiable

Slide courtesy of Richard Scheltema, Utrecht University
XlinkX Node: Identification - MS3

Algorithm:
- **No** deisotoping **but do** TopX filter
- We match both with standard settings (a[1-2], b, and y)
- Target / decoy search – score is a standard likelihood calculation as developed by Olson et al

Frequent mass differences from backbone cleavage
Potential specific neutral losses?

Frequently detected masses
Potential immonium ions?

Spectral binning; Kelstrup et al; JPR; 2014
Precision mapping of the metabolome; Breitling et al; Trends in Biotech; 2006
FDR processes MS2 and MS3 independently

Percolator based FDR

Score based FDR (Simple)

MS3 features
1. Score
2. Score ratio
3. Mono isotopic mass
4. Charge
5. Delta mass
6. Absolute delta mass
7. Missed cleavages
8. Variable modification ratio
9. Unique variable modifications
10. Number modifiable residues
11. Number modified residues
12. Total intensity
13. Matched intensity
14. Matched intensity percentage
15. Median fragment mass error
16. Median fragment mass ppm error
17. Fraction ions matched [ion-series]
18. Matched ion intensity [ion-series]

MS2 features
1. Score
2. Score ratio
3. Mono isotopic mass
4. Charge
5. Delta mass
6. Absolute delta mass
7. Missed cleavages [A / B]
8. Number matches [A / B]
9. Sequence coverage [A / B]
10. Missed cleavages [A / B]
11. Fraction ions matched [free / linked]
12. Median fragment mass error
13. Total intensity
14. Matched intensity
15. Crosslink position

Percolator; Käll et al; Nat meth; 2007

+30% identifications...
Data set
1. SCX fractionated full lysate (11 fractions)
2. 2hr gradients – mass difference triggered
3. Total 211,916 MS2 scans
4. Total 296,560 MS3 scans

Fasta
1. Uniprot human
2. Total 44,110 protein sequences

Computer
1. 8-core intel i7-4790 processor @ 3.6HhZ (64-bit)
2. 16Gb internal memory
3. 222Gb SSD drive

Max memory used: 3.9 Gb
CPU's used: 7

# identified MS2 scans 1764 (of 7474)
# identified MS3 scans 2835 (of 19698)
# crosslink spectrum matches 1761
# crosslinks 1198

Crosslink search

Left for OS tasks

Slide courtesy of Richard Scheltema, Utrecht University
XLinkX Node: Results

Consensus Workflow

- Peptide Validator
- Peptide and Protein Filter
- Protein Score
- Protein FDR Validator

Quantification (Xlink 2.0)

Full integration in PD
- Linked output tables
- Visualization
- Automatic export to xiNET

Automatic export to 3D structures (Xlink 2.0)
- Supporting pyMol
- Planned Haddock (modelling)
XLinkX Node: Results

- Protein Structure
- Protein Complex
- Interactomics

[Images of protein structures and networks]

XiNET
http://crosslinkviewer.org/
Proteome-wide XL-MS application (*E. coli*)- 1158 Unique XL*

Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification

Fan Liu¹,²,e, Philip Loszl¹,²,e, Richard Scheltema¹,², Rosa Viner³ & Albert J.R. Heck¹,²

*Liu F. et al. (2017.) Nat. Commun. doi: 10.1038/ncomms15473*
Acknowledgements

Ryan Bomgarden
Chris Etienne
Sergei Snovida
Kay Opperman
Mathias Muller
Derek Bailey
Kai Fritzemeier
Bernard Delanghe
Torsten Ueckert
Shijun Li
David Horn
Andreas Huhmer
Terry Zhang
Romain Huguet
Chris Mullen
Alexander Makarov
Linda van Driel
Huiling Liu
Jenny Ho
Vlad Zabrousakov

Universiteit Utrecht

Fan Liu
Richard Scheltema
Albert Heck

University of California, Irvine

Clinton Yi
Craig Gutierrez
Lan Huang