Cross-Linking Assisted Modeling

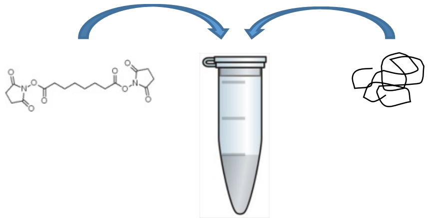
CASP 13 Tutorial
May 24, 2018

Alexander Leitner & Esben Trabjerg
Institute of Molecular Systems Biology, ETH Zurich

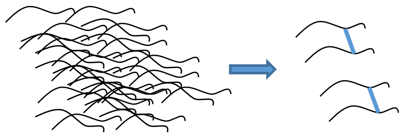
Experimental workflow of a cross-linking experiment

The workflow resembles a conventional proteomics experiment, with some modifications

Sample preparation and cross-linking reaction



Sample work-up: enzymatic digestion, clean-up, enrichment/fractionation (optional)



LC-MS/MS analysis

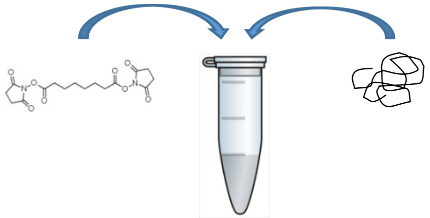
Data analysis using specialized software

Leitner et al., Mol. Cell. Proteomics, 2012
Leitner et al., Nat. Protoc., 2014

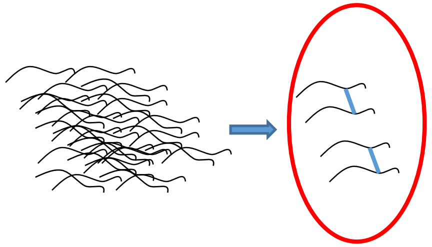
Experimental workflow of a cross-linking experiment

The workflow resembles a conventional proteomics experiment, with some modifications

Sample preparation and cross-linking reaction



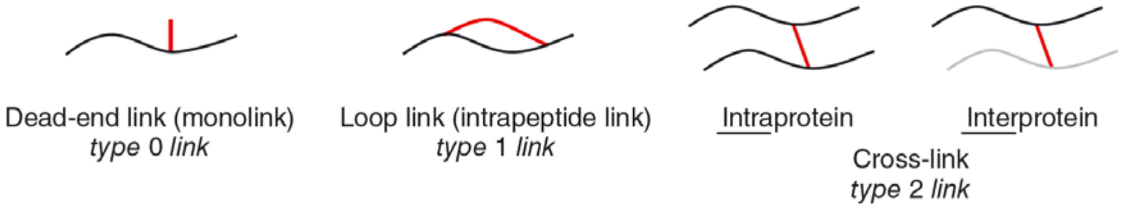
Sample work-up: enzymatic digestion, clean-up, enrichment/fractionation (optional)



Cross-linked peptides that need to be identified

LC-MS/MS analysis

Different products from a XL experiment

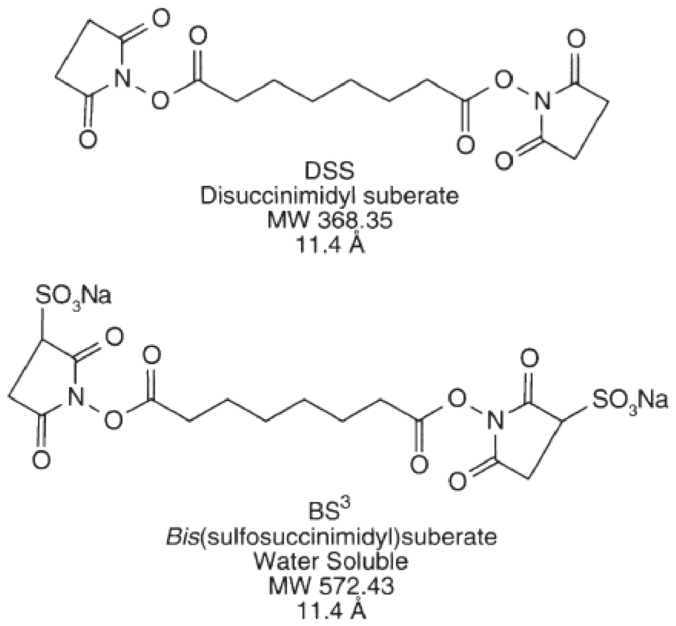


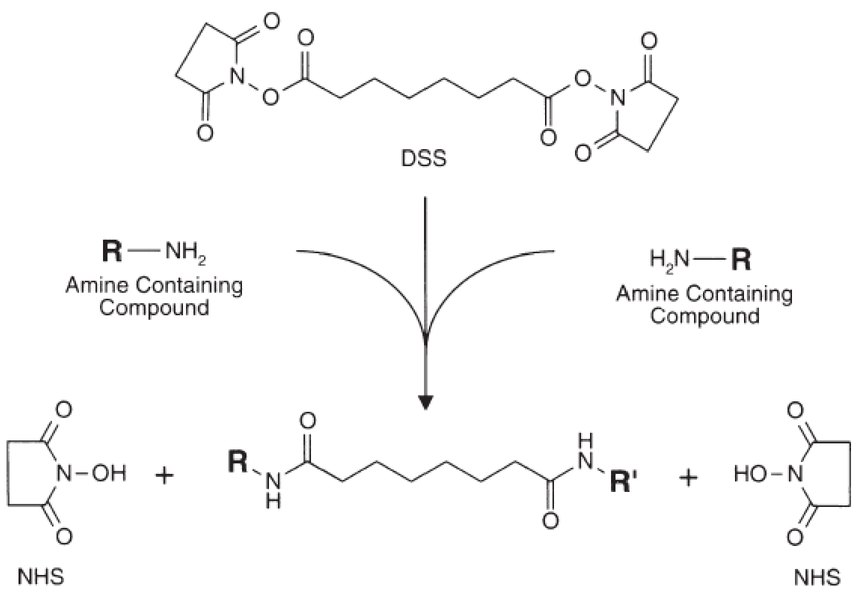
Data analysis using specialized software

Cross-linking chemistries

Cross-linking of primary amines (Lys, N-terminus) using succinimide esters, e.g. DSS, BS^3

- Most widely used chemistry in XL-MS
- Side-reactions with Ser/Thr/Tyr possible



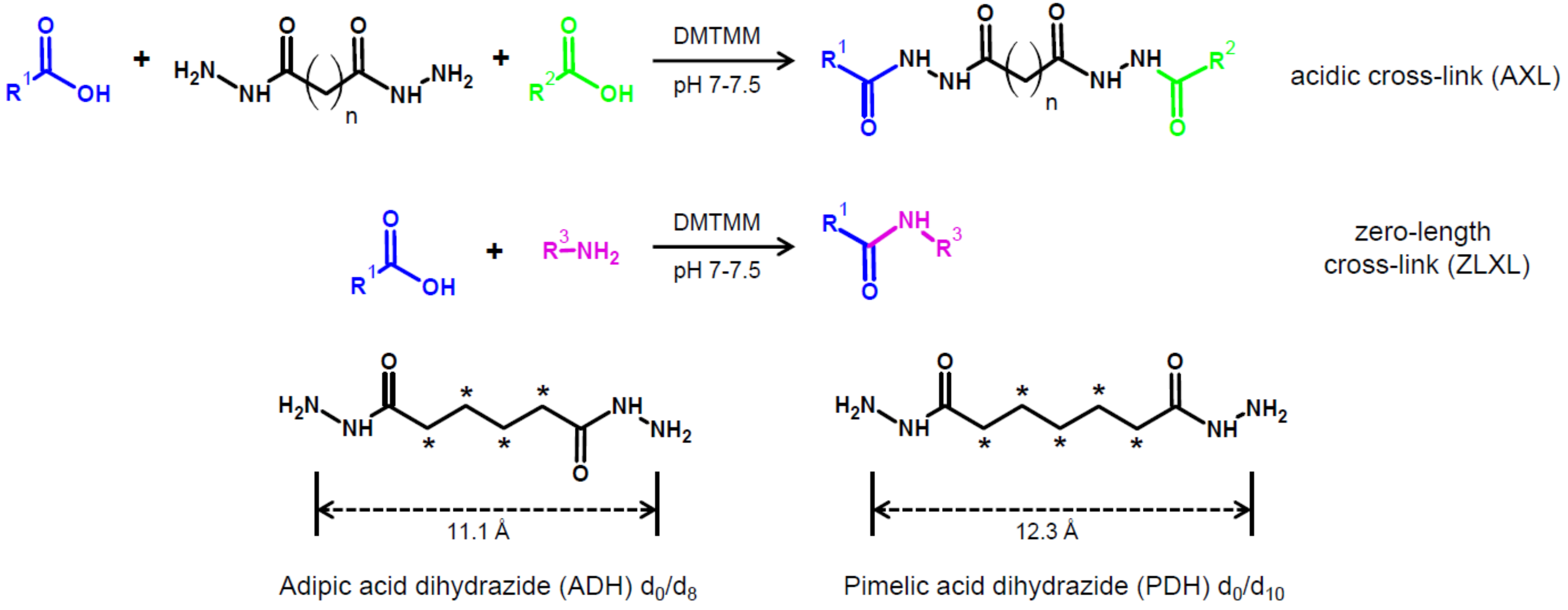


Hermanson, *Bioconjugate Techniques*

Cross-linking chemistries

Cross-linking of carboxyl groups (Asp, Glu, C-terminus) and of primary amines with carboxyl groups (without spacer)

- Combined reaction will yield two different reaction products
- Lower reaction yields, success depends more on target protein (complex)



Leitner et al., PNAS, 2014

Experimental considerations

To reflect the native state of the protein (complex), experimental conditions need to be controlled, e.g.

- Protein concentration
- Excess of reagent
- Buffer pH and composition
- Temperature

Yield of the cross-linking reaction will depend on the parameters listed above, but also on

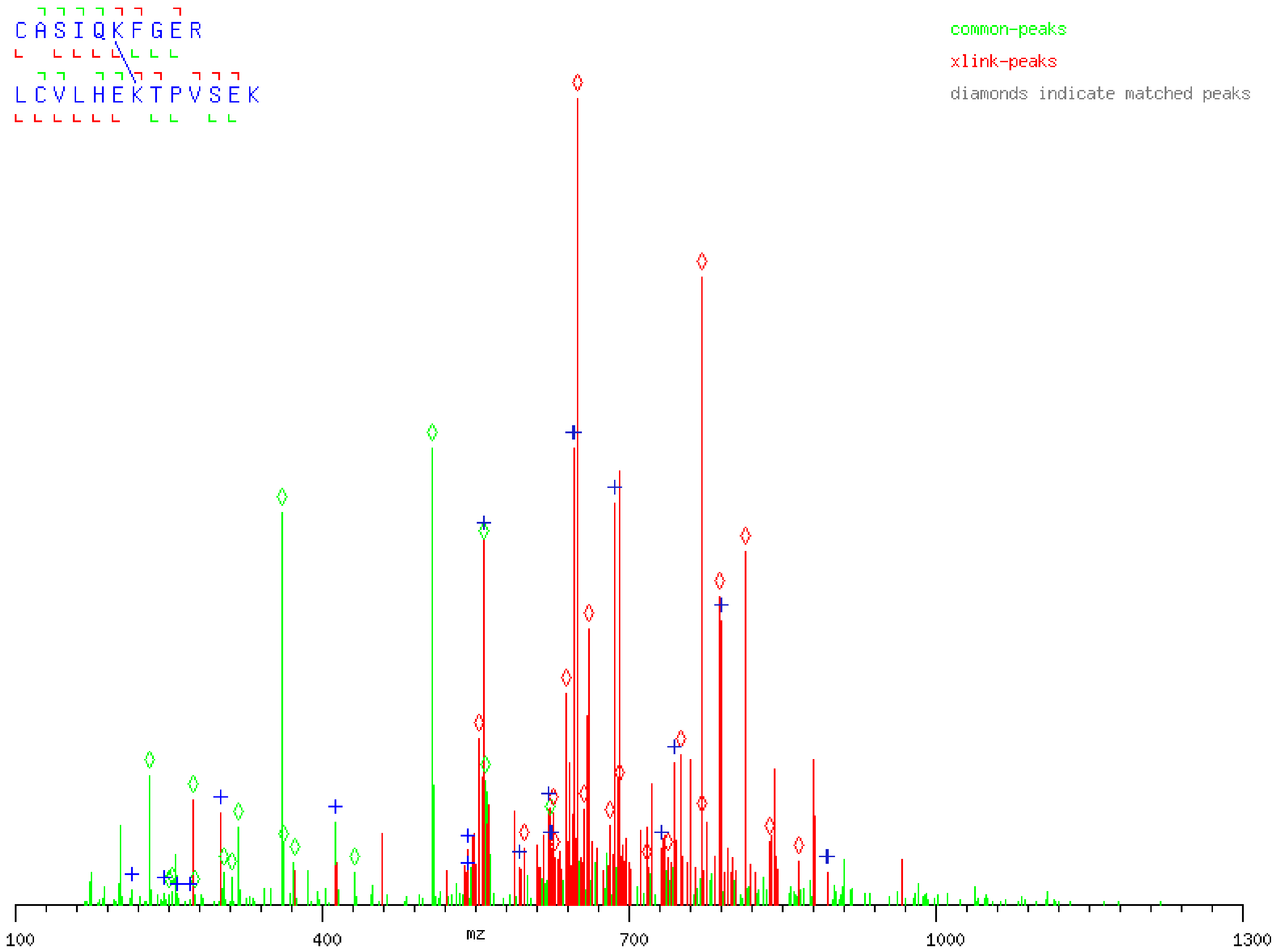
- Exposed (and reactive) target residues, for cross-linking to occur (mainly Lys)
- Sufficient size of the binding interface, to be able to probe intersubunit interactions
- Distribution of reactive sites and protease cleavage sites, for MS identification

In summary, not all structurally plausible contacts will be identified and data will be sparse!

Homo-oligomers provide ambiguous structural information (intra- or inter-subunit cross-links cannot be discriminated unless the sequences of the two peptides are identical or overlapping)

Computational analysis steps

MS/MS spectra of cross-linked peptides typically contain fragment ions from both chains

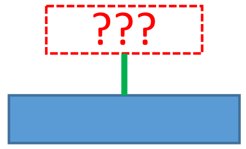


Computational analysis steps

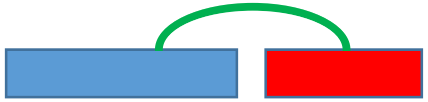
To deal with spectral complexity, different computational/bioinformatic strategies have been proposed

Linearization of the peptide sequences

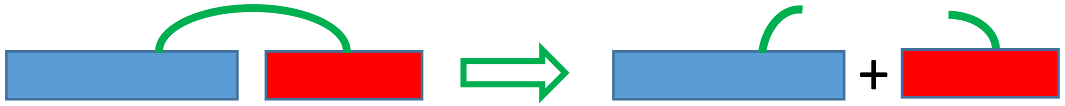
Treating one peptide as a **modification with unknown mass** of the other



Predicting and scoring **actual pairs of peptides** connected by the linker



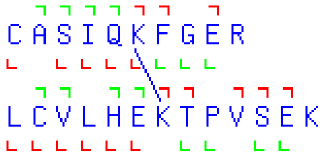
Using reagents with **cleavable linkers** in MS2- or MS3-based workflows



Computational considerations

The data analysis procedure tries to derive the following information from the experimental spectra

- Identity of the two connected peptide sequences
- Localization of the cross-linking sites within the peptides
- Quality of the match (mainly for identification, not site assignment) by assigning a score



All commonly used software for XL-MS relies on database search, i.e. the experimental spectra are compared against predicted spectra from sets of paired peptides derived from known protein sequences

The difficulty of the identification step scales with database size (small for CASP targets), the difficulty for site assignment depends on the chemical specificity of the cross-linking reagent

To estimate error rates, established statistical procedures (target/decoy competition) can be used, but for small data sets error rate determination is not very robust (we specify approx. 5% for our data)

Remember that at this stage, discrimination of native and non-native cross-links is not possible!

Additional information from XL-MS experiments

In addition to cross-linked peptides, single peptide chains that are modified by the cross-linking reagents may be identified

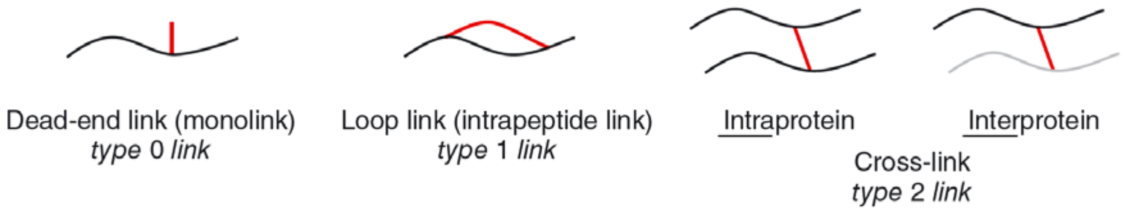
- Reflects solvent exposure of modified residues
- Computational identification relatively robust, but interpretation more ambiguous (from free/bound protein?)

The choice of protease also determines which regions of the protein sequences may not be accessible for MS analysis

- Particularly, if peptides are too long (20+ residues), identification rates decrease rapidly
- For membrane proteins and other hydrophobic proteins, these regions may span a considerable part of the sequence
- Additional/complementary proteases could be used, but this requires more material and time

We will provide both types of information for CASP targets!

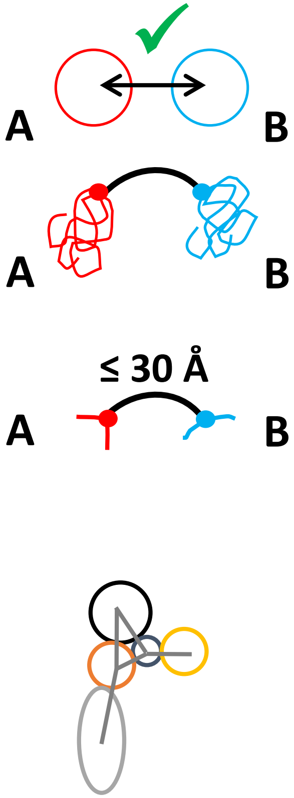
Different products from a XL experiment

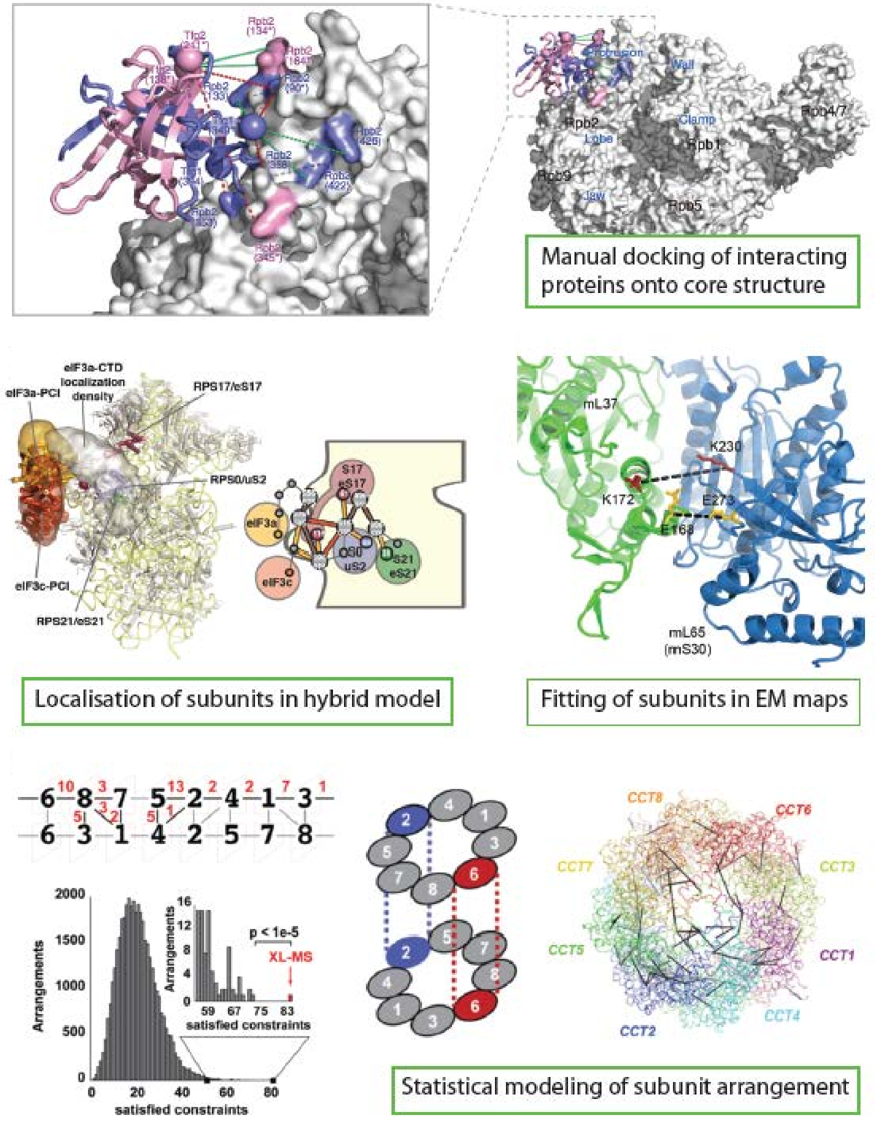


Information content of XL-MS data

What can cross-linking data tell you?

- **Direct evidence** of physical interactions
- **Spatial contact information** on subunit or domain level
- **Distance restraint information** (via spacer length)
  as input for modeling
- Information about **subunit arrangements, binding sites, conformational changes**

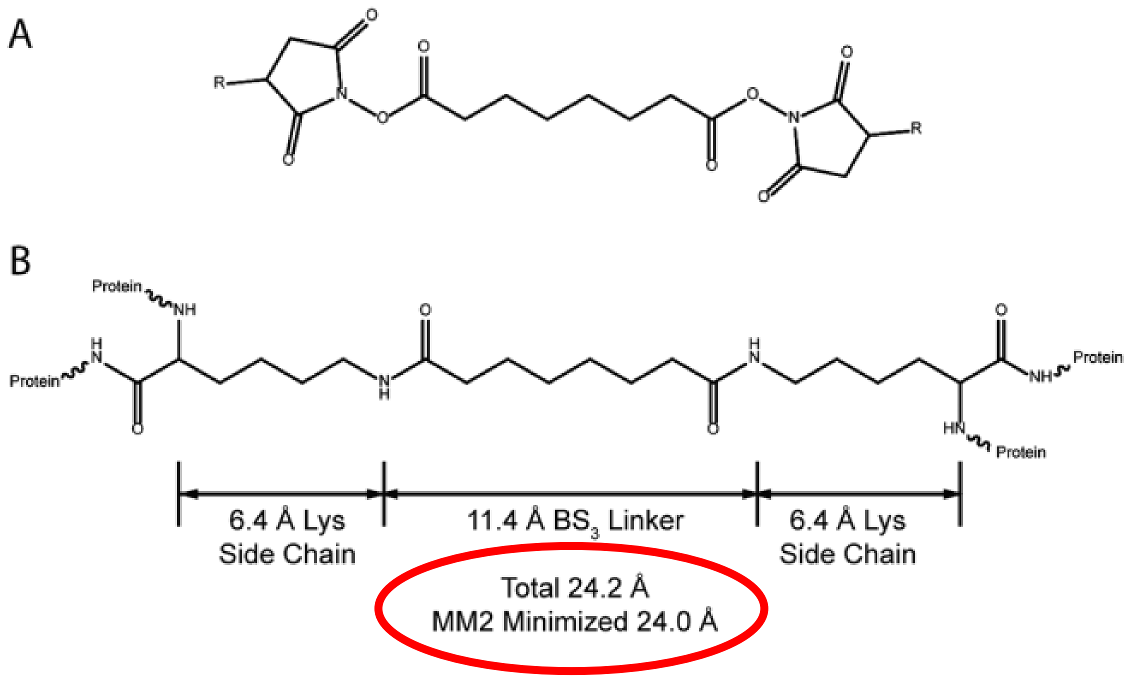




Leitner et al., Trends Biochem. Sci., 2016

How to calculate actual distance restraints?

The theoretical distance that a cross-linker can bridge can be easily calculated, e.g. for DSS / BS^3:

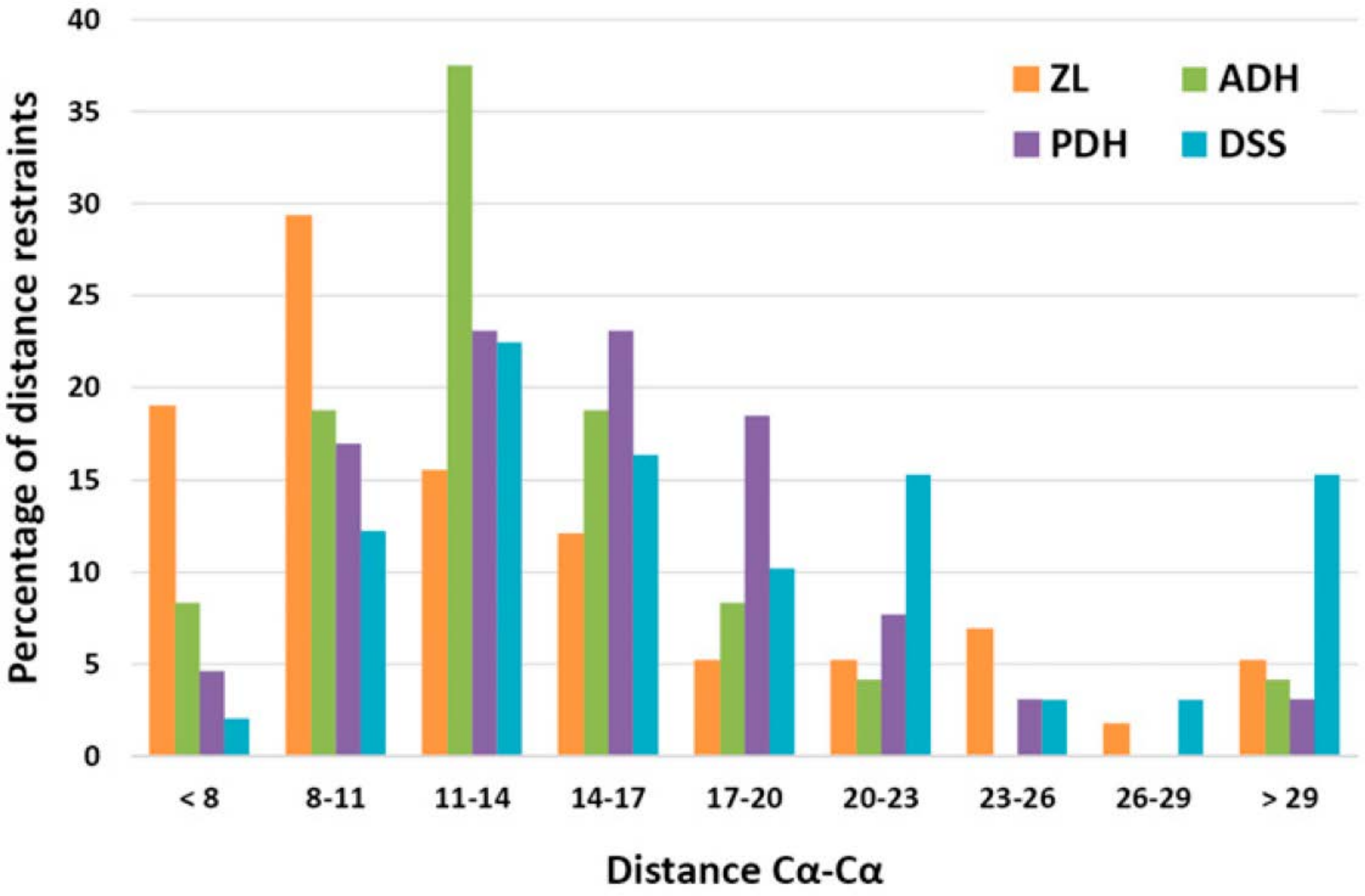


Merkley et al., Protein Sci., 2014

How to calculate actual distance restraints?

Practically, larger distances are observed, e.g. up to 30 Å and more (for proteins with known 3D structure)

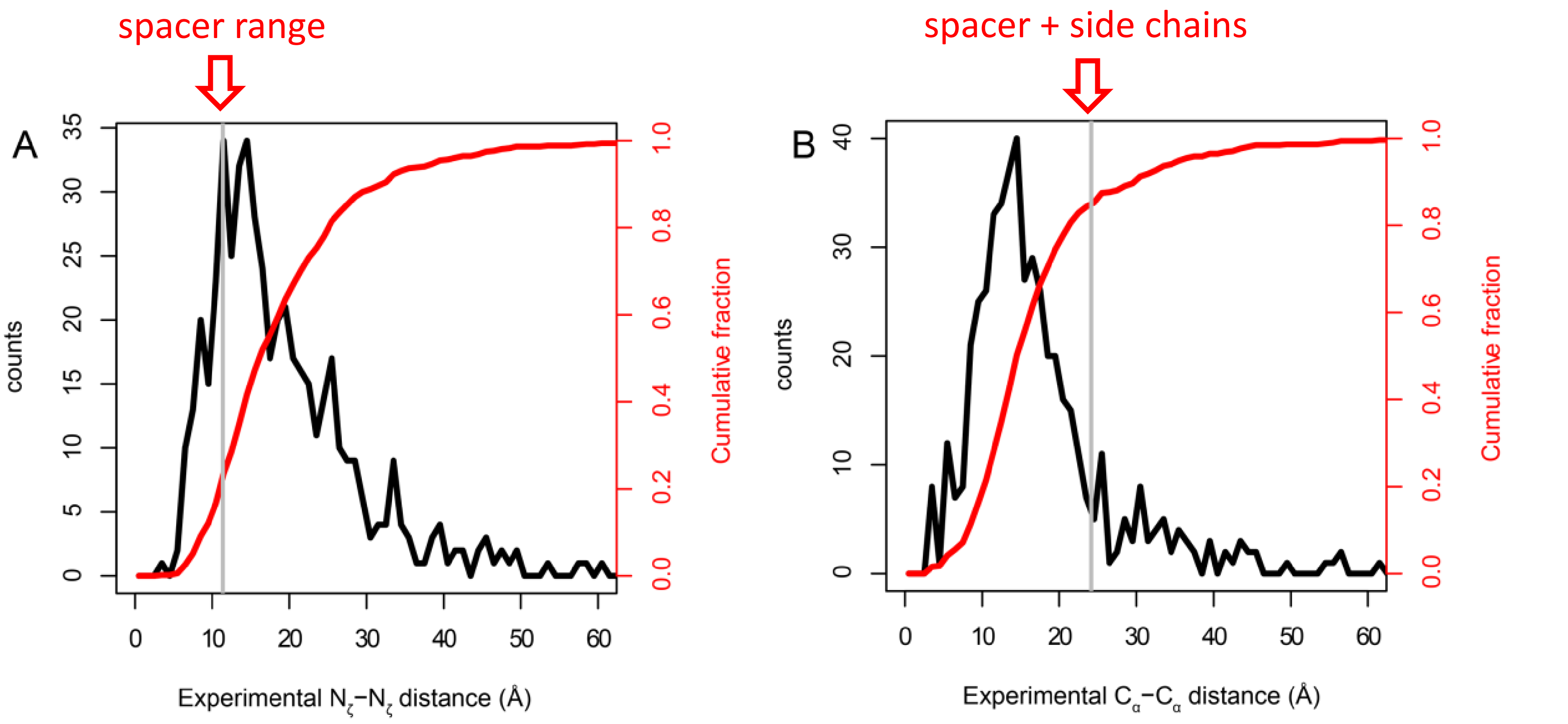
Note that ZL cross-links bridge shorter distances, but by only approx. 5 Å!



Leitner et al., PNAS, 2014

How to calculate actual distance restraints?

Practically, larger distances are observed, e.g. up to 30 Å and more (for proteins with known 3D structure)

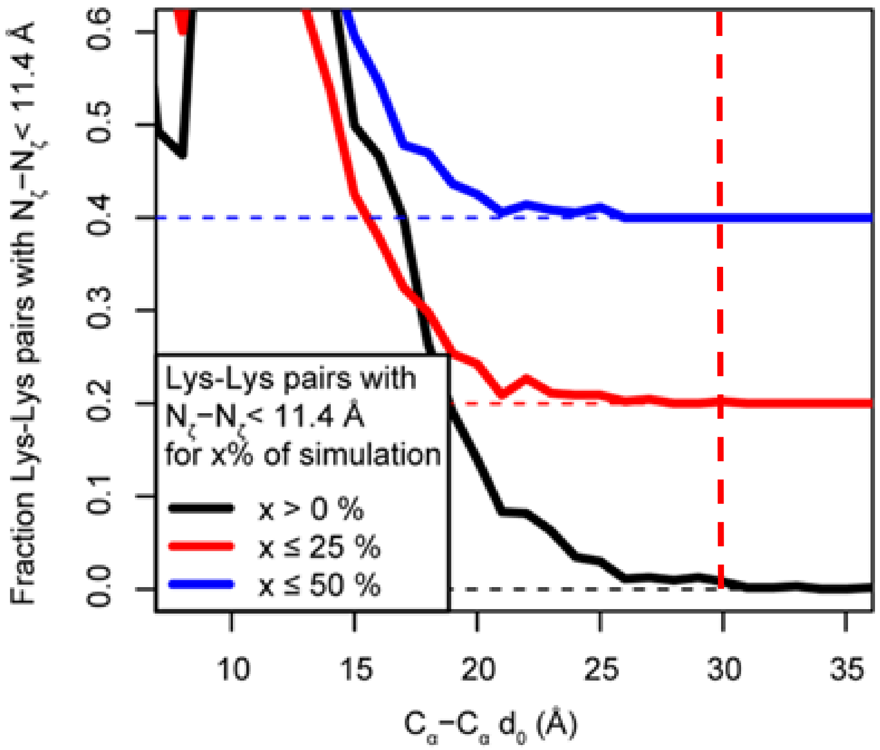


DSS/BS^3 data: Approx. 90% < 30 Å for both scenarios

Merkley et al., Protein Sci., 2014

How to calculate actual distance restraints?

This can be explained by the flexibility of proteins, as confirmed by molecular dynamics simulations
→ a 30 Å cut-off seems reasonable for DSS / BS^3



Merkley et al., Protein Sci., 2014

Higher flexibility possible for

- Terminal regions, flexible loops within proteins
- Very large assemblies («molecular machines»)

In addition to linear (Euclidean) distances, distances can also be calculated over the protein surfaces

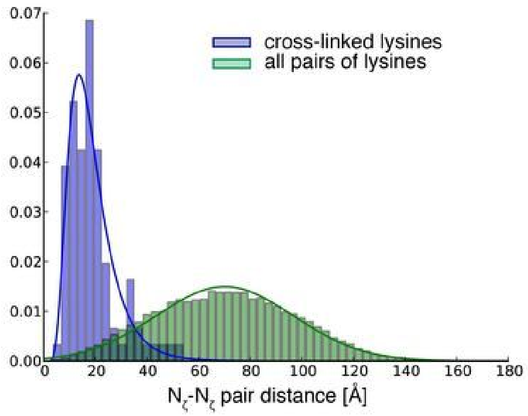
Use of distance restraints in modeling

Cross-linking derived restraints can therefore be treated as

- **Hard cut-offs** with fixed upper distance thresholds or distance ranges, e.g. < 30 Å = compatible, 5-30 Å = compatible
- **Soft cut-offs** with a penalty function, e.g.

The effect of a DSS crosslinker, which is specific for primary amines, including amino groups of lysine residues and protein amino termini, is mimicked by the combination of two types of interactions: the log-harmonic restraints of the elastic network that maintain the Nζ atoms approximately in their starting position with respect to the nearby Cα atoms and the Nζ–Nζ log-harmonic potential.

(Ferber et al., Nat. Methods, 2016)



If the SASD is under 33 Å, it is scored positively, taking into account its probability distribution, which is given by a normal distribution (the mean and variance are calculated from all the SASDs ≤33 Å from the XLdb). If the SASD exceeds 33 Å (indicating inconsistency with the native structure) it is scored with a flat penalty of −0.1.

(Bullock et al., Mol. Cell. Proteomics, 2016)

In addition, restraints can be

- **directly considered during the modeling stage** or
- **used for validation / filtering purposes** to rank models obtained independently