CASP 13 target: X0957

Cross-linking mass spectrometry data for H0957

**Protein information (as provided)**

CASP13 target: X0957

Protein Name: not available

Organism Name: Escherichia coli 3006

Amino acid sequence: CPX200209

**A**  
SNSFEVSSLPDANGKNHITAVKGDAKIPVDKIELYMRGKASGDLDSLQAEYNSLKDARISSQKEFAKDPNNAKRMEVLEKQIHNIERSQDMARVLEQAGIVNTASNNSMIMDKLLDSAQGATSANRKTSVVVSGPNGNVRIYATWTILPDGTKRLSTVTGTFK

**B**  
SNAMINVNSTAKDIEGLESYLANGYVEANSFNDPEDDALECLSNLLVKDSRGGLSFCKKILNSNNIDGVFIKGSALNFLLLSEQWSYAFEYLTSNADNITLAELEKALFYFYCAKNETDPYPVPEGLFKKLMKRYEELKNDPDAKFYHLHETYDDFSKAYPLN

**Methods**

The target protein complex was cross-linked and analyzed by mass spectrometry as described here:

*Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline*. Leitner, Walzthoeni and Aebersold. *Nature Protocols*, 2014. DOI: 10.1038/nprot.2013.168

*Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes*. Leitner, Joachimiak, Unverdorben, Walzthoeni, Frydman, Förster and Aebersold. *Proceedings of the National Academy of Sciences of the United States of America,* 2014. DOI: 10.1073/pnas.1320298111

The concentration the protein complex was adjusted to avoid over-cross-linking, e.g. introduction of non-native oligomerization states.

All cross-linking reactions were followed by SDS-PAGE.

**Cross-links identified by mass spectrometry \***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Interprotein XL** | |  |  |  |  |
| **Protein1** | **Protein2** | **AbsPos1** | **AbsPos2** | **ld-Score** | **Chemistry\*\*** |
| A | B | 127 | 67 | 43.83 | ZL |
| A | B | 50 | 1 | 41.37 | ZL |
| A | B | 163 | 141 | 37.18 | ZL |
| B | A | 119 | 63 | 33.88 | ZL |
| B | A | 145 | 90 | 29.12 | ZL |
| B | A | 67 | 127 | 27.21 | ZL |
| B | A | 1 | 67 | 25.63 | DSS |
| A | B | 163 | 143 | 23.38 | ZL |
|  |  |  |  |  |  |
| **Intra protein XL** | |  |  |  |  |
| **Protein1** | **Protein2** | **AbsPos1** | **AbsPos2** | **ld-Score** | **Chemistry** |
| A | A | 127 | 80 | 50.94 | DSS |
| A | A | 39 | 31 | 40.31 | DSS |
| A | A | 39 | 80 | 38.93 | DSS |
| A | A | 116 | 22 | 36.59 | ZL |
| A | A | 80 | 73 | 35.52 | DSS |
| A | A | 39 | 39 | 33.29 | DSS \*\*\* |
| A | A | 39 | 33 | 33.2 | ZL |
| A | A | 112 | 22 | 27.61 | ZL |
|  |  |  |  |  |  |
| B | B | 59 | 1 | 32.59 | DSS |
| B | B | 158 | 139 | 30.16 | DSS |
| B | B | 129 | 139 | 24.61 | DSS |

\* The score is a measure of confidence for the identification of the two connected peptides (i.e. computational assignment) that are identified by MS (the higher, the better). It is generally NOT correlated with the distance between the cross-linked residues. In addition, physicochemical properties of the peptides may affect the identification, so that some cross-linked peptides intrinsically have lower scores.

For the reported data, we expect a **false positive rate of identification of approximately 5%.**

\*\*Cross-linking chemistries:

DSS: disuccinimidyl suberate – a lysine specific cross-linker.  
ZL: Zero-length cross-links formed between lysine and an aspartate/glutamate residue by the coupling reagent 4-(4,6-dimethoxy-1,3,5- triazin-2-yl)-4-methylmorpholinium chloride (DMTMM).  
PDH: pimelic acid dihydrazide – a carboxylic acid specific cross-linker (Aspartate and glutamate).

For experimentally observed distance restraints, see the following plot (ADH is not used here):



(taken from Leitner et al., PNAS, 2014)

\*\*\* This cross-link connects two identical residues, pointing to a homo-dimeric contact (native or non-native).

**Sub-optimal sequence regions for conventional cross-linking mass spectrometry**

Red residues: Lysine residues. Can be cross-linked by DSS and the zero-length cross-linking reagent DMTMM. Cleavage sites for trypsin (protease used in the process).

Black residues: Arginine residues. Cleavage sites for trypsin.

Green residues. Aspartate and Glutamate residues. Can be cross-linked by PDH and the zero-length cross-linking reagent DMTMM.

Residues highlighted in yellow are sub-optimal regions for mass spectrometric analysis.

**A**

10 20 30 40 50 60

SNSF**E**VSSLP **D**ANG**K**NHITA V**K**G**D**A**K**IPV**D K**I**E**LYM**R**G**K**A SG**D**L**D**SLQA**E** YNSL**KD**A**R**IS

70 80 90 100 110 120

SQ**KE**FA**KD**PN NA**KR**M**E**VL**EK** QIHNI**ER**SQ**D** MA**R**VL**E**QAGI VNTASNNSMI M**DK**LL**D**SAQG

130 140 150 160

ATSAN**RK**TSV VVSGPNGNV**R** IYATWTILP**D** GT**KR**LSTVTG TF**K**

B

10 20 30 40 50 60

SNAMINVNST A**KD**I**E**GL**E**SY LANGYV**E**ANS FN**D**P**EDD**AL**E** CLSNLLV**KD**S **R**GGLSFC**KK**I

70 80 90 100 110 120

LNSNNI**D**GVF I**K**GSALNFLL LS**E**QWSYAF**E** YLTSNA**D**NIT LA**E**L**EK**ALFY FYCA**K**N**E**T**D**P

130 140 150 160

YPVP**E**GLF**KK** LM**KR**Y**EE**L**K**N **D**P**D**A**K**FYHLH **E**TY**DD**FS**K**AY PLNN

**Residues labelled by cross-linking reagents**

Red residues: residues labeled by either by DSS (Lysine reactive) or pimelic acid dihydrazide (PDH, reactive towards carboxylic acids), but not cross-linked. These residues are expected to be solvent exposed.

Notes:

Complex formation in the cross-linked sample may not be quantitative, so that exposed regions could only be accessible in the free binding partners.

Absence of a modification may also mean that the corresponding modified peptide is present, but not identified by MS.

Green residues: reactive unlabeled residues.

**A**

10 20 30 40 50 60

**S**NSF**E**VSSLP **D**ANG**K**NHITA V**K**G**D**A**K**IPV**D K**I**E**LYMRG**K**A SG**D**L**D**SLQA**E** YNSL**KD**ARIS

70 80 90 100 110 120

SQ**KE**FA**KD**PN NA**K**RM**E**VL**E**K QIHNI**E**RSQ**D** MARVL**E**QAGI VNTASNNSMI M**D**KLL**D**SAQG

130 140 150 160

ATSANR**K**TSV VVSGPNGNVR IYATWTILP**D** GT**K**RLSTVTG TF**K**

**B**

10 20 30 40 50 60

**S**NAMINVNST A**KD**IEGL**E**SY LANGYV**E**ANS FN**D**PE**DD**ALE CLSNLLV**KD**S RGGLSFC**KK**I

70 80 90 100 110 120

LNSNNI**D**GVF I**K**GSALNFLL LSEQWSYAFE YLTSNA**D**NIT LA**E**L**EK**ALFY FYCA**K**N**E**T**D**P

130 140 150 160

YPVP**E**GLF**KK** LM**K**RYEEL**K**N **D**P**D**A**K**FYHLH **E**TY**DD**FS**K**AY PLNN