CASP 13 internal ID: T0981/X0981

Cross-linking mass spectrometry data

**Protein information (as provided)**

CASP13 internal ID: regular target T0981, XL-MS assisted X0981.

Protein Name: gp146

Organism Name: Staphylococcus phage

Amino acid sequence:

**gp146**
 10 20 30 40 50 60
MGSSHHHHHH SSGLVPRGSH MASMTGGQQM GRGSMAFNYT PLTETQKLKD MYPKVNDIGN

 70 80 90 100 110 120
FLKTEVNLSD VKQISQPDFN NILASIPDSG NYYVTNSKGA PSGEATAGFV RLDKRNVNYY

 130 140 150 160 170 180
KIYYSPYSSN KMYIKTYANG TVYDWISFKL DEGSLYNEGN TLNVKELTES TTQYATLVNP

 190 200 210 220 230 240
PKENLNTGWV NYKESKNGVS SLVEFNPVNS TSTFKMIRKL PVQEQKPNLL KDSLFVYPET

 250 260 270 280 290 300
SYSNIKTDNW DTPPFWGYSS NSGRSGVRFR GENTVQIDDG SDTYPSVVSN RFKMGKELSV

 310 320 330 340 350 360
GDTVTVSVYA KINDPALLKD NLVYFELAGY DTVDDTSKNP YTGGRREITA SEITTEWKKY

 370 380 390 400 410 420
SFTFTIPENT IGASGVKVNY VSLLLRMNCS SSKGNGAVVY YALPKLEKSS KVTPFITHEN

 430 440 450 460 470 480
DVRKYDEIWS NWQEVISKDE LKGHSPVDIE YNDYFKYQWW KSEVNEKSLK DLAMTVPQGY

 490 500 510 520 530 540
HTFYCQGSIA GTPKGRSIRG TIQVDYDKGD PYRANKFVKL LFTDTEGIPY TLYYGGYNQG

 550 560 570 580 590 600
WKPLKQSETS TLLWKGTLDF GSTEAVNLND SLDNYDLIEV TYWTRSAGHF STKRLDIKNT

 610 620 630 640 650 660
SNLLYIRDFN ISNDSTGSSV DFFEGYCTFP TRTSVQPGMV KSITLDGSTN TTKVASWNEK

 670
ERIKVYNIMG INRG

**Methods**

The target protein 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

The gp146 protein could not successfully be analyzed by the DMTMM/PDH cross-linking chemistry due to buffer incompatibilities.

The concentration of the protein or 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. The cross-linked protein migrated as a trimer exclusively.

**Cross-links identified by mass spectrometry**

|  |
| --- |
| **Inter-chain cross-links\*** |
| **Protein1** | **Protein2** | **AbsPos1** | **AbsPos2** | **ld-Score\*** | **Chemistry\*\*** |
| gp146 | gp146 | 121 | 131 | 35.64 | DSS |
| gp146 | gp146 | 593 | 593 | 35.36 | DSS |
| gp146 | gp146 | 516 | 516 | 35.07 | DSS |
| gp146 | gp146 | 467 | 467 | 31.28 | DSS |
| gp146 | gp146 | 660 | 660 | 28.74 | DSS |
| gp146 | gp146 | 193 | 193 | 25.33 | DSS |

|  |
| --- |
| **Intra-chain cross-links** |
| **Protein1** | **Protein2** | **AbsPos1** | **AbsPos2** | **ld-Score\*** | **Chemistry\*\*** |
| gp146 | gp146 | 226 | 408 | 45.95 | DSS |
| gp146 | gp146 | 411 | 358 | 45.02 | DSS |
| gp146 | gp146 | 664 | 660 | 43.56 | DSS |
| gp146 | gp146 | 598 | 593 | 41.82 | DSS |
| gp146 | gp146 | 411 | 226 | 39.68 | DSS |
| gp146 | gp146 | 508 | 461 | 39.53 | DSS |
| gp146 | gp146 | 359 | 377 | 38.62 | DSS |
| gp146 | gp146 | 358 | 219 | 38.23 | DSS |
| gp146 | gp146 | 196 | 131 | 37.69 | DSS |
| gp146 | gp146 | 467 | 516 | 36.48 | DSS |
| gp146 | gp146 | 508 | 593 | 36.28 | DSS |
| gp146 | gp146 | 442 | 424 | 36.22 | DSS |
| gp146 | gp146 | 319 | 338 | 34.53 | DSS |
| gp146 | gp146 | 641 | 660 | 34.34 | DSS |
| gp146 | gp146 | 593 | 516 | 33.76 | DSS |
| gp146 | gp146 | 641 | 653 | 33.28 | DSS |
| gp146 | gp146 | 598 | 664 | 32.63 | DSS |
| gp146 | gp146 | 121 | 131 | 32.2 | DSS |
| gp146 | gp146 | 660 | 293 | 31.55 | DSS |
| gp146 | gp146 | 358 | 226 | 29.15 | DSS |
| gp146 | gp146 | 196 | 121 | 28.96 | DSS |
| gp146 | gp146 | 411 | 219 | 28.01 | DSS |
| gp146 | gp146 | 182 | 424 | 26.68 | DSS |
| gp146 | gp146 | 182 | 442 | 26.47 | DSS |
| gp146 | gp146 | 424 | 438 | 26.21 | DSS |

\* These cross-links connect two identical residues or peptides with overlapping sequence, pointing to a inter-subunit contact (native or non-native). Cross-links listed as intra-chain could also be inter-chain for this homotrimer.

\*\* The score is a measure of confidence for the identification of the tw-o 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, PDH and ZL are not used here):



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

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

Red residues: Lysine residues. Can be cross-linked by DSS and they are cleavage sites for trypsin (protease used in the process).

Black residues: Arginine residues. Cleavage sites for trypsin.

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

**gp146**
 10 20 30 40 50 60

MGSSHHHHHH SSGLVP**R**GSH MASMTGGQQM G**R**GSMAFNYT PLTETQ**K**L**K**D MYP**K**VNDIGN

 70 80 90 100 110 120

FL**K**TEVNLSD V**K**QISQPDFN NILASIPDSG NYYVTNS**K**GA PSGEATAGFV **R**LD**KR**NVNYY

 130 140 150 160 170 180

**K**IYYSPYSSN **K**MYI**K**TYANG TVYDWISF**K**L DEGSLYNEGN TLNV**K**ELTES TTQYATLVNP

 190 200 210 220 230 240

P**K**ENLNTGWV NY**K**ES**K**NGVS SLVEFNPVNS TSTF**K**MI**RK**L PVQEQ**K**PNLL **K**DSLFVYPET

 250 260 270 280 290 300

SYSNI**K**TDNW DTPPFWGYSS NSG**R**SGV**R**F**R** GENTVQIDDG SDTYPSVVSN **R**F**K**MG**K**ELSV

 310 320 330 340 350 360

GDTVTVSVYA **K**INDPALL**K**D NLVYFELAGY DTVDDTS**K**NP YTGG**RR**EITA SEITTEW**KK**Y

 370 380 390 400 410 420

SFTFTIPENT IGASGV**K**VNY VSLLL**R**MNCS SS**K**GNGAVVY YALP**K**LE**K**SS **K**VTPFITHEN

 430 440 450 460 470 480

DV**RK**YDEIWS NWQEVIS**K**DE L**K**GHSPVDIE YNDYF**K**YQWW **K**SEVNE**K**SL**K** DLAMTVPQGY

 490 500 510 520 530 540

HTFYCQGSIA GTP**K**G**R**SI**R**G TIQVDYD**K**GD PY**R**AN**K**FV**K**L LFTDTEGIPY TLYYGGYNQG

 550 560 570 580 590 600

W**K**PL**K**QSETS TLLW**K**GTLDF GSTEAVNLND SLDNYDLIEV TYWT**R**SAGHF ST**KR**LDI**K**NT

 610 620 630 640 650 660

SNLLYI**R**DFN ISNDSTGSSV DFFEGYCTFP T**R**TSVQPGMV **K**SITLDGSTN TT**K**VASWNE**K**

 670

E**R**I**K**VYNIMG IN**R**G

**Residues labelled by cross-linking reagents**

Red residues: residues labeled by DSS (Lysine and N-terminus reactive).These residues are expected to be solvent exposed.

Green residues: reactive unlabeled residues.

Notes:

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

**gp146**
 10 20 30 40 50 60

MGSSHHHHHH SSGLVPRGSH MASMTGGQQM GRGSMAFNYT PLTETQ**K**L**K**D MYP**K**VNDIGN

 70 80 90 100 110 120

FL**K**TEVNLSD V**K**QISQPDFN NILASIPDSG NYYVTNS**K**GA PSGEATAGFV RLD**K**RNVNYY

 130 140 150 160 170 180

**K**IYYSPYSSN **K**MYI**K**TYANG TVYDWISF**K**L DEGSLYNEGN TLNV**K**ELTES TTQYATLVNP

 190 200 210 220 230 240

P**K**ENLNTGWV NY**K**ES**K**NGVS SLVEFNPVNS TSTF**K**MIR**K**L PVQEQ**K**PNLL **K**DSLFVYPET

 250 260 270 280 290 300

SYSNI**K**TDNW DTPPFWGYSS NSGRSGVRFR GENTVQIDDG SDTYPSVVSN RF**K**MG**K**ELSV

 310 320 330 340 350 360

GDTVTVSVYA **K**INDPALL**K**D NLVYFELAGY DTVDDTS**K**NP YTGGRREITA SEITTEW**KK**Y

 370 380 390 400 410 420

SFTFTIPENT IGASGV**K**VNY VSLLLRMNCS SS**K**GNGAVVY YALP**K**LE**K**SS **K**VTPFITHEN

 430 440 450 460 470 480

DVR**K**YDEIWS NWQEVIS**K**DE L**K**GHSPVDIE YNDYF**K**YQWW **K**SEVNE**K**SL**K** DLAMTVPQGY

 490 500 510 520 530 540

HTFYCQGSIA GTP**K**GRSIRG TIQVDYD**K**GD PYRAN**K**FV**K**L LFTDTEGIPY TLYYGGYNQG

 550 560 570 580 590 600

W**K**PL**K**QSETS TLLW**K**GTLDF GSTEAVNLND SLDNYDLIEV TYWTRSAGHF ST**K**RLDI**K**NT

 610 620 630 640 650 660

SNLLYIRDFN ISNDSTGSSV DFFEGYCTFP TRTSVQPGMV **K**SITLDGSTN TT**K**VASWNE**K**

 670

ERI**K**VYNIMG INRG