Supplementary Information

Direct N- or C-terminal protein labeling via a sortase-mediated swapping approach

Min Cong[1*], Soheil Tavakolpour[1*], Lea Berland[1*], Hannah Glockner[1], Bohdan Andreiuk[1], Taha Rakhshandehroo[1], Safak Uslu[1], Shruti Mishra[1], Louise Clark[1], and Mohammad Rashidian[1**]

[1] Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02215, USA.

[*] These authors contributed equally to this work

[**] Correspondence: mohammad_rashidian@dfci.harvard.edu

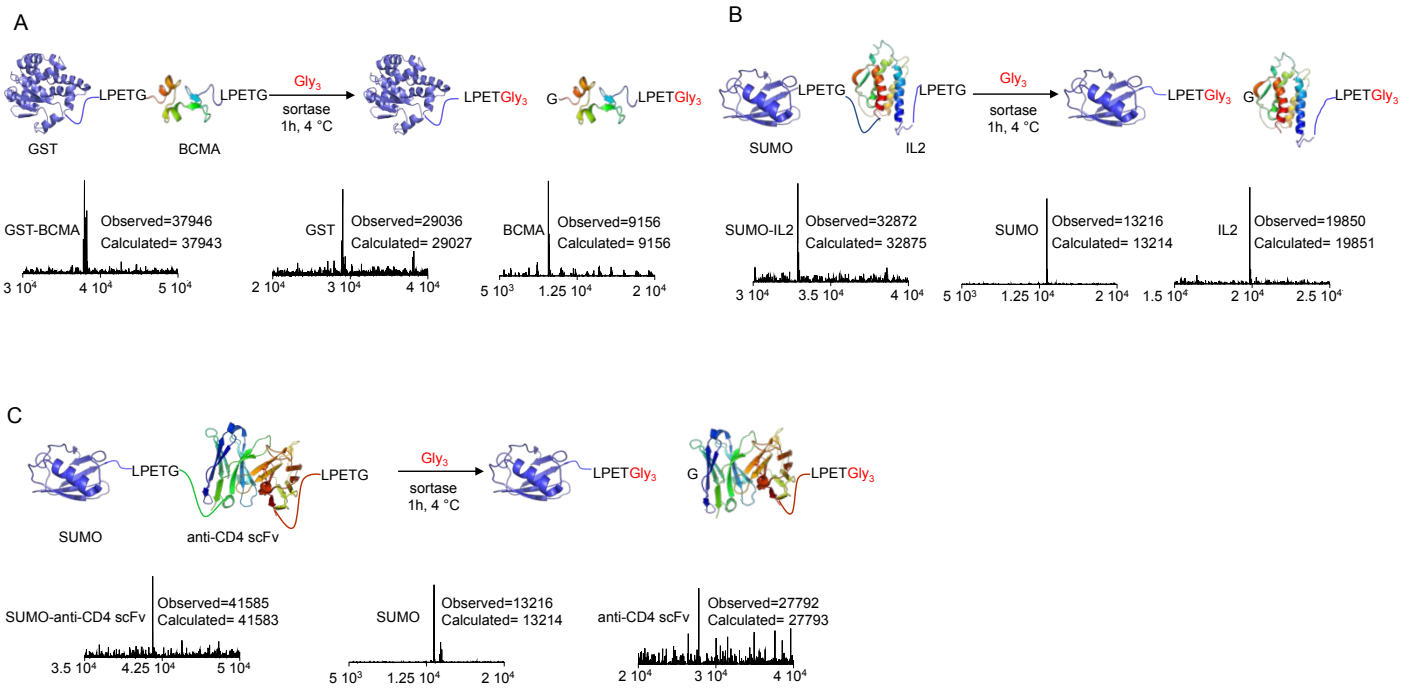


Figure S1. Fusion tag cleavage using sortase. A, B, C) GST-BCMA, SUMO-IL2, and SUMO-anti-CD4 scFv fusion proteins were rapidly and quantitatively cleaved via the addition of sortase and Gly3 substrate. LC-MS analyses confirmed the formation of the products.

1. Sequences of recombinant protein substrates produced in this study. Sortase recognition tags are underlined.

1.1. GFP:

MHHHHHH**LPETG**GGGGSGGSGGSGGSGGVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKL
TLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFE
GDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGD
GPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK

1.2. GST-BCMA:

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVT

HPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLE
VLFQGPGSHHHHHHHHHHGGGGS**LPETG**SDYKDDDDKGGSMLQMAGQCSQNEYFDSLLHACIPCQLRCSS
NTPPLTCQRYCNASVTNSVKGTNAGGGGSDYKDDDDK**LPETG**G

1.3. SUMO-BCMA:

MKYLLPTAAAGLLLLAAQPAMAGSMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME
AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGGGSDYKDDDDK**LPETG**GGGSGGGS
GGGDYKDDDDKMLQMAGQCSQNEYFDSLLHACIPCQLRCSSNTPPLTCQRYCNASVTNSVKGTNAGSDYKD
DDDKHHHHHH

*In red is the pelB sequence that is cleaved upon periplasmic expression.

1.4. SUMO-IL2

MKYLLPTAAAGLLLLAAQPAMAGSMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME
AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGGGSDYKDDDDK**LPETG**GGGSGGGS
GGGSGGGPTSSPTSSPTSSSTAEAQQQQQQQQHLEQLLMDLQELLSRMENYRNLKLPRMLTFKFYLPKQAT
ELKDLQCLEDELGPLRHVLDLTQSKSFQLEDAENFISNIRVTVVKLKGSDNTFECQFDDESATVVDFLRRWIAF
CQSIISTSPQDYKDDDDKHHHHHH

*In red is the pelB sequence that is cleaved upon periplasmic expression.

1.5. SUMO-anti-CD4 scFv

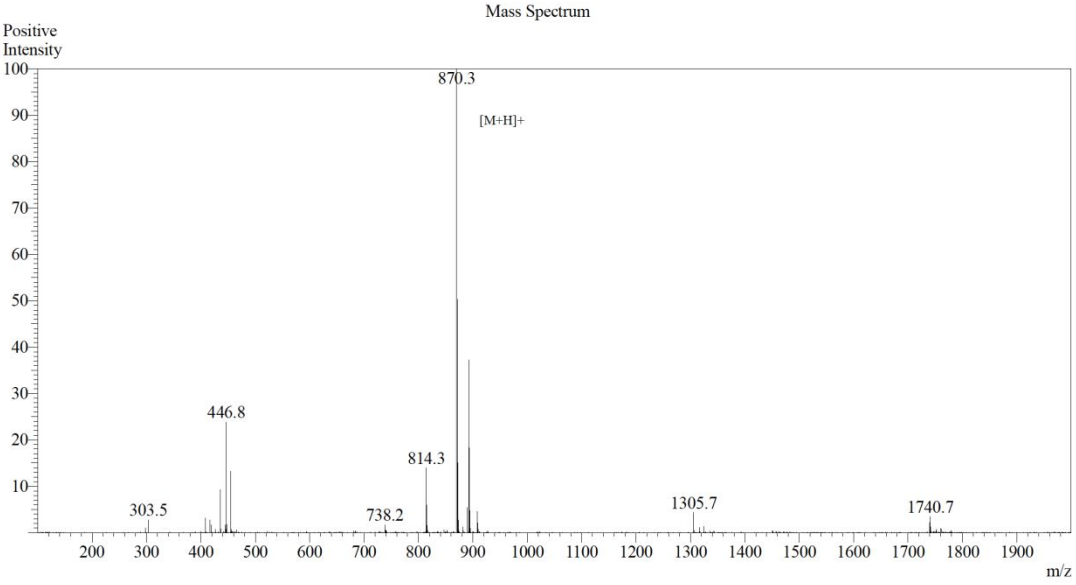
GSMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQ
ADQTPEDLDMEDNDIIEAHREQIGGGGSDYKDDDDK**LPETG**SSGSEVQLVQSGAELVKPGASVKLSCKVSDY
NIRRTYMHWVRQRPGKGLEWIGRIDPANGNTIYGEKFKSKATLTADTSSNTAYMQLSQLKSDDTAIYYCAIGVQ
YLDYWGQGTTVTVSSGGGGSGGGGSGGGGSDIVLTQSPALAVSPGERVTISCRATESVSTLIHWYQQRPGQ
QPKLLIYLTSHLDSGVPARFSGSGSGTDFTLTIDPVEADDTATYYCQQTWNDPWTFGGGTKLELKRGGGSDY
KDDDDK**LPETG**GHHHHHH

2. Structures, Mass Spectra, and HPLC analyses of the sortase substrates used in this study (peptides were custom synthesized at GenScript company):

2.1. azide-LPETG

Peptide sequence: {Lys(N3)}GSLPETGG; C-terminal: amide; Theoretical MW: 869.93

MS analysis:



Sample Information
Month-Day Processed : 12/13/20
Time Processed : 10:45:18 PM
Injection Volume : 0.4
Sample Name : azide-GSLPETGG--100mg
Sample ID : U055PFK300-3
Theoretical MW : 869.93
Observed MW : 869.3

Interface :ESI
Nebulizing Gas Flow :1.5L/min
CDL Temp :250
Block Temp :200

Equipment : GK11010007
Interface Bias : +4.5 kV
Drying Gas Flow :5 L/min
T.Flow :0.2 ml/min
B.conc :50%H2O/50%MeOH

HPLC analysis:

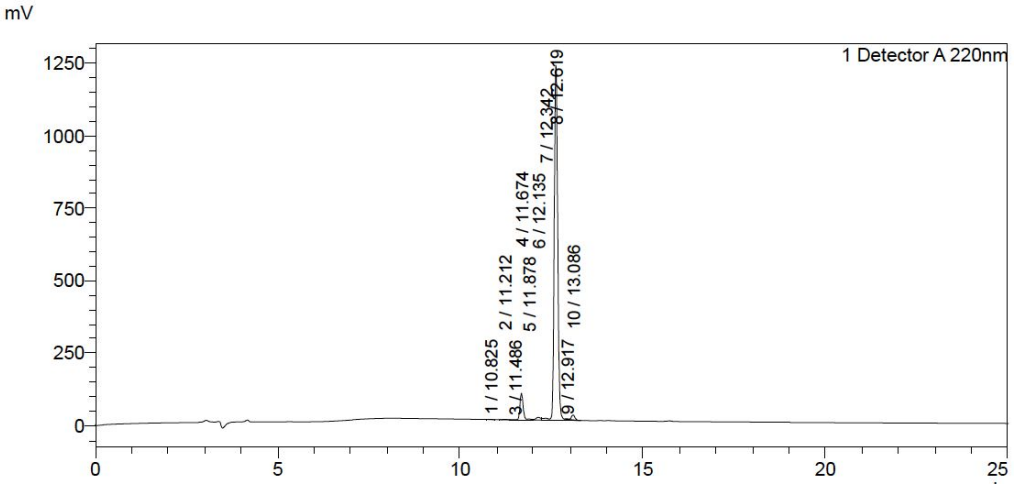
Purity: 90.1%

Pump A : 0.065% trifluoroacetic in 100% water (v/v)
Pump B : 0.05% trifluoroacetic in 100% acetonitrile (v/v)
Total Flow:1 ml/min
Wavelength:220 nm
<<LC Time Program>>

Time	Module	Command	Value
0.01	Pumps	B.Conc	5
25.00	Pumps	B.Conc	65
25.01	Pumps	B.Conc	95
27.00	Pumps	B.Conc	95
27.01	Pumps	B.Conc	5
35.00	Pumps	B.Conc	5
35.01	Controller	Stop	

<<Column Performance>>
<Detector A>
Column :Inertsil ODS-3 4.6 x 250 mm
Equipment: ZJ17010508

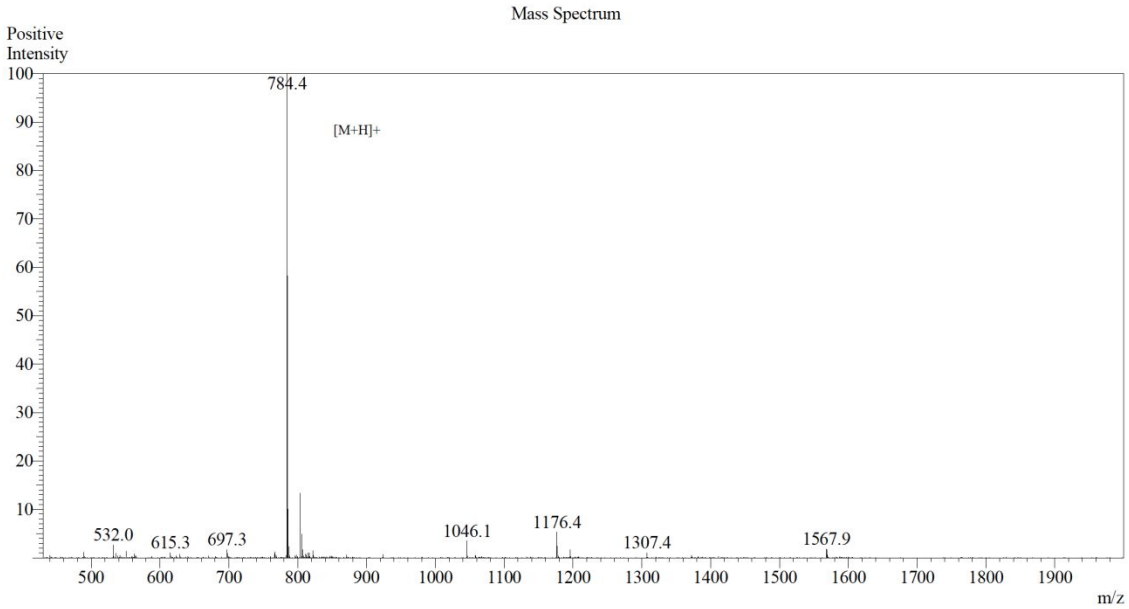
<Chromatogram>



2.2. alkyne-LPETG

Peptide sequence: {Pra}SLPETGS; C-terminal: amide; Theoretical MW: 783.83

MS analysis:



Sample Information
Month-Day Processed : 03/05/21
Time Processed : 7:28:42 PM
Injection Volume : 0.4
Sample Name : {Pra}SLPETGS
Sample ID : U764WGB180-5
Theoretical MW : 783.83
Observed MW : 783.4

Interface :ESI
Nebulizing Gas Flow :1.5L/min
CDL Temp :250
Block Temp :200

Equipment : GK11010007
Interface Bias : +4.5 kV
Drying Gas Flow :5 L/min
T.Flow :0.2 ml/min
B.conc :50%H2O/50%MeOH

HPLC analysis:

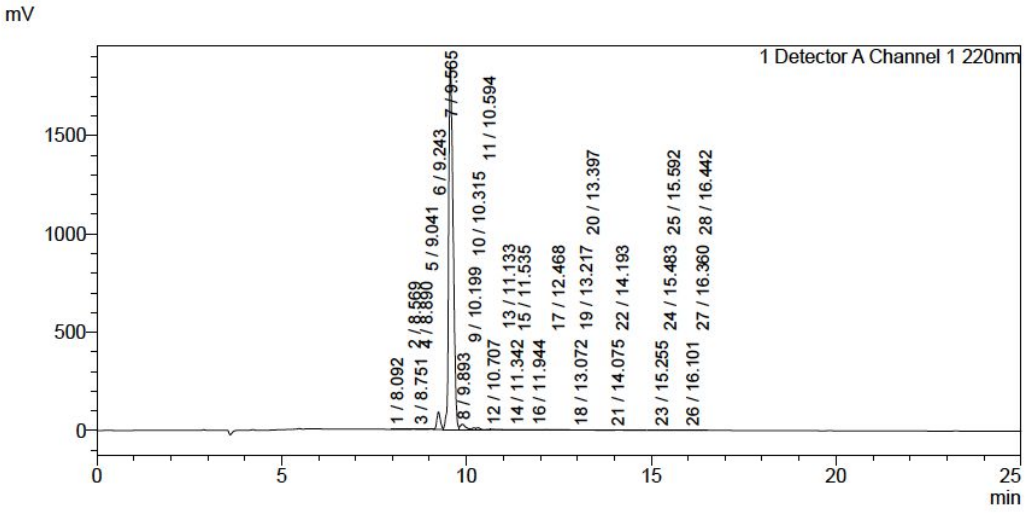
Purity: 93.3%

Pump A : 0.065% trifluoroacetic in 100% water (v/v)
Pump B : 0.05% trifluoroacetic in 100% acetonitrile (v/v)
Total Flow:1 ml/min
Wavelength:220 nm
<<LC Time Program>>

Time	Module	Command	Value
0.01	Pumps	B.Conc	5
25.00	Pumps	B.Conc	65
25.01	Pumps	B.Conc	95
27.00	Pumps	B.Conc	95
27.01	Pumps	B.Conc	5
35.00	Pumps	B.Conc	5
35.01	Controller	Stop	

<<Column Performance>>
<Detector A>
Column :Inertsil ODS-3 4.6 x 250 mm
Equipment: GK11010009

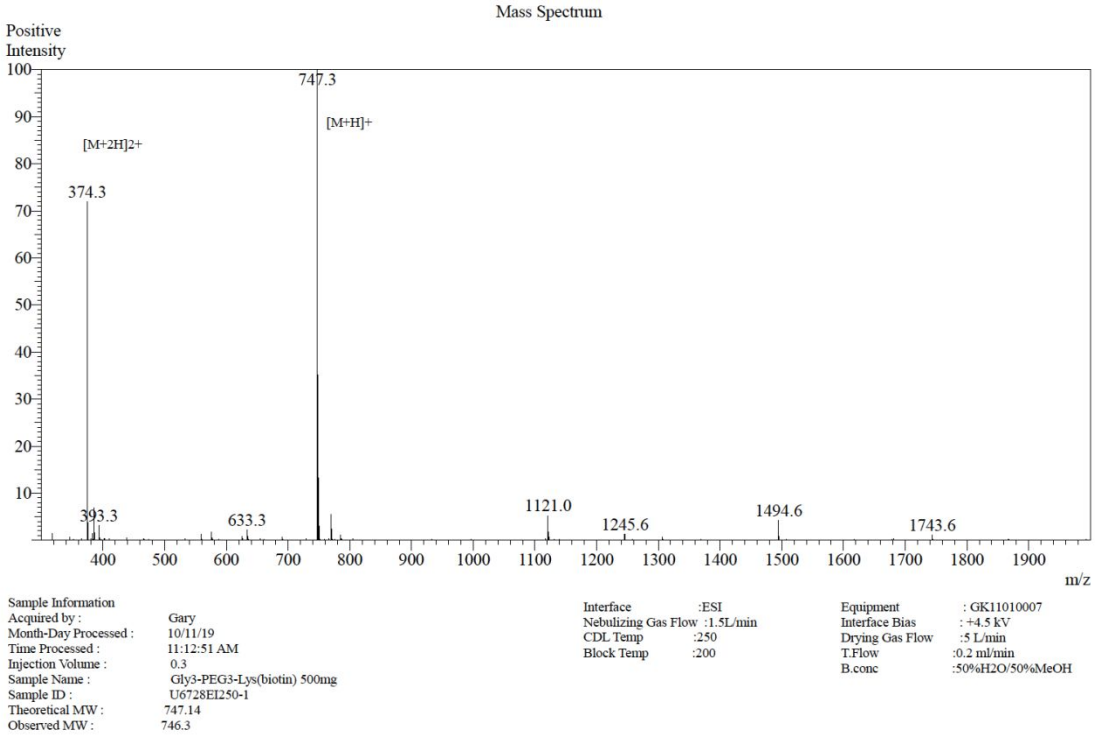
<Chromatogram>



2.3. Gly_3-Biotin

Peptide sequence: GGG{PEG3}{Lys(biotin)}; Theoretical MW: 747.14

MS analysis:



HPLC analysis:

Purity: 94.3%

Pump A : 0.065% trifluoroacetic in 100% water (v/v)
Pump B : 0.05% trifluoroacetic in 100% acetonitrile (v/v)
Total Flow:1 ml/min
Wavelength:220 nm
<<LC Time Program>>

Time	Module	Command	Value
0.01	Pumps	Solvent B Conc.	5
25.00	Pumps	Solvent B Conc.	65
25.01	Pumps	Solvent B Conc.	95
27.00	Pumps	Solvent B Conc.	95
27.01	Pumps	Solvent B Conc.	5
35.00	Pumps	Solvent B Conc.	5
35.01	Controller	Stop	

<<Column Performance>>
<Detector A>
Column : Inertsil ODS-3 4.6 x 250 mm
Equipment:ZJ17010508

<Chromatogram>