Supporting Information for

Solving Complex Biologics Truncation Problems by Top-Down Mass Spectrometry

Zhe Zhang*1, Christian Hug2, Yuanqi Tao1, Francis Bitsch2, and Yang Yang2

¹NIBR Biologics Center, Novartis Institutes for BioMedical Research, Cambridge, Massachusetts, 02139, United States

² NIBR Biologics Center, Novartis Institutes for BioMedical Research, Klybeckstrasse 141, CH-4057, Basel, Switzerland

Address: 100 Technology Square, Cambridge, Massachusetts, 02139, United States

Email: zhe.zhang@novartis.com

Table of Contents:

- Figure S1. MS spectrum of TP1 and chromatograph of TP1 and TP2
- Figure S2. Raw top-down MS/MS spectrum of TP1 and TP2
- Figure S3. TerSeqCov calculation and comparison using different numbers of terminal residues
- Figure S4. Sequence coverage analyzed by ProSight
- Figure S5. ETD and CID Sequence coverage analysis of TP1
- Figure S6. Western blot of TP2
- Full sequences of TP1 and TP2 chain A



Figure S1. Raw LC-MS spectrum of TP1 and TP2. (a) UV chromatograph of TP1 and TP2. Other byproducts of TP2 was studied by other methods and out of scope of this article. (b) Averaged TP1 MS spectrum of ~1 min RP-LCMS eluted peak of the produced material. (c) zoom-in view of the averaged spectrum showing one dominant species and others impurities in TP1 sample.



Figure S2. Raw top-down MS/MS spectrum of TP1 and TP2. (a) Raw data of EThcD on TP1, ETD reaction time = 10ms, hcD SA = 25%, isolation window = 900-1300 m/z, averaged over 0.8 min. (b) EThcD spectrum of TP2 under the same setting.



Figure S3. TerSeqCov calculation and comparison using different numbers of terminal residues, based on TP2 (a) n = 30, (b) n = 50. Blue bars represent N-TerSeqCovs, and orange bars represent C-TerSeqCovs. Blue and Orange crosses mean the no product ion detected from representative termini. Boxplot analysis shows G12-G247 as an outlier for 27,402 Da search compared with other sequences. The box portion shows 25-75% percentile; the whisker represents 1.5 interquartile range; the horizontal line in the box stands for median value.





Figure S4. Sequence coverage analyzed by ProSight. (a) S14-S436 from TP1, (b) S5-N427 from TP1, (c) G12-G247

from TP2.

а																											b																								
N-terminal	s	s	G	т	Q	N	K	A	L	₽	E]	N	V1	к	Y1	G1	I	v	L	D1	A]	G1	S	s	н		N-termina	S		s	G T	Q	N	K	A	L	P	Е	N	1	к	YG	I	v	L	D	A	G	s	S I	н
	т	1 S	11	1Y	II	1Y	K	W	P	A	E]	K1	E1	N	D	т	G	v	v	н	Q	v	Е	Е	С			т		S I	L Y	I	Y	ĸ	W	P	A	Е	K I	5 1	NI	DT	G	v	v	н	Q	v	E	E (C
	R	v	K	G	P	G	I	S	к	F	v	Q	ĸ	v	N	E	I	G	I	Y	L	т	D	C	м			R		VP	K G	P	G	I	S	ĸ	F	v	QI	< 1	V I	NE	I	G	I	Y	L	т	D	C I	M
	Е	R	A	R	Е	v	I	Ρ	R	s	Q	н	Q	Е	т	P	v	Y	L	G	A	т	A	G	М			E	1	R Z	AR	E	v	I	P	R	s	Q	H	2 1	E	T P	v	Y	L	G	А	т	A	GI	M
	R	L	L	R	М	E	S	Е	Е	L	A	D	R	v	L	D	v	v	Е	R	s	L	s	N	Y			R	1	LI	L R	M	E	S	Е	Е	L	A	DI	2 1	v	L D	v	v	Е	R	S	L	S	N	Y)
	P	F	D	F	Q	G	A	R	I	I	т	G	Q	E	E	G.	A	Y	G	W	I	т	I	N	Y			P	1	FI	F	· Q	G	A	R	I	I	т	G	2 1	E	EÌG	A	Y	G	W	I	т	I	N	Y
	L	L	G	K	F	S	Q	к	N	Q	Е	т	F	G	A	L	D	L	G	G	A	s	т	Q	v			L	. 1	LO	G K	F	S	Q	ĸ	N	Q	Е	т	7 (G	AL	D	L	G	G	A	S	т	Q '	v
	т	F	v	P	Q	N	Q	т	I	Е	S	P	D	N	A	L	Q	F	R	L	Y	G	к	D	Y			т	1	F١	/ P	2	N	Q	т	I	Е	s	PI) 1	N	AL	Q	F	R	L	Y	G	K	D	Y
	N	v	Y	т	H	S	F	L	С	Y	G	к	D	Q	A	L	W	Q	ĸ	L	A	к	D	I	Q			N	1	v	(I	H	S	F	L	С	Y	G	KI)	Q i	AL	W	Q	ĸ	L	A	к	D	I	2
	v	A	S	N	E	I	L	R	D	P	С	F	н	P	G	Y	ĸ	к	v	v	N	v	S	D	L			v	1	A S	5 N	E	I	L	R	D	P	С	FI	1	P	G Y	K	K	v	v	N	v	S	D	L
	Y	K	Т	P	С	т	к	R	F	Е	м	т	L	P	F	Q	Q	F	E	I	Q	G	I	G	N			Y	1	кл	C P	c	т	к	R	F	Е	м	т	5	P	FQ	Q	F	Е	I	Q	G	I	G 1	N
	Y	Q	Q	С	H	Q	s	I	L	Е	L	F	N	т	s	Y	С	P	Y	s	Q	С	A	F	N			Y		2 9	2 0	н	Q	s	I	L	Е	L	FI	1	т	S Y	С	P	Y	s	Q	С	A	F I	N
	G	I	F	L	P	P	L	Q	G	D	F	G	A	F	S	A	F	Y	F	v	м	ĸ	F	L	N			G	: :	IF	F L	P	P	L	Q	G	D	F	G	4	F	SA	F	Y	F	v	М	к	F	LI	N
	L	т	S	E	K	v	s	Q	Е	к	v	т	Е	M	м	ĸ	К	F	С	A	Q	Ρ	W	Е	Е			L		r s	S L E	K	v	S	Q	Е	к	v	т	2 1	MI	MK	K	F	С	A	Q	P	W	E	2
	I	K	т	S	Y	A	G	v	к	Е	к	Y	L	s	Е	Y	С	F	s	G	т	Y	I	L	s			I	1	кл	C S	Y	A	G	v	к	E	ĸ	Y I	4	S I	EY	C	F	s	G	т	Y	I	L	S
	L	L	L	Q	G	Y	н	F	т	A	D	s	W	Е	н	I	H	F	I	G	к	I	Q	G	S			L	. 1	LI	L C	G	Y	н	F	т	A	D	SI	1	E	HI	H	F	I	G	K	I	Q	G	S
	D	A	G	W	LT	L	G	Y	М	L	N	L	т	N	М	I	P.	A	Е	Q	P	L	S	<u>C-t</u>	ern	ninal		D	1	A C	G W	Т	L	G	Y	М	L	N	r ľ.	C L I	NU	MLI	L P	A	E	Q	P	L	S	C-te	rminal

Figure S5. ETD and CID Sequence coverage analysis of TP1. Bars stand for detected product ions: red bars represent c/z ions, and blue bars represent b/y ions (a) ETD, (b) CID.



Figure S6. Western blot shows potential degradation of α chain on the left. The middle band is not relevant to this manuscript; the right band is the molecular weight ladder.

The full sequence of TP1 is:

APTSSSTKKTQLTSSGTQNKALPENVKYGIVLDAGSSHTSLYIYKWPAEKENDTGVVHQVEECRVKGPGI SKFVQKVNEIGIYLTDCMERAREVIPRSQHQETPVYLGATAGMRLLRMESEELADRVLDVVERSLSNYPF DFQGARIITGQEEGAYGWITINYLLGKFSQKNQETFGALDLGGASTQVTFVPQNQTIESPDNALQFRLYGK DYNVYTHSFLCYGKDQALWQKLAKDIQVASNEILRDPCFHPGYKKVVNVSDLYKTPCTKRFEMTLPFQQF EIQGIGNYQQCHQSILELFNTSYCPYSQCAFNGIFLPPLQGDFGAFSAFYFVMKFLNLTSEKVSQEKVTEM MKKFCAQPWEEIKTSYAGVKEKYLSEYCFSGTYILSLLLQGYHFTADSWEHIHFIGKIQGSDAGWTLGYM LNLTNMIPAEQPLSTPLSHST

The full sequence of TP2 chain A is:

ADSGEGDFLAEGGGVRGPRVVERHQSACKDSDWPFCSDEDWNYKCPSGCRMKGLIDEVNQDFTNRIN KLKNSLFEYQKNNKDSHSLTTNIMEILRGDFSSANNRDNTYNRVSEDLRSRIEVLKRKVIEKVQHIQLLQKN VRAQLVDMKRLEVDIDIKIRSCRGSCSRALAREVDLKDYEDQQKQLEQVIAKDLLPSRDRQHLPLIKMKPV PDLVPGNFKSQLQKVPPEWKALTDMPQMRMELERPGGNEITRGGSTSYGTGSETESPRNPSSAGSWN SGSSGPGSTGNRNPGSSGTGGTATWKPGSSGPGSTGSWNSGSSGTGSTGNQNPGSPRPGSTGTWN PGSSERGSAGHWTSESSVSGSTGQWHSESGSFRPDSPGSGNARPNNPDWGTFEEVSGNVSPGTRRE YHTEKLVTSKGDKELRTGKEKVTSGSTTTTRRSCSKTVTKTVIGPDGHKEVTKEVVTSEDGSDCPEAMDL GTLSGIGTLDGFRHRHPDEAAFFDTASTGKTFPGFFSPMLGEFVSETESRGSESGIFTNTKESSSHHPGI AEFPSRGKSSSYSKQFTSSTSYNRGDSTFESKSYKMADEAGSEADHEGTHSTKRGHAKSRPVRGIHTSP LGKPSLSP