Extended Snake Venomics by Top-Down In-Source Decay: Investigating the Newly Discovered Anatolian Meadow Viper Subspecies, *Vipera anatolica senliki*•

Benjamin-Florian Hempel[a,†], Maik Damm[a,†], Mrinalini[b,†], Bayram Göçmen[c], Mert Karış[c,‡], Ayse Nalbantsoy[d], R. Manjunatha Kini[b,e] and Roderich D. Süssmuth[a,*]

[a] Department of Chemistry, Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany.

[b] Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543.

[c] Department of Biology, Faculty of Science, Ege University, 35100 Bornova, Izmir, Turkey.

[d] Department of Bioengineering, Faculty of Engineering, Ege University, Bornova, 35100 Izmir, Turkey.

[e] Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, 16, Medical Drive, Singapore 117600.

• This Paper is dedicated to the memory of Professor Bayram Göçmen

* Correspondence to Prof. Roderich D. Süssmuth; Phone: +49 (0)30 314-78774; Email: roderich.suessmuth@tu-berlin.de

† These authors contributed equally to this work

‡ Current address: Department of Chemistry and Chemical Process Technologies, Acıgöl Vocational High School of Technical Sciences, Nevşehir Hacı Bektaş Veli University, 50300 Nevşehir, Turkey.

The following supporting information are included.

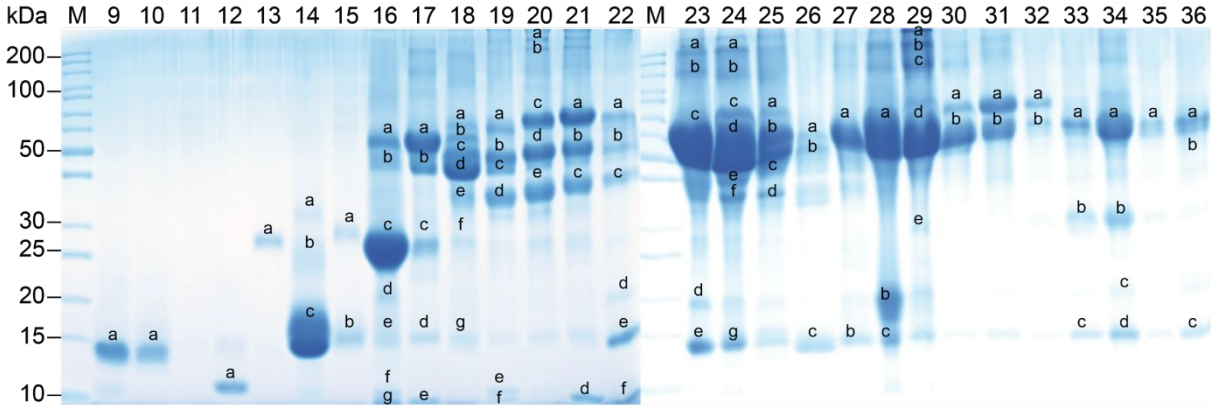

Figure S1. SDS-PAGE fractions of *Vipera anatolica senliki* venom under reducing conditions. RP-HPLC venom fractions shown in Figure 1 were further processed by SDS-PAGE analysis. Fraction numbers are indicated above the lanes. Nomenclature shows selected bands for tryptic in-gel digestion and subsequent bottom-up venomics.

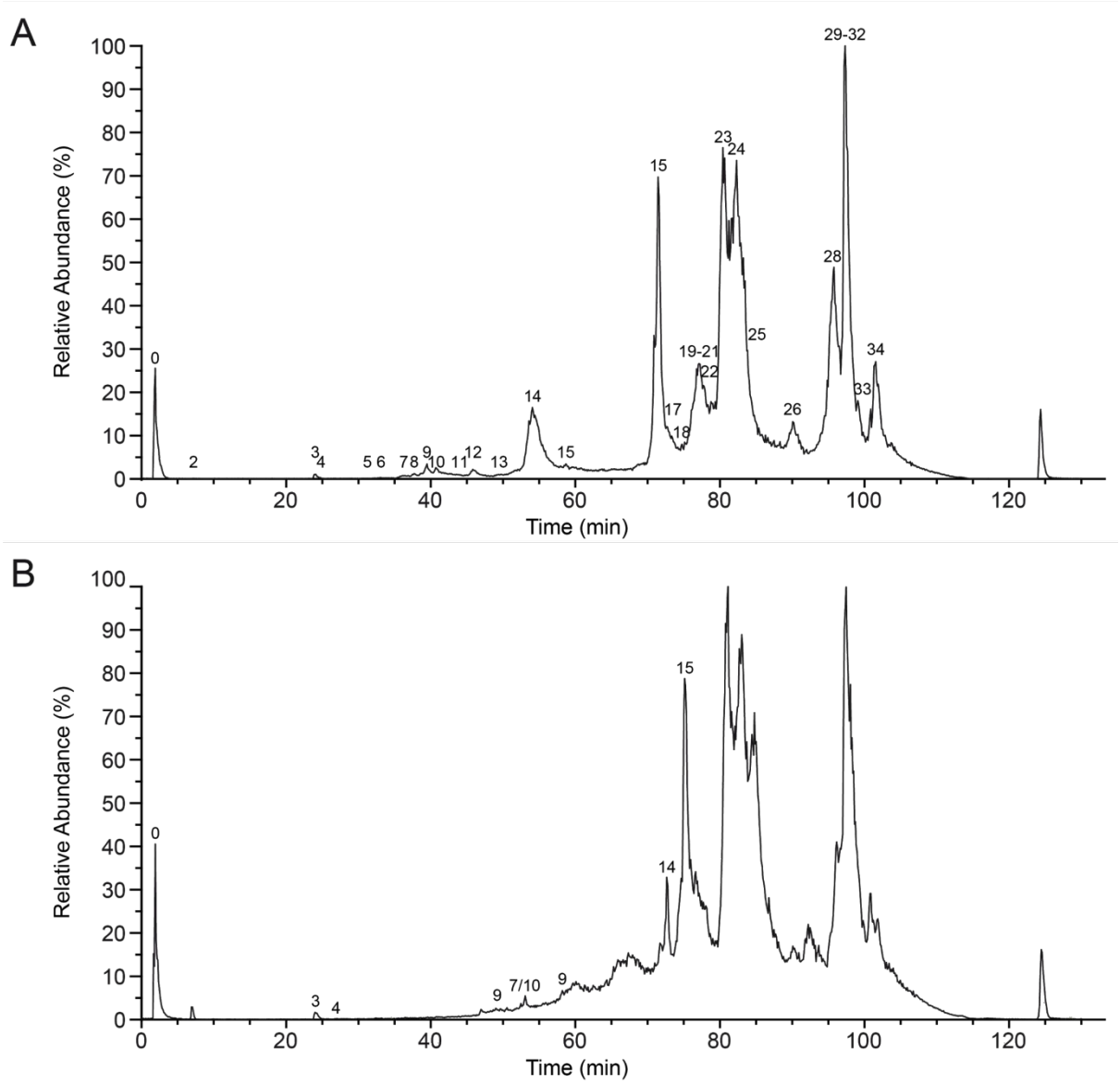

Figure S2. Extended snake venomic analysis of native and chemically reduced *Vipera anatolica senliki* crude venom. Total ion chromatogram (TIC) from (**A**) native and (**B**) reduced *V. a. senliki* venom for IMP. The total ion counts were measured by HPLC-ESI-MS and the relative abundance was set to 100% for the highest peak. Fraction nomenclature based on Figure 1.

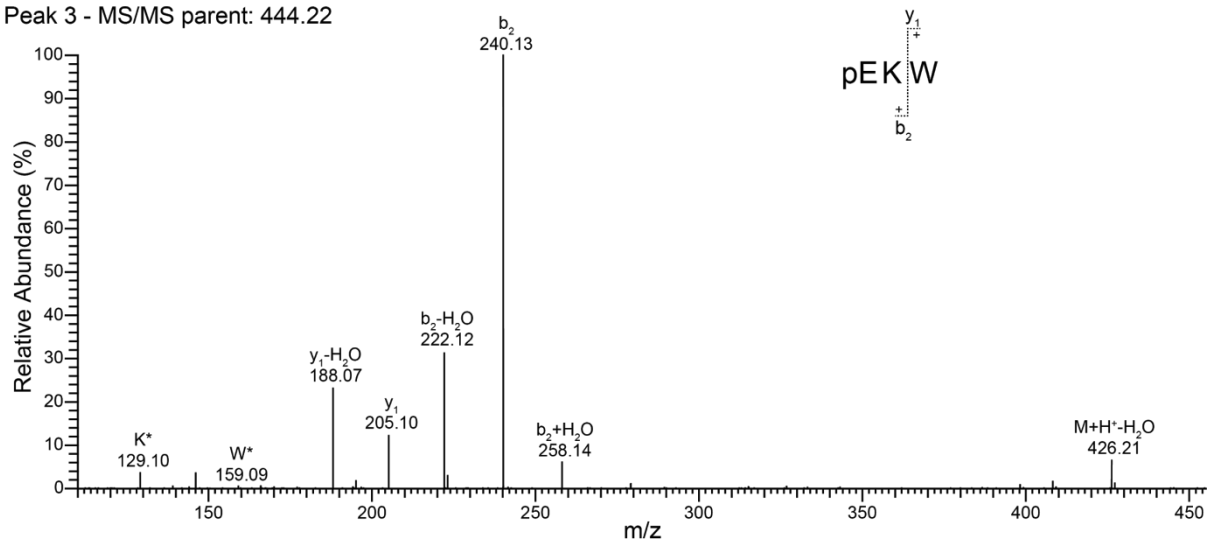

Figure S3. Tandem MS spectrum of the tripeptidic metalloprotease inhibitor pEKW. Representative MS/MS spectra of a small tripeptic svMP inhibitor (svMP-i) with *m/z* 444.22 precursor ion mass for *de novo* annotation in the *Vipera anatolica senliki* venom.

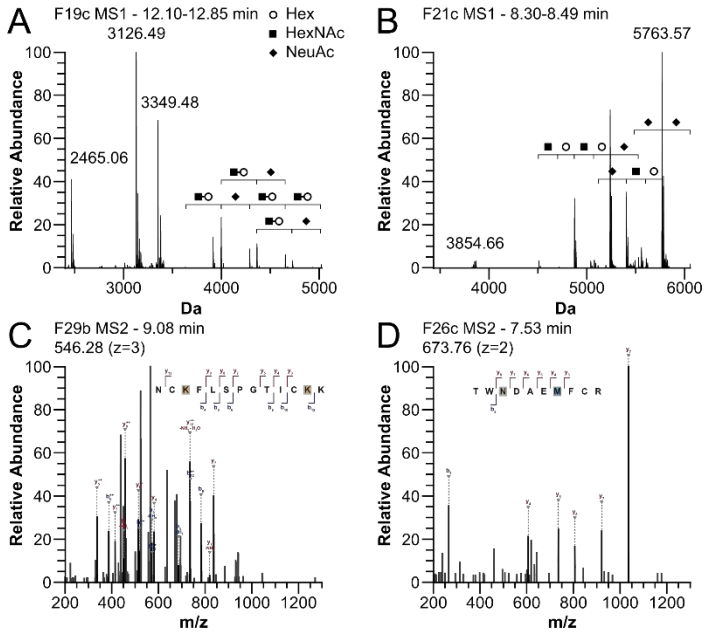

Figure S4. Representative MS/MS spectra of various bottom-up fractions showing post-translational modifications (PTMs). Deconvoluted MS spectra of two svSP fraction bands (**A**) F 19c and (**B**) F 21c with the appearance of three different glycosylation building blocks: hexose (Hex), *N*-acetyl-hexoseamine (HexNAc) and *N*-acetyl-neuraminic (NeuAc). The MS/MS spectra show (**C**) natural related Lys acetylations of a svMP and (**D**) other modifications such as experimental artefacts (e.g. Asn deamidation and Met oxidation).

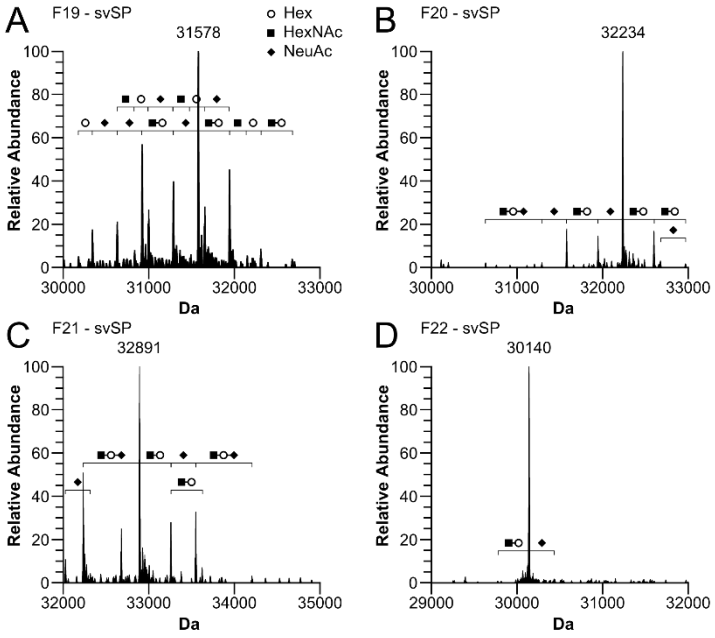

Figure S5. Identification of glycosylated svSP proteoforms by intact mass profiling. Several glycosylation building blocks, as hexose (Hex), *N*-acetyl-hexoseamine (HexNAc) and *N*-acetyl-neuraminic (NeuAc), were observed by IMP of snake venom serine proteases (svSP) in (**A**) F 19, (**B**) F 20, (**C**) F 21 and (**D**) F 22, with a typical glycosylation branch pattern HexNAc-Hex-NeuAc.

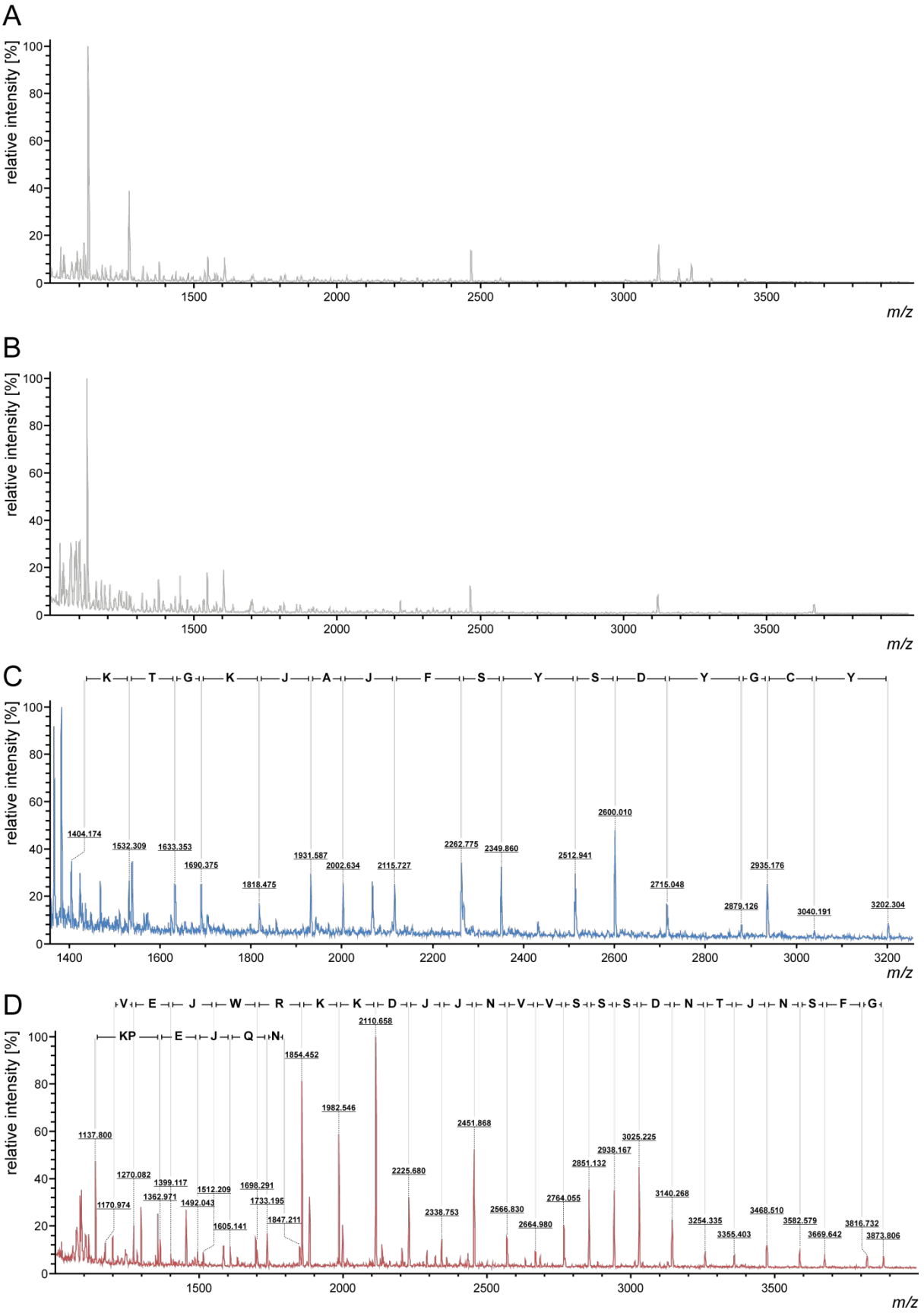

Figure S6. MALDI top-down sequencing by in-source decay of different venom components from *Vipera a. senliki.* (**A**) Examples of top-down ISD spectra of peptide fractions (F 8) and (**B**) (F 9/10) showing no distinct sequences. (**C**) Identification of a phospholipase A_2 (PLA_2) proteoform (ammodytin I2 (D)) by N-terminal sequence (F 14/15). (**D**) Identification of a short cysteine-rich venom protein (CRISP) peptide fragment by N-terminal sequencing, previously annotated in peak 16, as well as a snake venom metalloproteinase (svMP) proteoform by N-terminal sequencing (F 17). No distinction can be made between leucine and isoleucine (J = Leu or Ile).

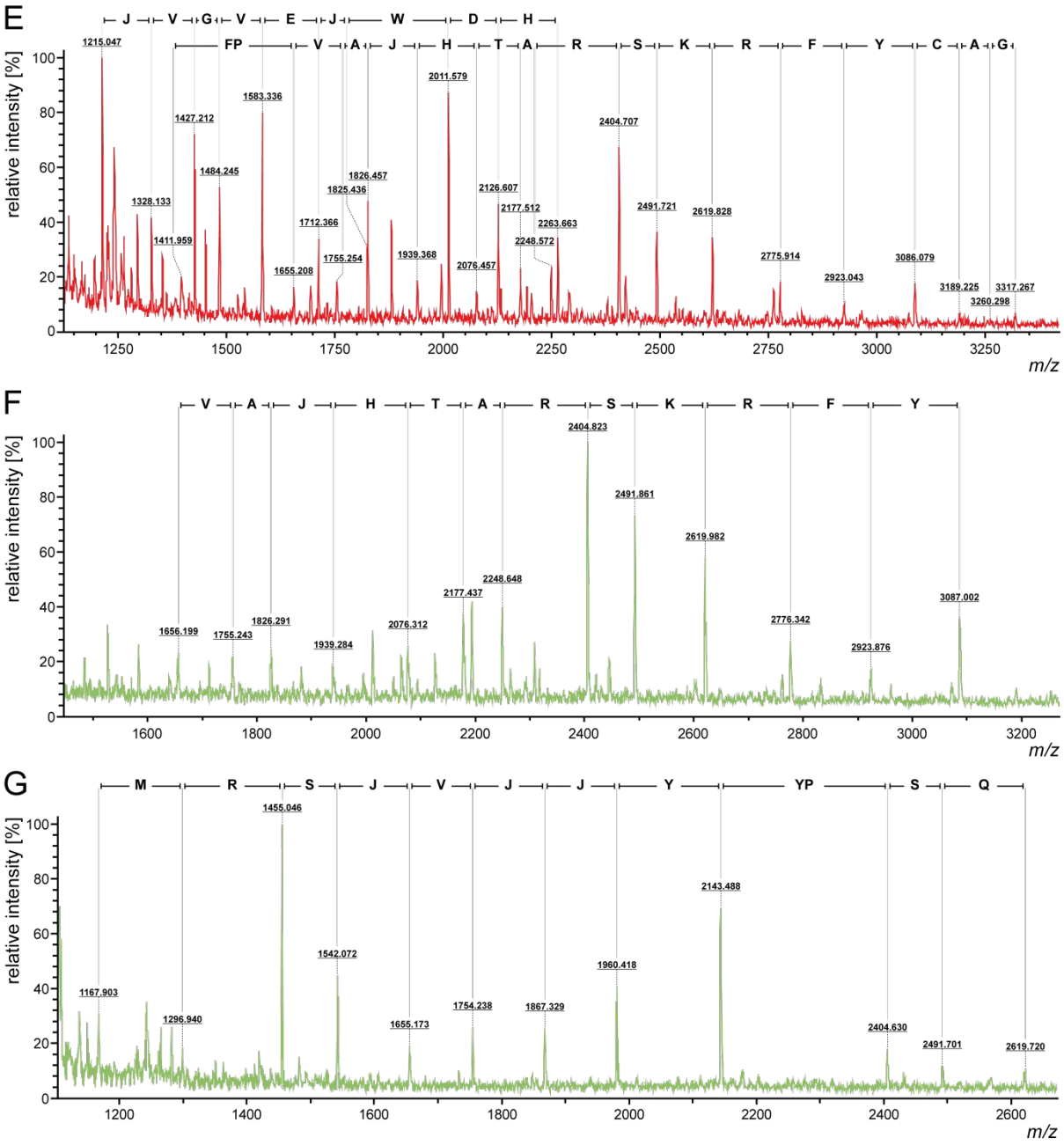

Figure S6. ***(continued)*** **MALDI top-down sequencing by in-source decay of different venom components from** ***Vipera a. senliki.*** (**E**) Identification of a svMP proteoform by N-terminal sequencing (F 18-22). (**F**) Identification of a snake venom serine protease (svSP) proteoform by N-terminal sequencing (F 23) with a specific transcriptome hit (MN831307). (**G**) Identification of a svSP proteoform by N-terminal sequencing (F 24). No distinction can be made between leucine and isoleucine (J = Leu or Ile).

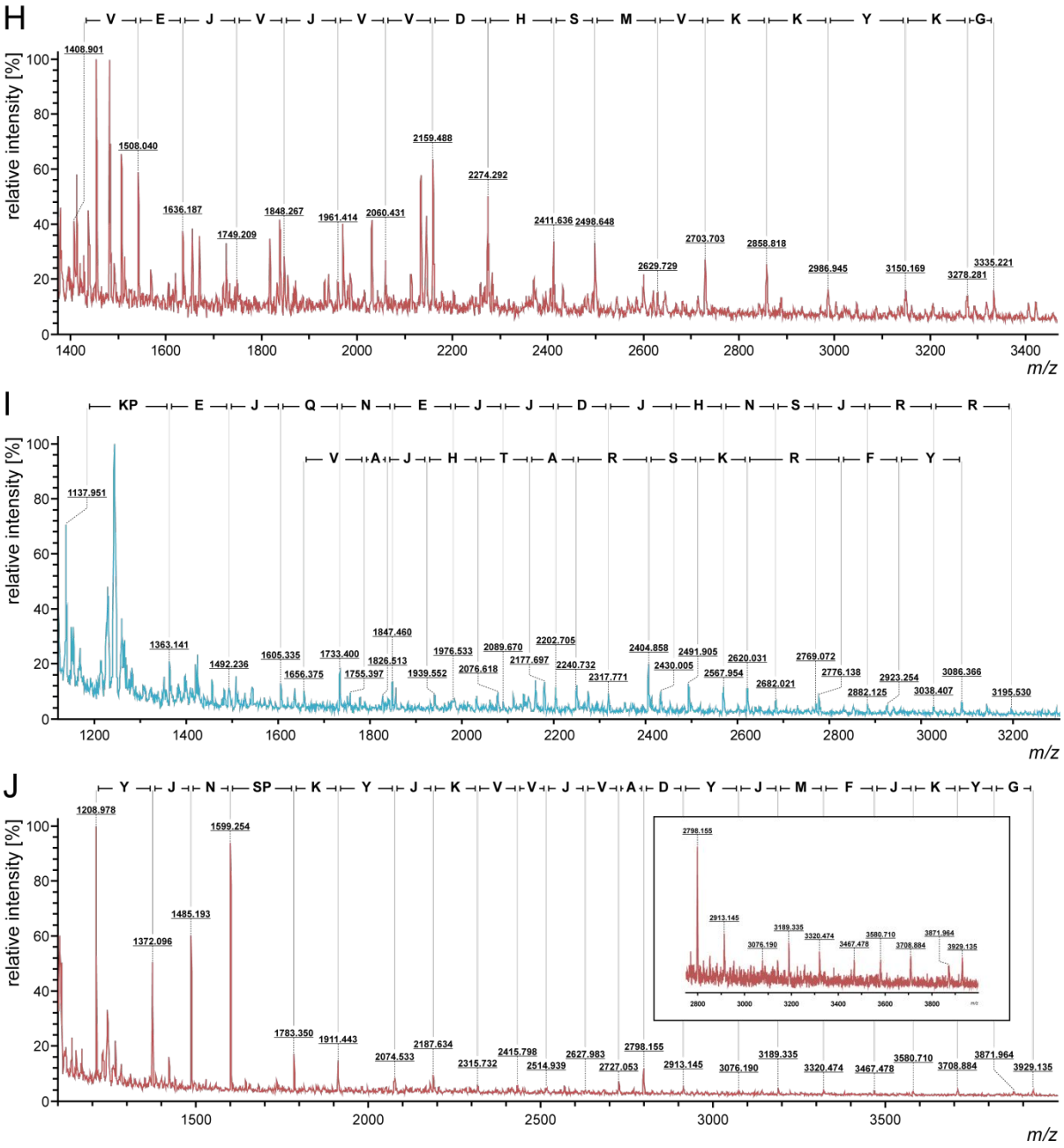

Figure S6. ***(continued)*** **MALDI top-down sequencing by in-source decay of different venom components from** ***Vipera a. senliki*****.** (**H**) Identification of a svMP proteoform by N-terminal sequencing (F 25) with a specific transcriptome hit (MN831329). (**I**) Identification of a CRISP proteoform with a specific transcriptome hit (MN831241) and a svSP proteoform with a specific database hit (MN831307) (F 27). (**J**) Identification of a svMP proteoform by N-terminal sequencing (F 33-35). No distinction can be made between leucine and isoleucine (J = Leu or Ile).

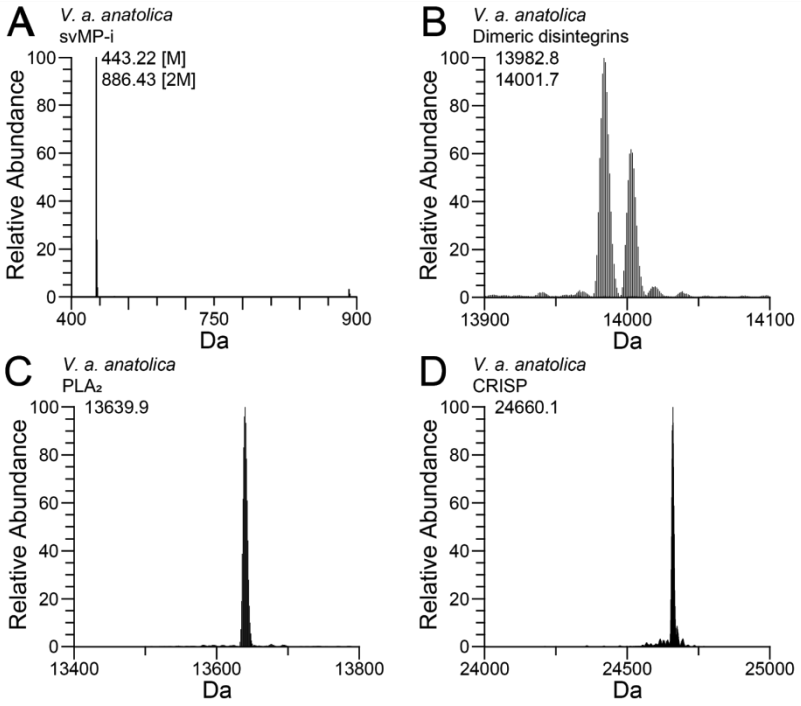

Figure S7. Intact mass profiling of exemplary *Vipera anatolica anatolica* venom components. *V. a. anatolica* (Göcmen et al.[1]) shows compared to *V. a. senliki* (**Figure 2** and **Table S2**) identical toxin masses, like (**A**) svMPI-i, (**B**) dimeric disintegrins, (**C**) PLA_2 and (**D**) CRISP.

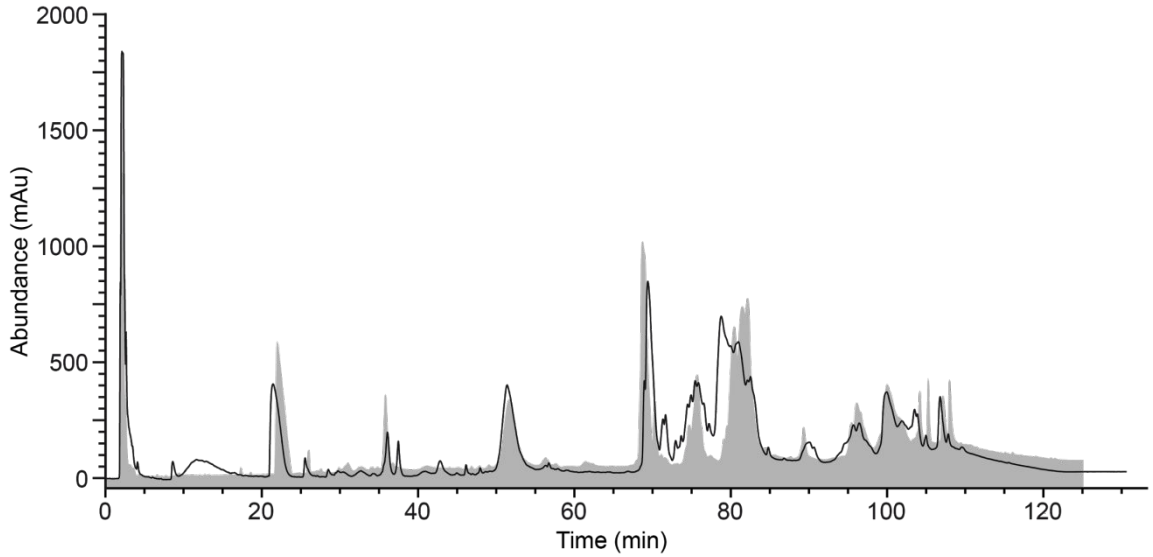

Figure S8. Overlay of C18 RP-HPLC venom profile from *Vipera anatolica senliki* and *Vipera anatolica anatolica*. HPLC venom profile of *V. a. senliki* (black line) is shown compared to the *V. a. anatolica* (grey) analysis by Göçmen et al.[1]. Same venom amounts of venom were measured on identical devices and column.

Table S1. Transcriptome assembled sequences identified by snake venom gland transcriptomics from *Vipera anatolica senliki*. List of full length and partial length toxin transcripts identified by de novo transcriptome assembly of the *Vipera anatolica senliki* venom gland. GenBank accession numbers and functional annotations are provided along with the average and monoisotopic masses of peptide hits from proteomics analysis.

Table S2. Venom proteins and peptides identified from *Vipera anatolica senliki.* Assignments of venomic components by crude venom intact mass profiling (IMP, method A), IMP of a single RP-HPLC fraction with low molecular mass (method B), bottom-up (BU, method C) and in-source decay annotation (ISD, method D; identified sequences marked in green). Fraction numbers are based on the RP-HPLC chromatogram (**Figure 1**). Annotation was performed *de novo* and by peptide spectrum matching from in-gel digested protein bands (**Figure S1**). Peak 0 corresponds with injection peak Identification was carried out against a non-redundant *Viperidae* protein database (taxid: 8689), our custom transcriptome database and a set of proteins found as common contaminants (cRAP). SDS-PAGE and intact mass profile analysis provided the average molecular weight. Most abundant mass for IMP analysis s is marked by *. IMP performed by charge-state deconvolution was carried out with MagicTransformer (MagTran; marked by #) and isotopic resoloved deconvolution with Thermo Xtract.

Table S3. Compositional venom lineup of two *Vipera anatolica* subspecies. The most abundant toxin families in the venoms of *Vipera anatolica senliki* (this study) and *Vipera anatolica anatolica* (Göçmen *et al.*[1]) are compared by their HPLC retention time. Venoms were measured on identical devices and column. Identical identified masses in the IMP are mentioned in the correspondent row.

V. a. senliki			***V. a. anatolica***		**identical IMP masses in both venoms in Da**
Peak No.	**most abundant fraction components**	**t_R in min**	**t_R in min**	**most abundant fraction components**	
0	peptides	2	2	peptides	
1	peptides	9	-	-	-
2	peptides	12	-	-	-
-	-	-	17	peptides	-
-	-	-	18	peptides	-
3	svMP-i (pEKW)	22	23	svMP-i (pEKW)	443.2; 886.4
4	BPP; peptides	26	26	peptides	429.2; 752.4; 1128.6
-	-	-	28	peptides	-
5	BPP; peptides	29	30	peptides	-
6	unknown peptides	30	31	peptides	808.39; 1128.61; 3943.81
7	Kunitz-inhibitor; peptides	33	33	Kunitz-inhibitor, peptides	6738.0; 7280.2
8	Kunitz-inhibitor; peptides	34	34	peptides	3119.46; 3233.50; 3421.59
9	DI	36	36	DI	13982.8; 14001.8
10	DI	38	-	-	-
11	Kunitz-inhibitor	41	-	-	-
12	Kunitz-inhibitor	43	-	-	-
13	unknown protein	46	47	peptides	1101.5
-	-	-	49	peptides	-
14	PLA_2	52	52	PLA_2	13639.9
15	PLA_2	57	56	unknown protein	11143.6; 11394.6; 13639.9
-	-	-	62	PLA_2	-
16	CRISP	70	69	CRISP	24660.1
-	-	-	70	CRISP	-
17	svMP; CRISP	72	71	svMP; CRISP	-
18	svMP	73	-	-	-
19	svMP; svSP	75	75	unknown protein; svSP	-
20	svMP; svSP	76	76	svMP	-
21	svMP; svSP	77	-	-	-
22	svMP; svSP; CTL	78	78	unknown protein	-
23	svMP; CTL	79	-	-	-
-	-	-	80	svMP	-
24	svMP	81	81	svMP	-
25	svMP	83	82	svMP	-
26	svMP; CTL	85	85	svMP	-
27	svMP; CTL	90	89	unknown protein	-
-	-	-	95	unknown protein	-
28	svMP; CTL	97	97	CTL	-
-	-	-	100	CTL	-
29	svMP	101	101	svMP	-
30	svMP	102	103	svMP	-
31	svMP	104	-	-	-
32	svMP	104	104	svMP	-
33	svMP; CTL	105	105	unknown protein	-
34	svMP; CTL	107	108	svMP	-
35	svMP	108	109	unknown protein	-
36	svMP; CTL	110	-	-	-

References

(1) Göçmen, B.; Heiss, P.; Petras, D.; Nalbantsoy, A.; Süssmuth, R. D. Mass spectrometry guided venom profiling and bioactivity screening of the Anatolian Meadow Viper, Vipera anatolica. *Toxicon : official journal of the International Society on Toxinology* **2015**, *107*, 163–174.