Orthogonal separation and identification of long-chain peptides from scorpion *Buthus martensi Karsch* venom by using two-dimensional mixed-mode reversed phase-reversed phase chromatography coupled to mass spectrometry
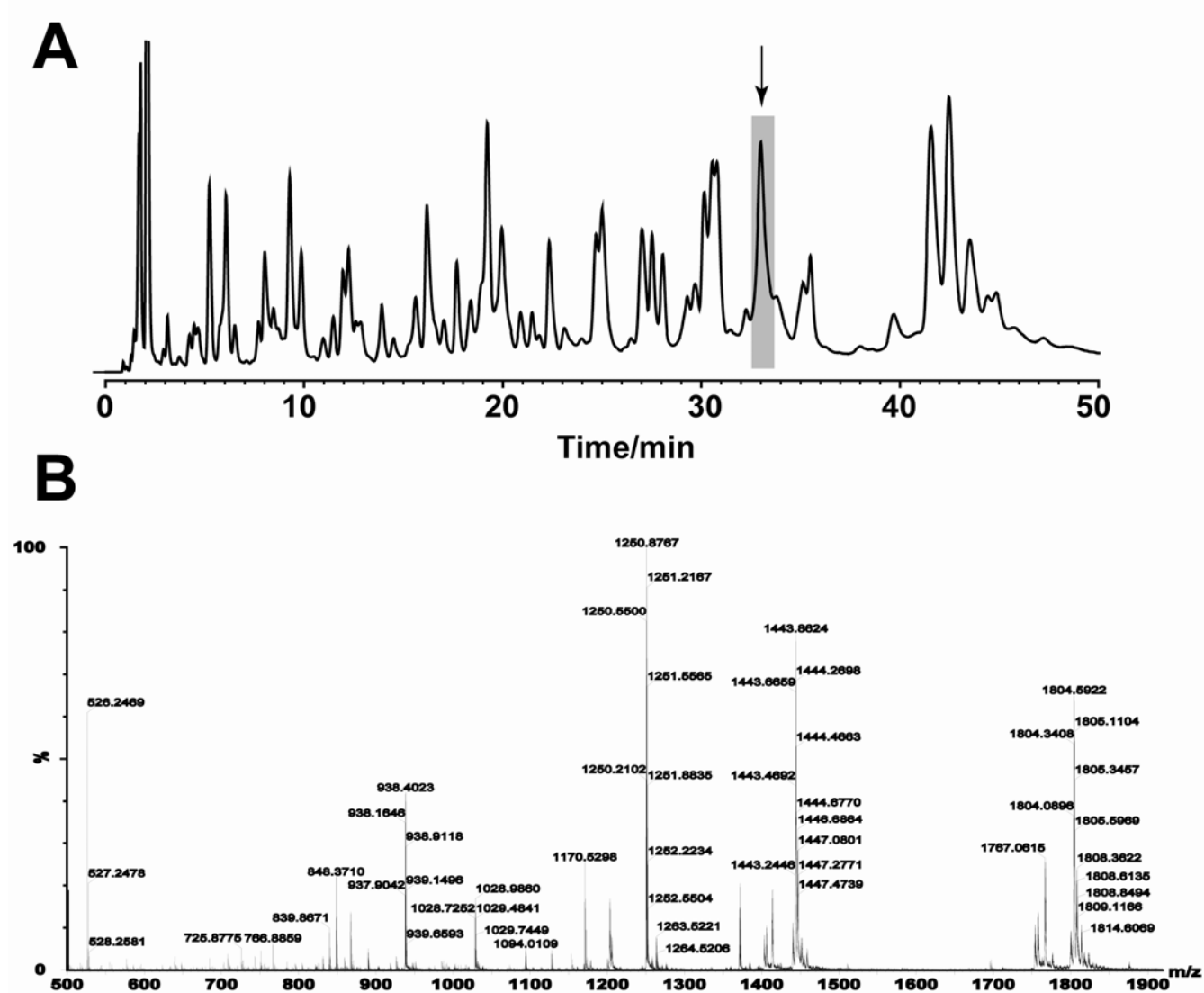

Fig. S1. (A), Sunfire C18 chromatogram of *Bm*K venom using pH 1.8. (B), ESI MS spectrum of the collected fraction (arrow denoted).
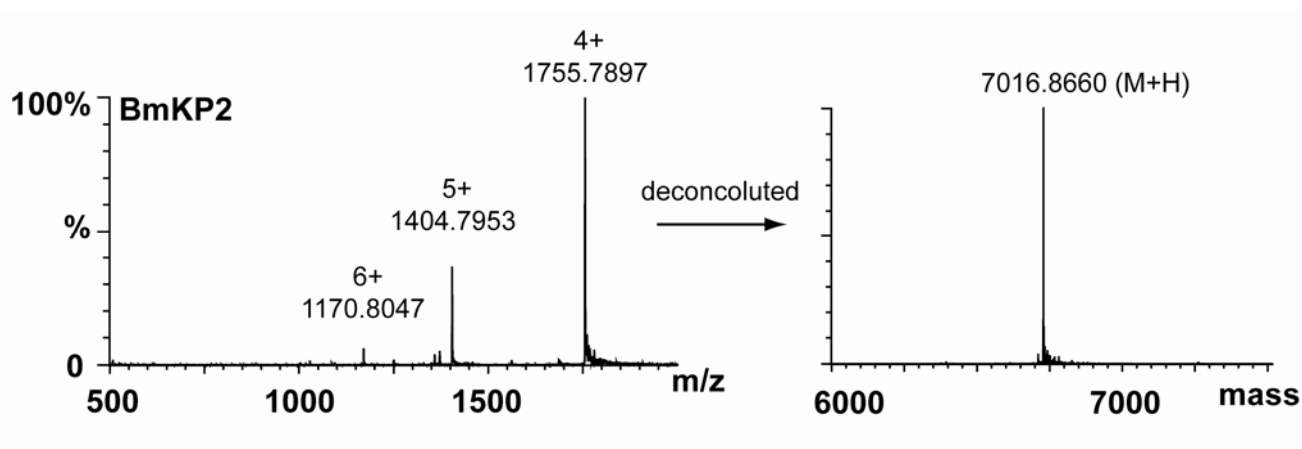
Fig. S2. ESI mass spectrum of native BmKP2 and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software.
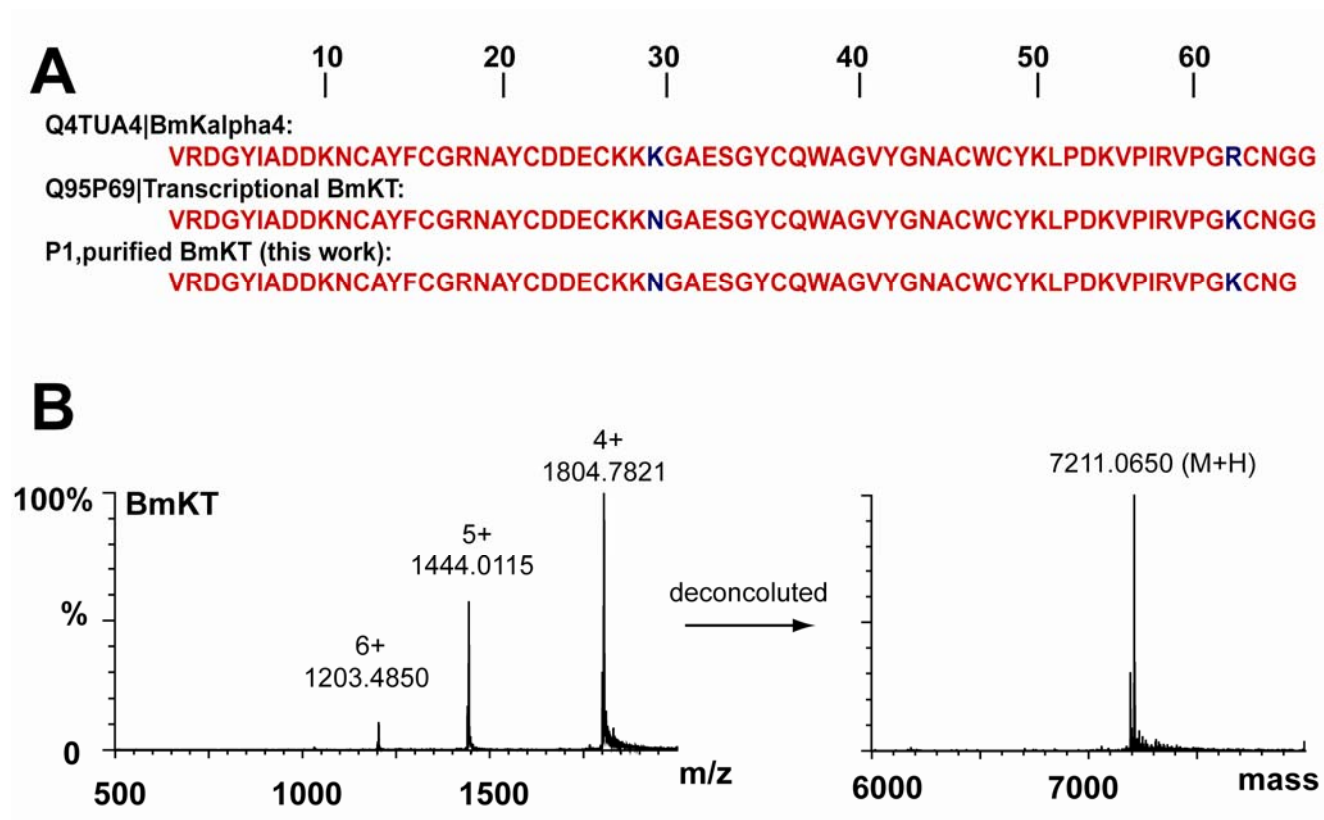

Fig. S3. (A), Sequences comparison of the purified BmKT, transcriptional BmKT and BmKalpha4.

(B), ESI mass spectrum of native BmKT and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software.

A

	10	20	30	40	50	60

Q17231|Transcriptional BmKIT3:
DGYIRGSNGCKVSCLWGNEGCNKECRAYGASYGYCWTWGLACWCEGLPDDKTWKSESNTCGRKK

P5, purified BmKIT3 (this work):
DGYIRGSNGCKVSCLWGNEGCNKECRAYGASYGYCWTWGLACWCEGLPDDKTWKSESNTCG*

Q17230|Transcriptional BmKIT4:
DGYIRGSNGCKISCLWGNEGCNKECKGFGAYYGYCWTWGLACWCEGLPDDKTWKSESNTCGRKK

P6, purified BmKIT4 (this work):
DGYIRGSNGCKISCLWGNEGCNKECKGFGAYYGYCWTWGLACWCEGLPDDKTWKSESNTCG*

B

BmKIT3

4+
1683.9342

5+
1347.3267

deconcoluted →

6728.7139 (M+H)

C

BmKIT4

4+
1691.6632

5+
1353.7238

6+
1128.4275

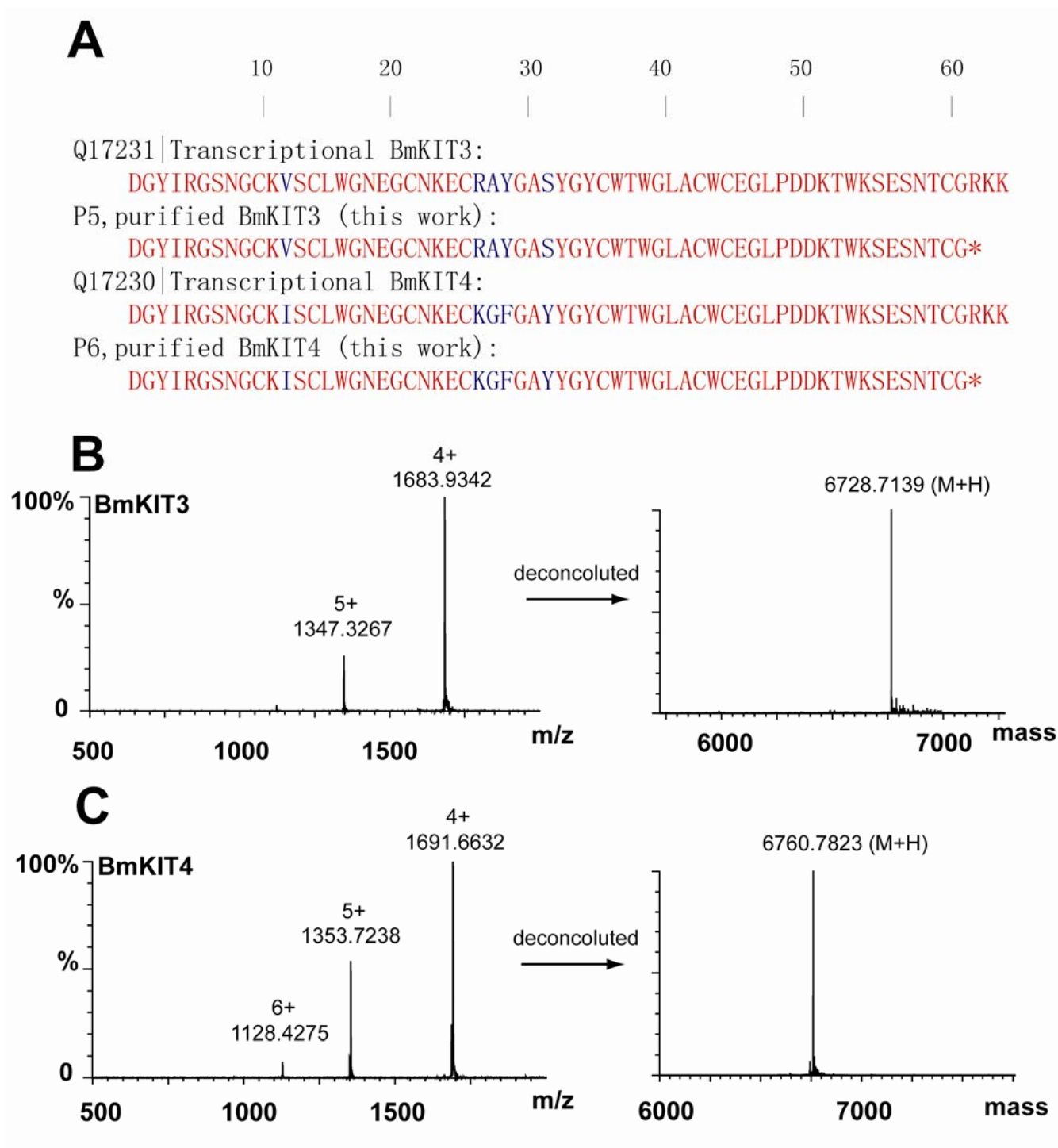deconcoluted →

6760.7823 (M+H)

Fig. S4. (A), Sequences comparison of the transcriptional BmKIT3, purified BmKIT3, transcriptional BmKIT4 and purified BmKIT4. (B), ESI mass spectrum of native BmKIT3 and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software. (C), ESI mass spectrum of native BmKIT4 and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software.
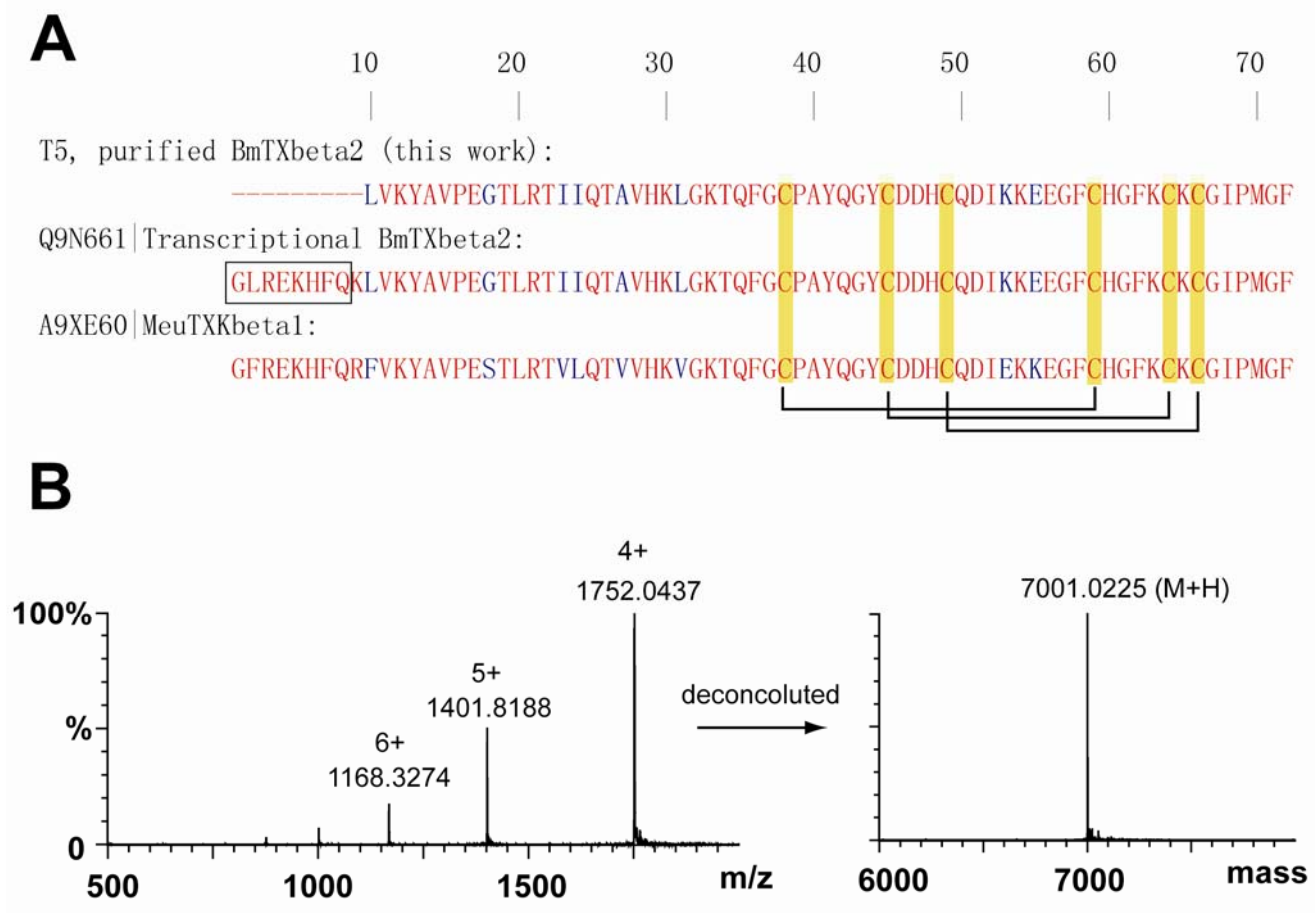
Fig. S5. (A), Sequences comparison of the purified BmTXbeta2, transcriptional BmTXbeta2 and MeuTXKbeta1(from *Buthus eupeus* scorpion). (B), ESI mass spectrum of native BmTXbeta2 and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software.
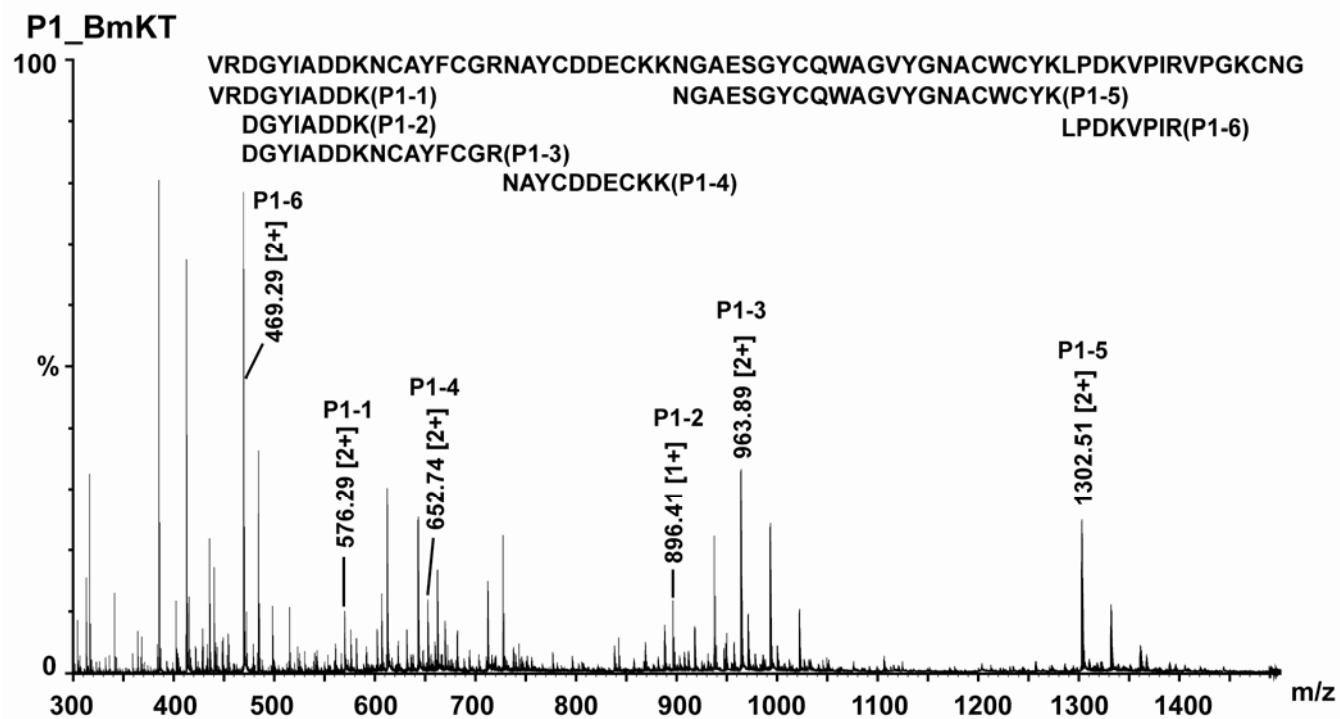
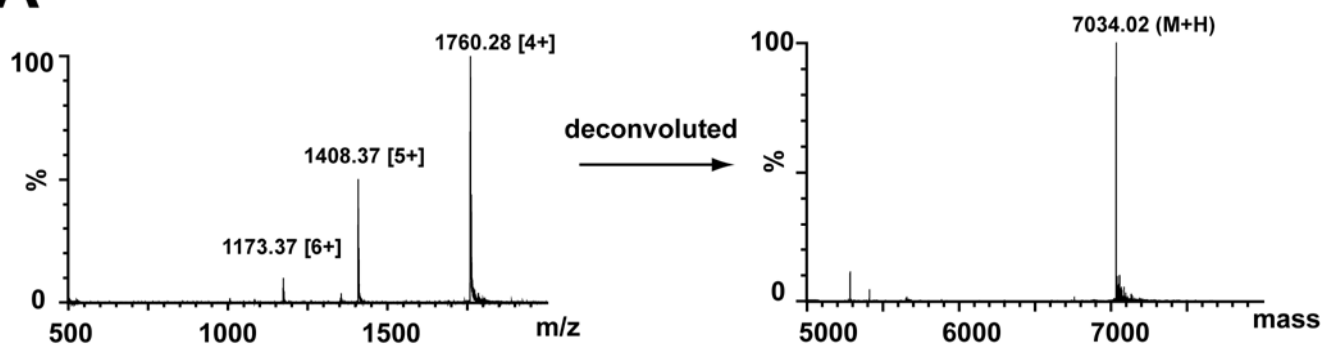Fig. S6. ESI MS spectrum of tryptic *S*-carboxymethylated BmKT.
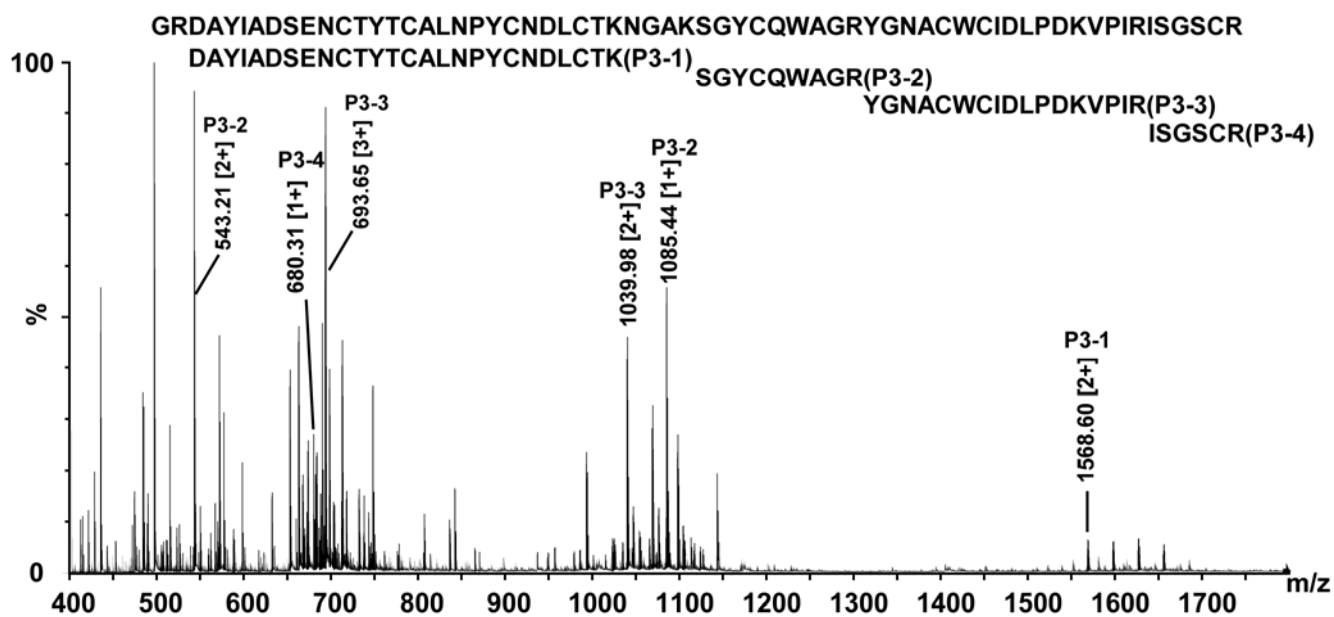
Fig. S7. (A), ESI mass spectrum of native Makatoxin1 and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software. (B), ESI MS spectrum of tryptic *S*-carboxymethylated Makatoxin1.
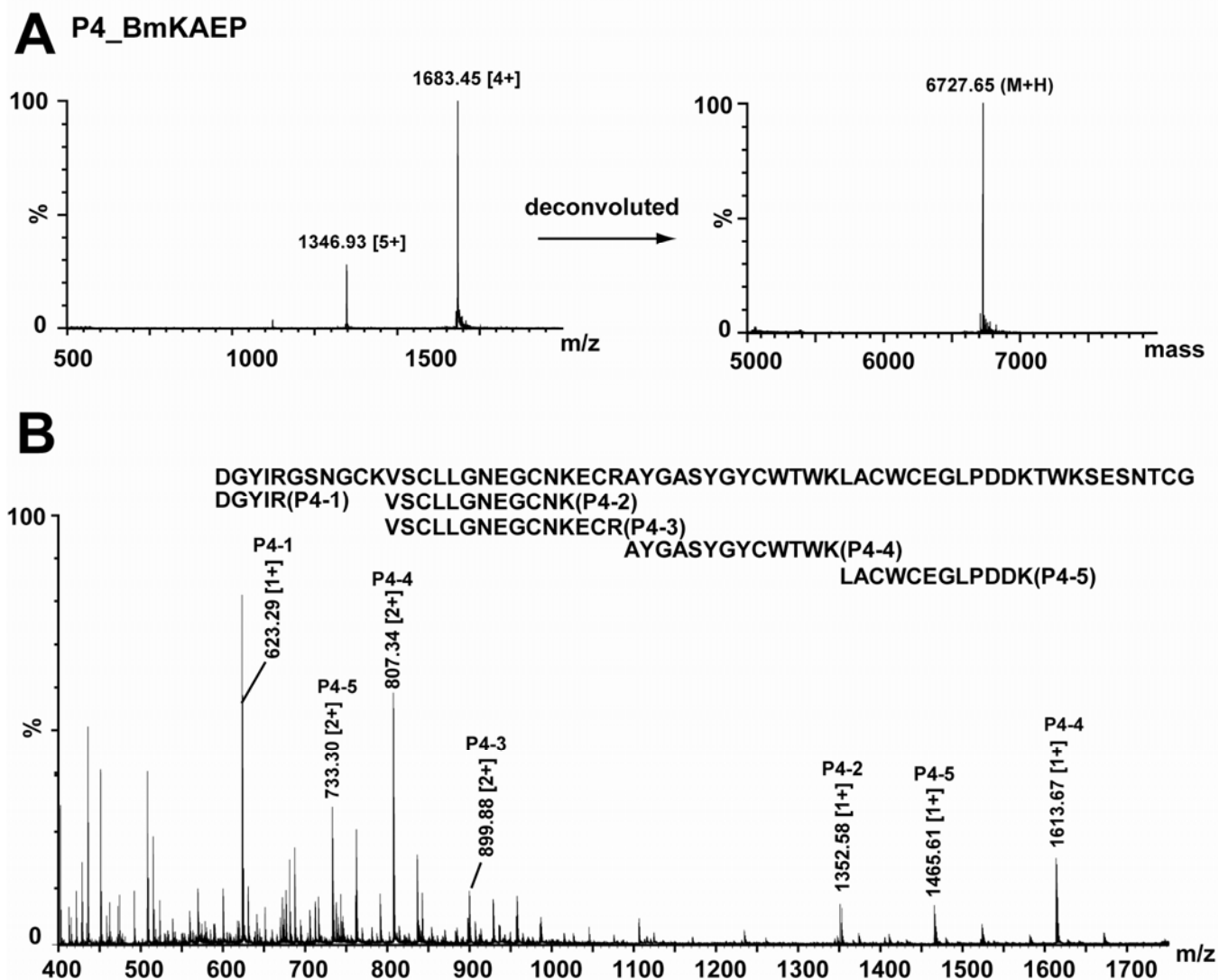
6

Fig. S8. (A), ESI mass spectrum of native BmKAEP and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software. (B), ESI MS spectrum of tryptic *S*-carboxymethylated BmKAEP.

P5_BmKIT3

DGYIRGSNGCKVSCLWGNEGCNKECRAYGASYGYCWTWGLACWCEGLPDDKTWKSESNTCG
DGYIR(P5-1)　　　VSCLWGNEGCNK(P5-2)
　　　　　　　　VSCLWGNEGCNKECR(P5-3)
　　　　　　　　　　　　　AYGASYGYCWTWGLACWCEGLPDDK(P5-4)

Fig. S9. ESI MS spectrum of tryptic *S*-carboxymethylated BmKIT3.

P6_BmKIT4

DGYIRGSNGCKISCLWGNEGCNKECKGFGAYYGYCWTWGLACWCEGLPDDKTWKSESNTCG
DGYIR(P6-1)　　　ISCLWGNEGCNK(P6-2)
　　　　　　　ISCLWGNEGCNKECK(P6-3)
　　　　　　　　　GFGAYYGYCWTWGLACWCEGLPDDK(P6-4)

Fig. S10. ESI MS spectrum of tryptic *S*-carboxymethylated BmKIT4.

8

A T1_BmP09

B

DNGYLLNKYTGCKIWCVINNESCNSECKLRRGNYGYCYFWKLACYCEGAPKSELWAYETNKCNGKM (M : M Sulfoxide)
DNGYLLNK(T1-1)   SELWAYETNK(T1-6)
            IWCVINNESCNSECK(T1-2)
                      RGNYGYCYFWK(T1-3)
                      GNYGYCYFWK(T1-4)
                          LACYCEGAPK(T1-5)

Fig. S11. (A), ESI mass spectrum of native BmP09 and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software. (B), ESI MS spectrum of tryptic *S*-carboxymethylated BmP09.
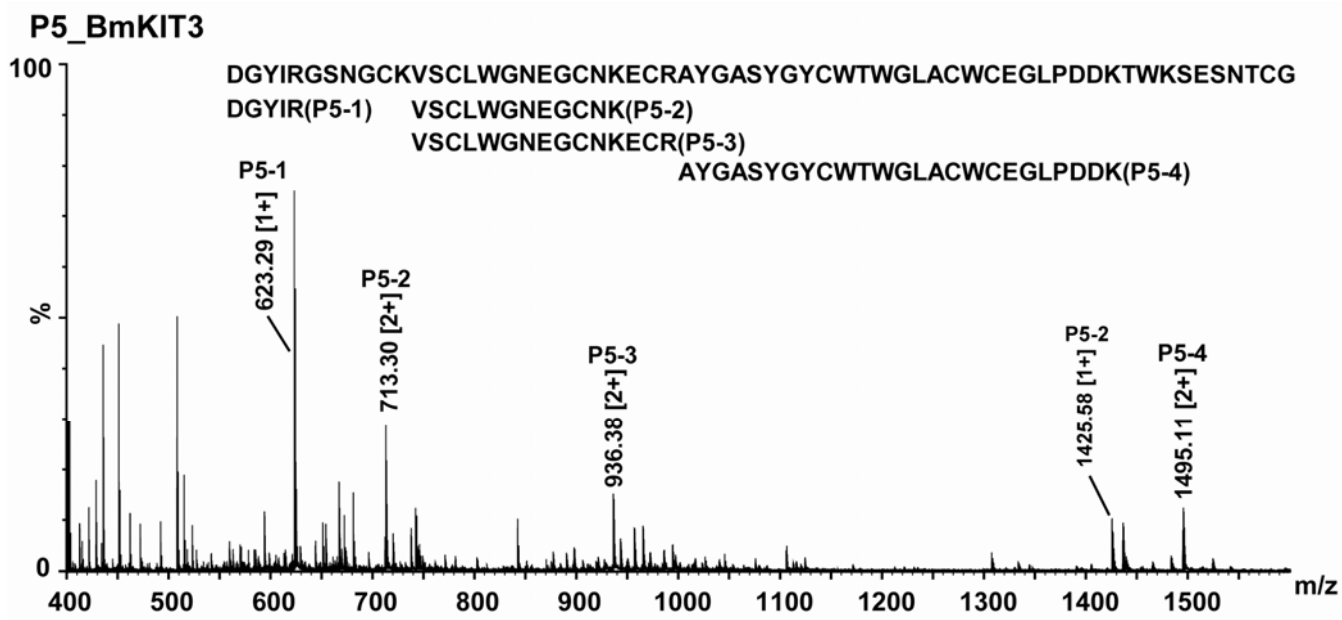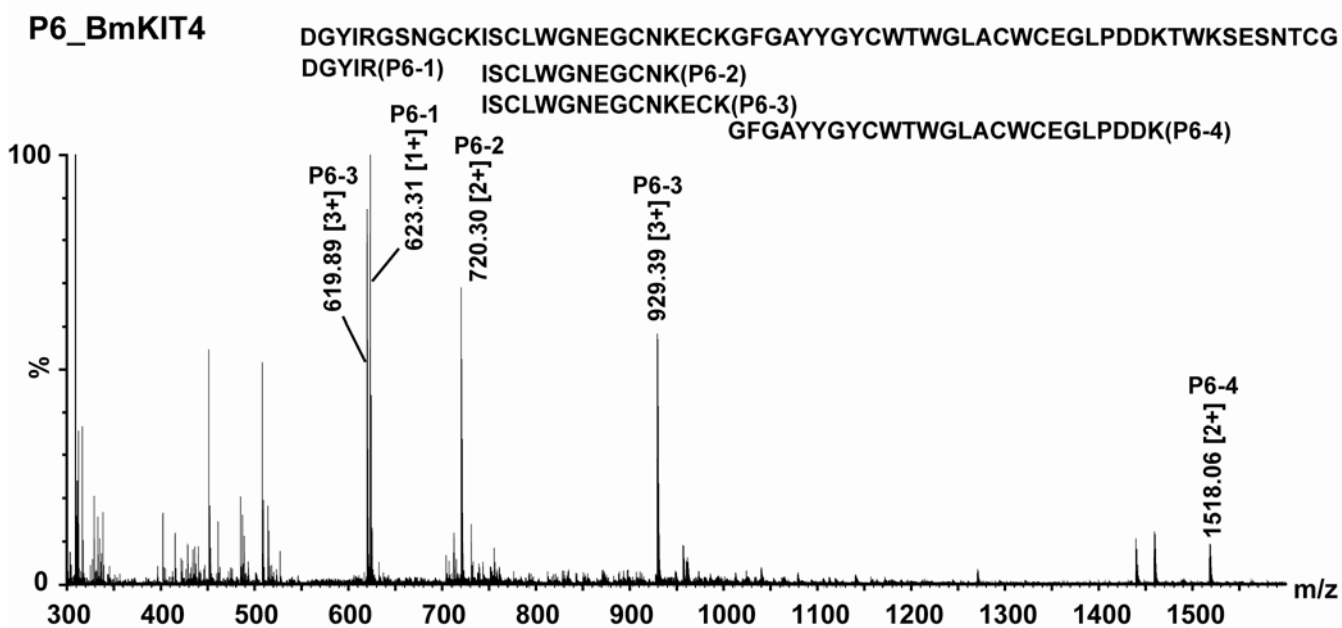
A T2_BmKAS1

B

DNGYLLNKYTGCKIWCVINNESCNSECKLRRGNYGYCYFWKLACYCEGAPKSELWAYETNKCNGKM
DNGYLLNK(T2-1)   LACYCEGAPK(T2-6)
YTGCK(T2-2)   SELWAYETNK(T2-7)
IWCVINNESCNSECK(T2-3)
RGNYGYCYFWK(T2-4)
GNYGYCYFWK(T2-5)

Fig. S12. (A), ESI mass spectrum of native BmKAS1 and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software. (B), ESI MS spectrum of tryptic *S*-carboxymethylated BmKAS1.

Fig. S13. (A), ESI mass spectrum of native BmKabT and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software. (B), ESI MS spectrum of tryptic *S*-carboxymethylated BmKabT.

Fig. S14. (A), ESI mass spectrum of native BmTXbeta2-b and its monoisotopic mass [M+H] deconvoluted by MaxEnt3 software. (B), ESI MS spectrum of tryptic *S*-carboxymethylated BmTXbeta2-b.

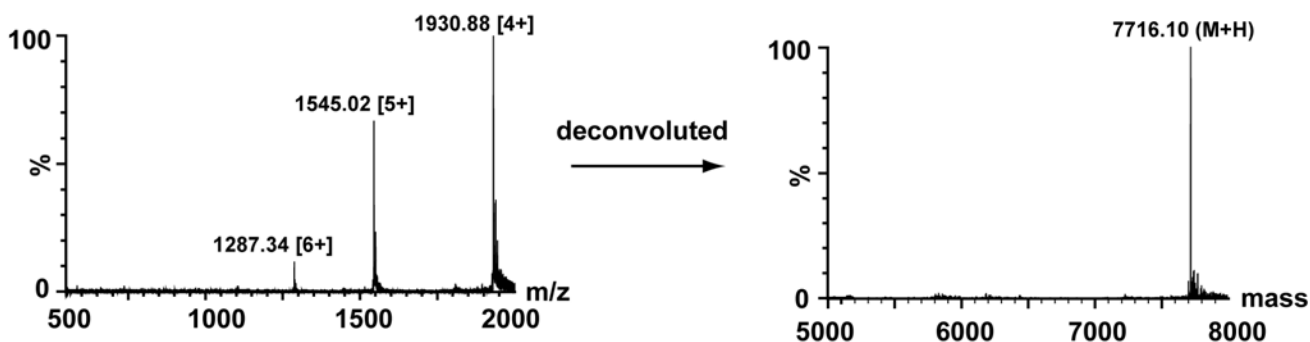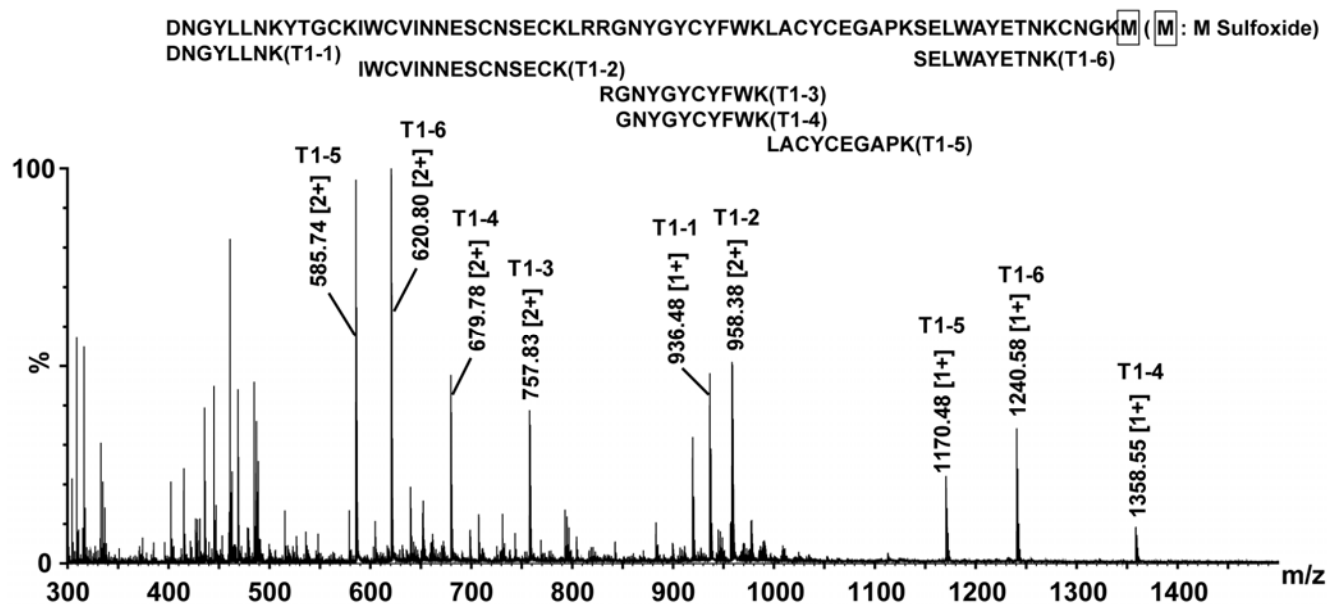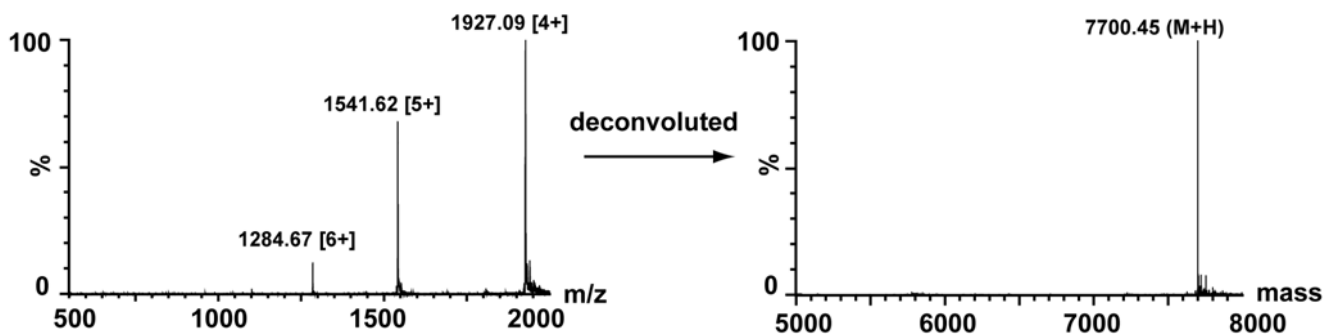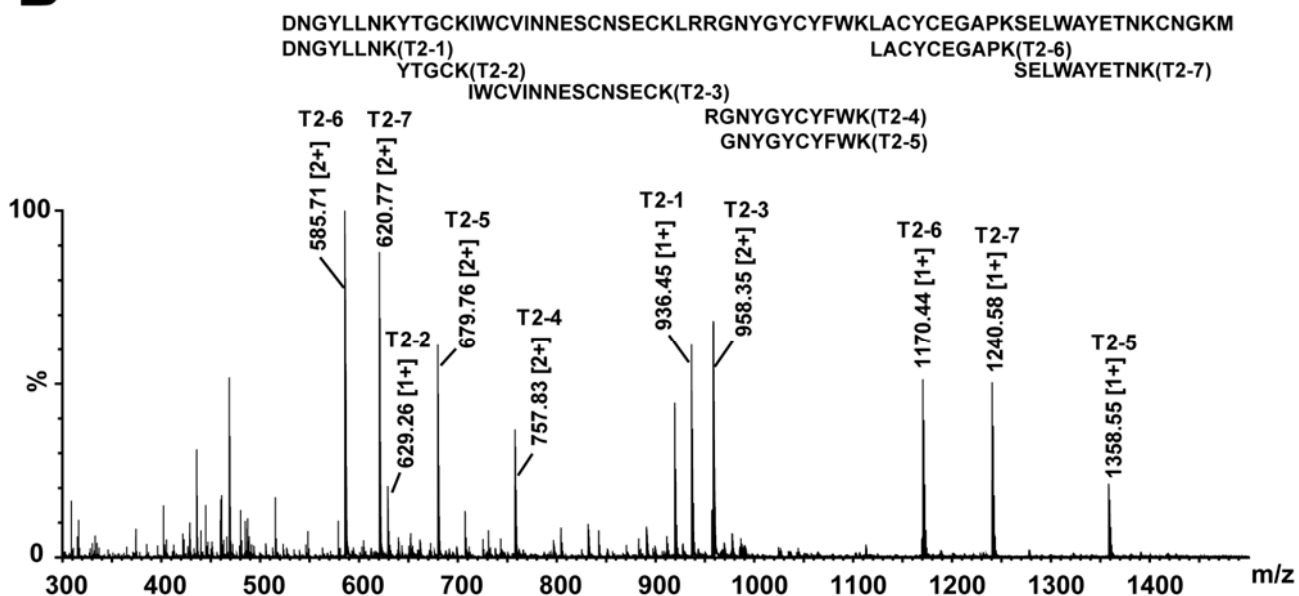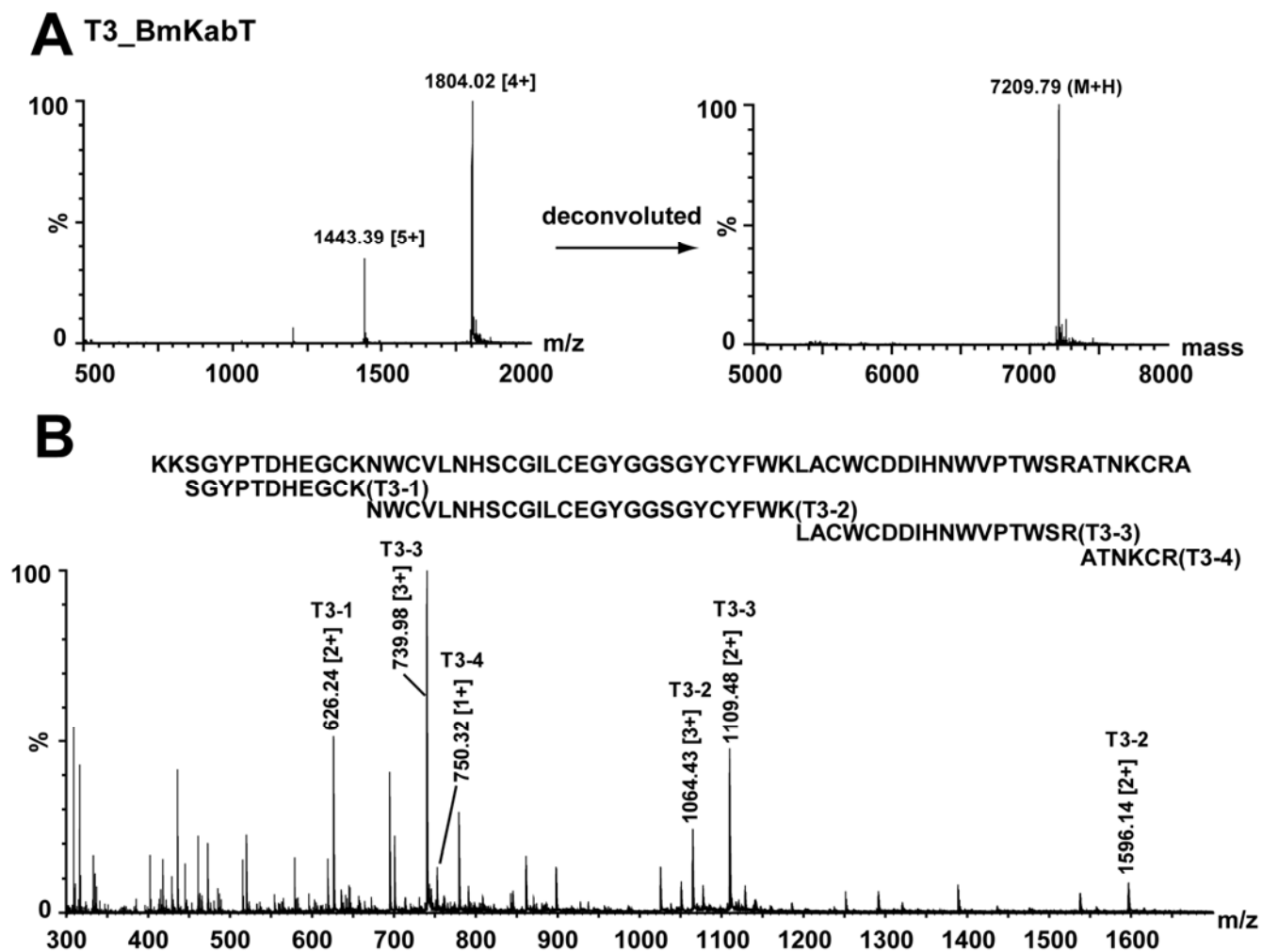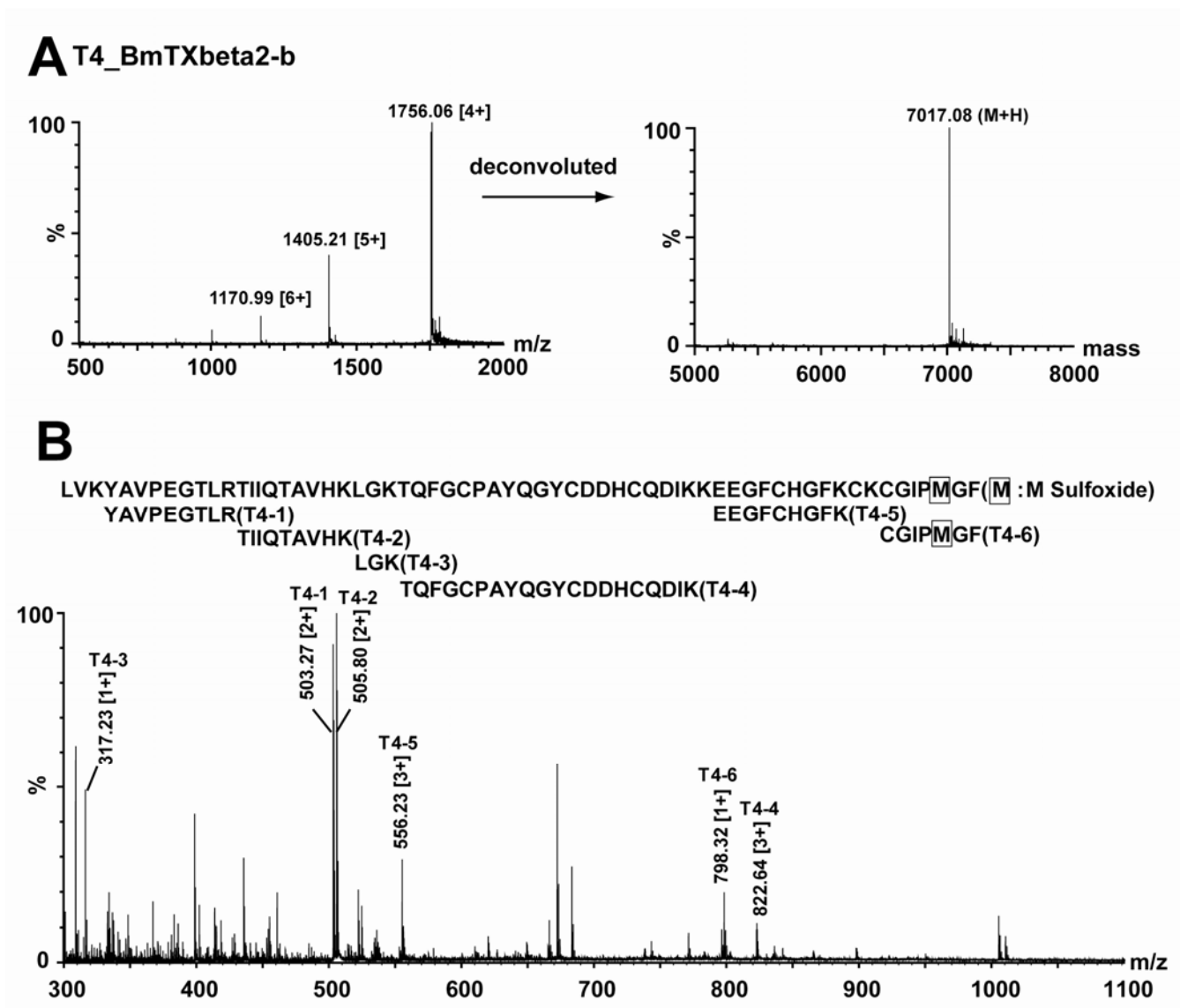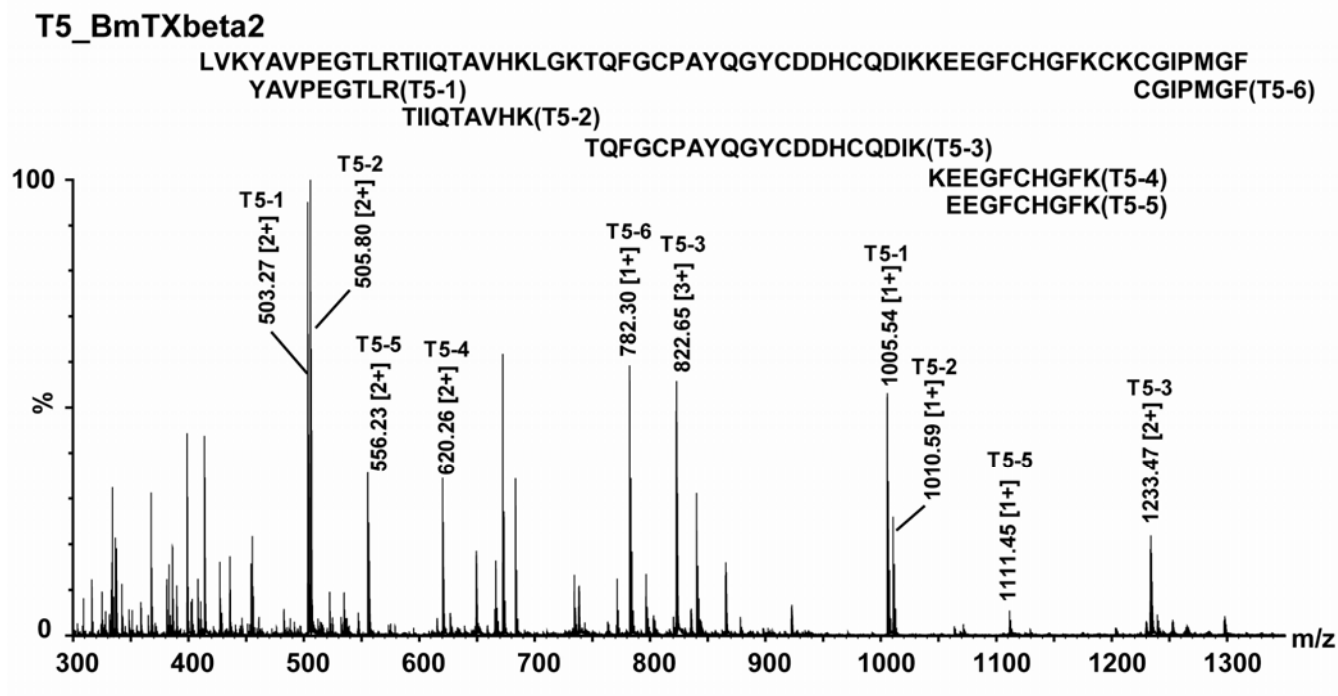Fig. S15. ESI MS spectrum of tryptic *S*-carboxymethylated BmTXbeta2.

13

Table S1. List of the purified peptides

[a] the monoisotopic molecular mass [M+H]$^{+}$ was determined by Q-TOF and the data was processed by MaxEnt3 software

[b] the theoretical monoisotopic molecular mass [M+H]$^{+}$

[c] number of disulfide bonds

[d] the observed *m/z* and charges when using NanoLC-Q-TOF and [e] the corresponding sequence

[f] Sequence of peptide, (NH$_2$) = amidation

[g] the name of peptide and [h] the protein existence level in the database

MW[M+H]$^{+}$ meas.[a](Da)	MW[M+H]$^{+}$ theor.[b](Da)	S-S[c]	*m/z* (charge)[d]	Observed trypsin fragment[e]	Full Sequence[f]	Peptide Matched[g]	Protein existence[h]
7211.06	7211.13	4	576.29(2+); 896.41(1+); 963.89(2+); 652.74(2+); 1302.51(2+); 469.29(2+);	VRDGYIADDK; DGYIADDK; DGYIADDKNCAYFCGR; NAYCDDECKK; NGAESGYCQWAGVYGNACWCYK; LPDKVPIR;	VRDGYIADDKNCAYFCGRNA YCDDECKKNGAESGYCQWAG VYGNACWCYKLPDKVPIRVP GKCNG(NH$_2$)	BmKT	Transcript-level
7016.86	7016.08	4	1117.11(3+); 1568.61(2+); 1170.46(3+); 529.22(2+); 693.65(3+); 690.32(1+);	GRDAYIADSENCTYTCALNPYCNDLCTK; DAYIADSENCTYTCALNPYCNDLCTK; DAYIADSENCTYTCALNPYCNDLCTKNGAK; SGYCQWAGK; YGNACWCIDLPDKVPIR; ISGPCR;	GRDAYIADSENCTYTCALNPY CNDLCTKNGAKSGYCQWAGK YGNACWCIDLPDKVPIRISGPC R	BmKP2	Novel-peptide

7034.02	7034.07	4	1568.60(2+);	DAYIADSENCTYTCALNPYCNDLCTK;	GRDAYIADSENCTYTCALNPY CNDLCTKNGAKSGYCQWAGR YGNACWCIDLPDKVPIRISGSC R	Makatoxin1	Protein-level
			543.21(2+);	SGYCQWAGR;			
			693.65(3+);	YGNACWCIDLPDKVPIR;			
			680.31(1+);	ISGSCR;			
6727.65	6727.84	4	623.29(1+);	DGYIR;	DGYIRGSNGCKVSCLLGNEGC NKECRAYGASYGYCWTWKL ACWCEGLPDDKTWKSESNTC G(NH$_2$)	BmKAEP	Protein-level
			1352.58(1+);	VSCLLGNEGCNK;			
			899.88(2+);	VSCLLGNEGCNKECR;			
			807.34(2+);	AYGASYGYCWTWK;			
			733.30(2+);	LACWCEGLPDDK;			
6728.71	6728.76	4	623.29(1+);	DGYIR;	DGYIRGSNGCKVSCLWGNEG CNKECRAYGASYGYCWTWG LACWCEGLPDDKTWKSESNT CG(NH$_2$)	BmKIT3	Transcript-level
			713.30(2+);	VSCLWGNEGCNK;			
			936.38(2+);	VSCLWGNEGCNKECR;			
			1495.11(2+);	AYGASYGYCWTWGLACWCEGLPDDK;			
6760.78	6760.80	4	623.31(1+);	DGYIR;	DGYIRGSNGCKISCLWGNEGC NKECKGFGAYYGYCWTWGL ACWCEGLPDDKTWKSESNTC G(NH$_2$)	BmKIT4	Transcript-level
			720.30(2+);	ISCLWGNEGCNK;			
			929.39(3+);	ISCLWGNEGCNKECK;			
			1518.06(2+);	GFGAYYGYCWTWGLACWCEGLPDDK;			
7716.10	7716.39	4	936.48(1+);	DNGYLLNK;	DNGYLLNKYTGCKIWCVINNE SCNSECKLRRGNYGYCYFWK LACYCEGAPKSELWAYETNK CNGKX (X: M Sulfoxide)	BmP09	Protein-level
			958.38(2+);	IWCVINNESCNSECK;			
			757.83(2+);	RGNYGYCYFWK;			
			679.78(2+);	GNYGYCYFWK;			
			585.74(2+);	LACYCEGAPK;			
			620.80(2+);	SELWAYETNK;			

7700.45	7700.40	4	936.45(1+); 629.26(1+); 958.35(2+); 757.83(2+); 679.76(2+); 585.71(2+); 620.77(2+);	DNGYLLNK; YTGCK; IWCVINNESCNSECK; RGNYGYCYFWK; GNYGYCYFWK; LACYCEGAPK; SELWAYETNK;	DNGYLLNKYTGCKIWCVINNE SCNSECKLRRGNYGYCYFWK LACYCEGAPKSELWAYETNK CNGKM	BmKAS-1	Protein-level
7209.79	7209.12	4	626.24(2+); 1064.43(3+); 739.98(3+); 750.32(1+);	SGYPTDHEGCK; NWCVLNHSCGILCEGYGGSGYCYFWK; LACWCDDIHNWVPTWSR; ATNKCR;	KKSGYPTDHEGCKNWCVLNH SCGILCEGYGGSGYCYFWKLA CWCDDIHNWVPTWSRATNKC RA	BmKabT	Protein-level
7017.08	7017.32	3	503.27(2+); 505.80(2+); 317.23(1+); 822.64(3+); 556.23(2+); 798.32(1+);	YAVPEGTLR; TIIQTAVHK; LGK; TQFGCPAYQGYCDDHCQDIK; EEGFCHGFK; CGIPXGF;	LVKYAVPEGTLRTIIQTAVHKL GKTQFGCPAYQGYCDDHCQDI KKEEGFCHGFKCKCGIPXGF(X : M Sulfoxide)	BmTXbeta2-b	Transcript-level
7001.02	7001.33	3	503.27(2+); 505.80(2+); 822.65(3+); 620.62(2+); 556.23(2+); 782.30(1+);	YAVPEGTLR; TIIQTAVHK; TQFGCPAYQGYCDDHCQDIK; KEEGFCHGFK; EEGFCHGFK; CGIPMGF;	LVKYAVPEGTLRTIIQTAVHKL GKTQFGCPAYQGYCDDHCQDI KKEEGFCHGFKCKCGIPMGF	BmTXbeta2	Transcript-level