Follow the Leader:

the Use of Leader Peptides to Guide Natural Product Biosynthesis

Supplementary Information

Trent J. Oman and Wilfred A. van der Donk*

Department of Chemistry, Howard Hughes Medical Institute, and Institute for Genomic Biology.
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

Telephone: 217 244 5360; e-mail: vddonk@illinois.edu

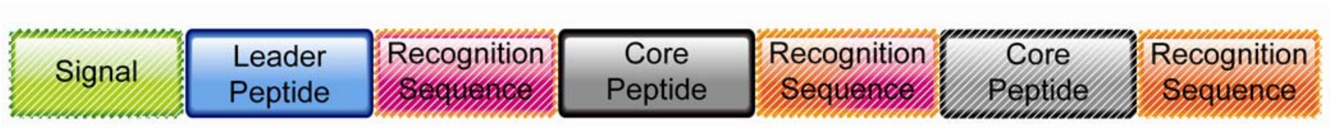
Nomenclature used in the literature for the classes of compounds discussed in relation to the nomenclature used in this review.

Historically, proteins and peptides that are generated in an inactive form and then proteolytically processed are called proproteins or propeptides. Similarly, peptides that contain a localization signal peptide (to an organelle or for secretion in unicellular organisms) are called prepeptides or preproteins. Peptides that have localization sequences and additional other sequences that do not end up in the final products are called prepropeptides. Unfortunately, in the field of posttranslationally modified peptide natural products where the modifications are not just proteolytic processes, the nomenclature used has been non-uniform between different classes of compounds and even within different classes of compounds. For instance, the term prepropeptide has also been used for the entire precursor including the segment that is transformed to the final structure (e.g. [1,2]). Similarly, the term propeptide has been used for the segment that does not end up in the final product (e.g. for microcins[3,4] and conopeptides[5]), or for the segment that does end up in the final product (e.g. for lantibiotics, mycotoxins, and cyclic bacteriocins [1,2,6]). Similarly, the name structural peptide has been used for the full precursor peptide[1] (as the translation product of the structural gene) as well as for the segment that ultimately forms the final product (e.g. [7-9]).

Because of this non-consistent use of terms and because it is not known for many of the natural products discussed in this review what the exact roles are for the segments that do not end up in the final product, we propose a standard uniform nomenclature and have used this throughout the review. We call the ribosomally synthesized peptides precursor peptides (we do not use prepeptides as this term has different meanings in the different fields). Any clear localization signals that direct the peptide to a certain organelle are referred to as signal peptides. The peptide sequence that is transformed into the final product is called core peptide (we do not use the terms propeptide or structural peptide because, again, these terms have different meaning in different fields), with the segment located N-terminal to the first core peptide termed leader peptide. For cases where the core peptide is flanked by conserved regions at both N- and C-terminus, we call these conserved sequences recognition sequences. The resulting nomenclature is shown in **Fig. 1** of the main text and also in **Supplementary Figure 1**. How these terms used in this review relate to the nomenclature used in each group of natural products discussed is also shown in **Supplementary Figure 1.**

Supplementary Figure 1. Nomenclature for various segments of the precursor peptides used in this review in relation to nomenclature in the literature.

Product Class	Terms used in literature	Terms used in this review
Lantibiotics[6]	prepeptide	precursor peptide
	leader peptide	leader peptide
	structural peptide	core peptide
	propeptide	core peptide
Microcins[3]	promicrocin[3]	precursor peptide
	premicrocin[10]	precursor peptide
	propeptide	leader peptide
Cytolysins[11]	precursor peptide	precursor peptide
	leader peptide	leader peptide
	propeptide	core peptide
Cyanobactins	leader peptide	leader peptide
	structural peptides	core peptide
	cassettes	core peptide
	recognition sequences	recognition sequences
Thiopeptides[1,12]	structural peptide[1]	precursor peptide
	leader peptide	leader peptide
	structural peptide[12]	core peptide
	core peptide	core peptide
Conopeptides[5]	N-terminal signal sequence	signal peptide
	propeptide	leader peptide
	toxin region	core peptide
	post-peptide	recognition sequence
Amatoxins and Phallotoxins		
	propeptide	core peptide
Cyclotides[13]	propeptide/prosequence	leader peptide
	NTR (N-terminal repeats)	recognition sequences
Microviridins[14,15]	prepeptide	precursor peptide
	leader peptide	leader peptide
	core region	core peptide
Cyclic bacteriocins[2]	propeptide	core peptide
	prepropeptide	precursor peptide
	prepeptide	core peptide



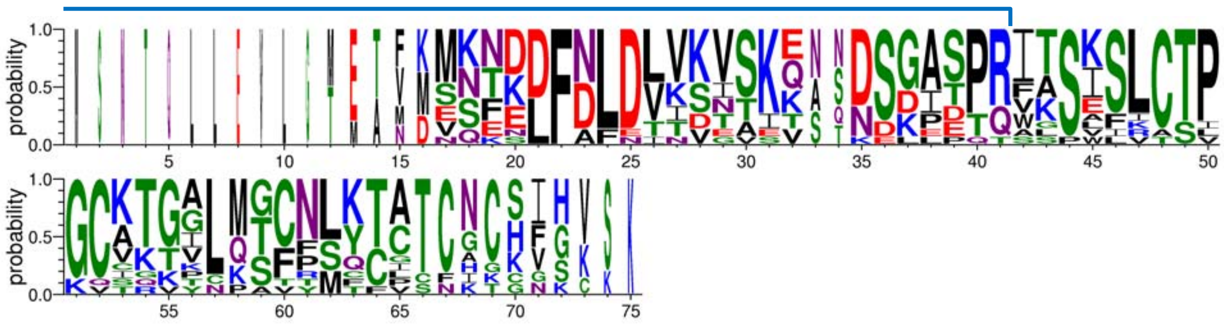
Peptide segments in striped boxes are not present in all examples of precursor peptides discussed in this review.

Sequence data for **Supplementary Figures 2-12** was obtained from the NCBI Database (http://www.ncbi.nlm.nih.gov/). The name of the precursor peptide and accession number for each sequence is indicated to the left of the alignment. Alignments were generated using AlignX software (Vector NTI, Invitrogen). Sequence Logos[16] were created using WebLogo software (internet version 3)[17].

Importantly, these alignments are not meant to be comprehensive, but to reveal commonalities, differences, and general topology of the precursor peptides.

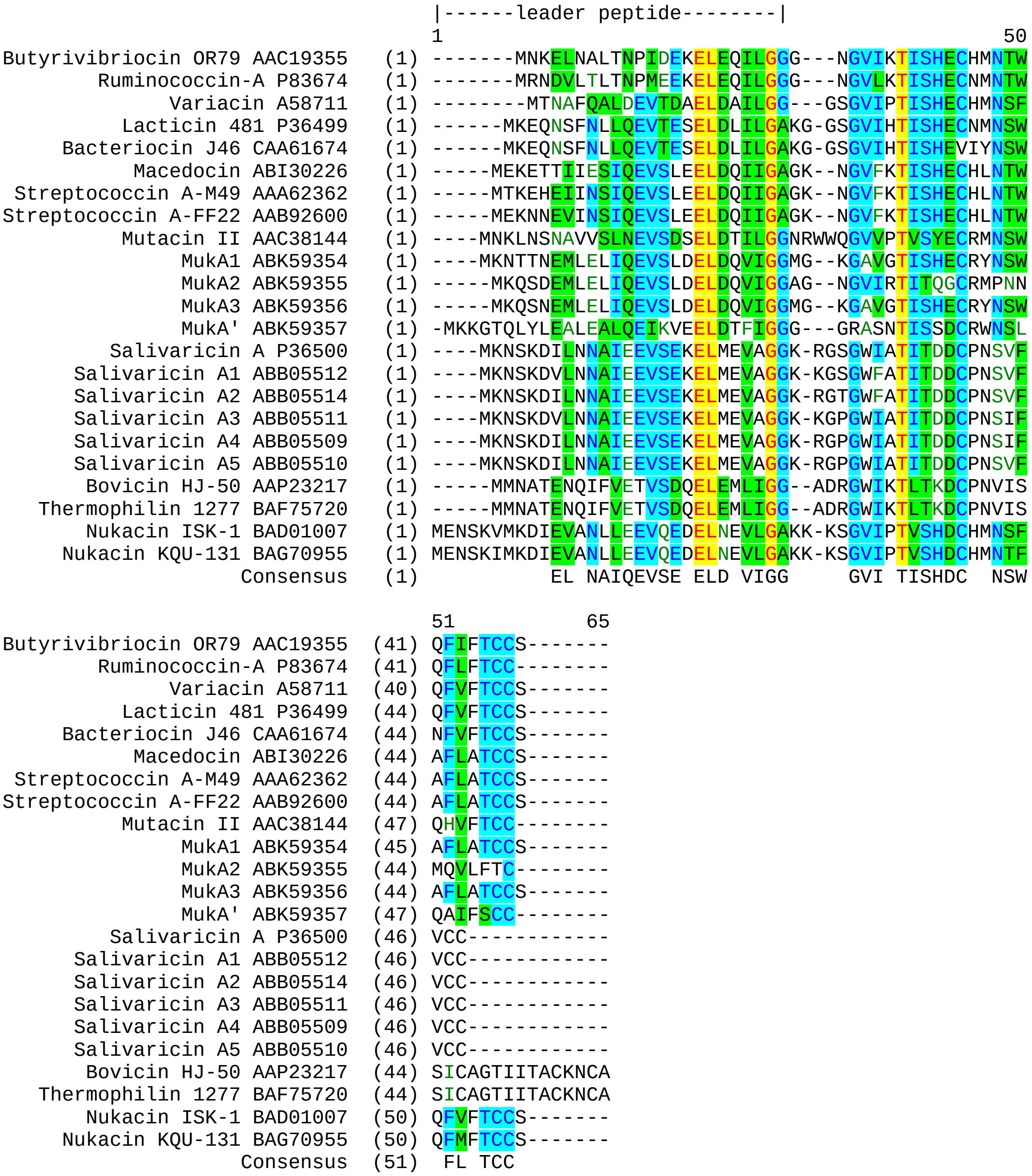
Supplementary Figure 2. Sequence alignments of the precursor peptides for selected class I lantibiotics. The logo below the alignments shows the probability of each amino acid (height of the letter) and is scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters). The blue line above sequence logo indicates leader peptide. An FXLD motif is found at positions 22-25, and a Pro is often found in the P2 position.

```
                               |------------leader peptide-------------|
                               1                                                50
            Nisin A P13068  (1) ----------------MSTKDFNLDLVSVSKK--DSGASPRITSISLCTP
            Nisin Z P29559  (1) ----------------MSTKDFNLDLVSVSKK--DSGASPRITSISLCTP
          Nisin Q BAD05046  (1) ----------------MSTKDFNLDLVSVSKT--DSGASTRITSISLCTP
          Nisin F ABU45463  (1) ----------------MSTKDFNLDLVSVSKK--DSGASPRITSISLCTP
            Nisin U Q2QBT0  (1) ----------------MNNEDFNLDLIKISKEN-NSGASPRITSKSLCTP
                  Nisin U2  (1) ----------------MNNEDFNLDLIKISKEN-NSGASPRVTSKSLCTP
   epilancin K7 AAA79236  (1) ----------------MNNSLFDLNLNKGVET-QKSDLSPQSASVLKTSI
         Subtilin P10946  (1) ---------------MSKFDDFDLDVVKVSKQ--DSKITPQWKSESLCTP
            Pep5 CAA90023  (1) --------------MKNNKNLFDLEIKKETSQNTDELEPQTAGPAIRASV
        Ericin S AAL15569  (1) ---------------MSKFDDFDLDVVKVSKQ--DSKITPQWKSESVCTP
        Ericin A AAL15567  (1) ------------MTNMSKFDDFDLDVVKVSKQ--DSKITPQVLSKSLCTP
       Epidermin P08136  (1) -----------MEAVKEKNDLFNLDVKVNAKESNDSGAEPRIASKFICTP
     Gallidermin P21838  (1) -----------MEAVKEKNELFDLDVKVNAKESNDSGAEPRIASKFLCTP
Mutacin 1140 (III) O68586  (1) MSNTQLLEVLGTETFDVQEDLFAFDTTDTTIVASNDDPDTRFKSWSLCTP
       Mutacin I AAG48564  (1) MSNTQLLEVLGTETFDVQEDLFAFDTTDTTIVASNDDPDTRFSSLSLCSL
                Consensus  (1)                 M  DDF LDLV VSK   DSGASPRI S SLCTP

                               51                     75
            Nisin A P13068 (33) GCKTGALMGCNMKTATCHCSIHVSK
            Nisin Z P29559 (33) GCKTGALMGCNMKTATCNCSIHVSK
          Nisin Q BAD05046 (33) GCKTGVLMGCNLKTATCNCSVHVSK
          Nisin F ABU45463 (33) GCKTGALMGCNMKTATCNCSVHVSK
            Nisin U Q2QBT0 (34) GCKTGILMTCPLKTATCGCHFG---
                  Nisin U2 (34) GCKTGILMTCPLKTATCGCHFG---
   epilancin K7 AAA79236 (34) KVSKKYCKGVTLTCGCNITGGK---
         Subtilin P10946 (34) GCVTGALQTCFLQTLTCNCKISK--
            Pep5 CAA90023 (37) KQCQKTLKATRLFTVSCKGKNGCK-
        Ericin S AAL15569 (34) GCVTGVLQTCFLQTITCNCHISK--
        Ericin A AAL15567 (37) GCITGPLQTCYLCFPTFAKC-----
       Epidermin P08136 (40) GCAKTG--SFNSYCC----------
     Gallidermin P21838 (40) GCAKTG--SFNSYCC----------
Mutacin 1140 (III) O68586 (51) GCARTG--SFNSYCC----------
       Mutacin I AAG48564 (51) GCTGVKNPSFNSYCC----------
                Consensus (51) GC TG L SCNL T TC C
```

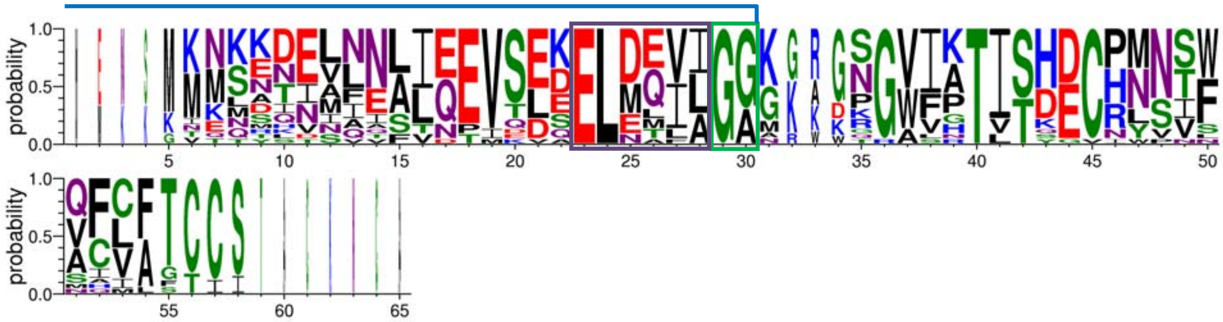


Supplementary Figure 3. Sequence alignment of the precursor peptides of selected class II lantibiotics. The logos show the probability of each amino acid (height of the letter) and are scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters). Blue lines above sequence logos indicate the leader peptide. The ELXXBXG motif is boxed in grey and the double Gly motifs are boxed in green (Note the first Gly in the double Gly motif is the last Gly in the ELXXBXG motif). In non-lantibiotic bacteriocins, this motif is E(L/B)XXBXG (see **Supplementary Fig. 4**). The motif is not found in two-peptide lantibiotics.

a. Single peptide lantibiotics

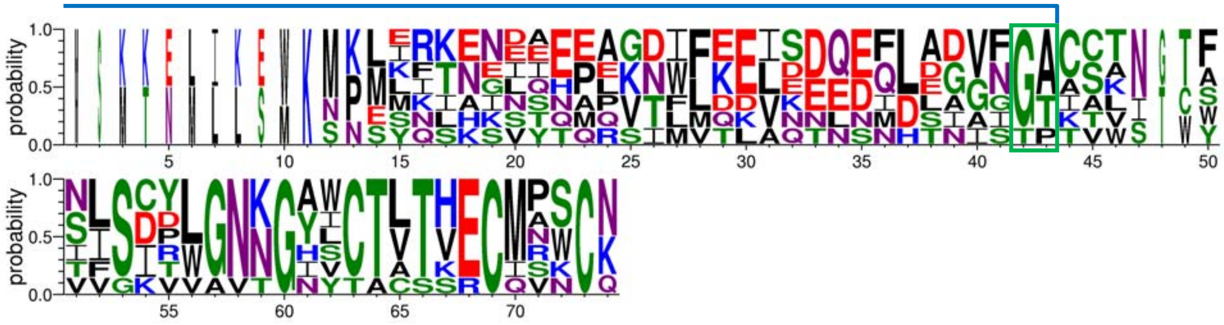


```
                                      |------leader peptide--------|
                                      1                                                50
Butyrivibriocin OR79 AAC19355   (1) -------MNKELNALTNPIDEKELEQILGGG---NGVIKTISHECHMNTW
        Ruminococcin-A P83674   (1) -------MRNDVLTLTNPMEEKELEQILGGG---NGVLKTISHECNMNTW
              Variacin A58711   (1) --------MTNAFQALDEVTDAELDAILGG---GSGVIPTISHECHMNSF
          Lacticin 481 P36499   (1) ------MKEQNSFNLLQEVTESELDLILGAKG-GSGVIHTISHECNMNSW
     Bacteriocin J46 CAA61674   (1) ------MKEQNSFNLLQEVTESELDLILGAKG-GSGVIHTISHEVIYNSW
           Macedocin ABI30226   (1) -----MEKETTIIESIQEVSLEELDQIIGAGK--NGVFKTISHECHLNTW
 Streptococcin A-M49 AAA62362   (1) -----MTKEHEIINSIQEVSLEELDQIIGAGK--NGVFKTISHECHLNTW
Streptococcin A-FF22 AAB92600   (1) -----MEKNNEVINSIQEVSLEELDQIIGAGK--NGVFKTISHECHLNTW
          Mutacin II AAC38144   (1) ----MNKLNSNAVVSLNEVSDSELDTILGGNRWWQGVVPTVSYECRMNSW
               MukA1 ABK59354   (1) ----MKNTTNEMLELIQEVSLDELDQVIGGMG--KGAVGTISHECRYNSW
               MukA2 ABK59355   (1) -----MKQSDEMLELIQEVSLDELDQVIGGAG--NGVIRTITQGCRMPNN
               MukA3 ABK59356   (1) -----MKQSNEMLELIQEVSLDELDQVIGGMG--KGAVGTISHECRYNSW
               MukA' ABK59357   (1) -MKKGTQLYLEALEALQEIKVEELDTFIGGG---GRASNTISSDCRWNSL
        Salivaricin A P36500   (1) ----MKNSKDILNNAIEEVSEKELMEVAGGK-RGSGWIATITDDCPNSVF
     Salivaricin A1 ABB05512   (1) ----MKNSKDVLNNAIEEVSEKELMEVAGGK-KGSGWFATITDDCPNSVF
     Salivaricin A2 ABB05514   (1) ----MKNSKDILNNAIEEVSEKELMEVAGGK-RGTGWFATITDDCPNSVF
     Salivaricin A3 ABB05511   (1) ----MKNSKDVLNNAIEEVSEKELMEVAGGK-KGPGWIATITDDCPNSIF
     Salivaricin A4 ABB05509   (1) ----MKNSKDILNNAIEEVSEKELMEVAGGK-RGPGWIATITDDCPNSIF
     Salivaricin A5 ABB05510   (1) ----MKNSKDILNNAIEEVSEKELMEVAGGK-RGPGWIATITDDCPNSVF
       Bovicin HJ-50 AAP23217   (1) -----MMNATENQIFVETVSDQELEMLIGG--ADRGWIKTLTKDCPNVIS
   Thermophilin 1277 BAF75720   (1) -----MMNATENQIFVETVSDQELEMLIGG--ADRGWIKTLTKDCPNVIS
       Nukacin ISK-1 BAD01007   (1) MENSKVMKDIEVANLLEEVQEDELNEVLGAKK-KSGVIPTVSHDCHMNSF
     Nukacin KQU-131 BAG70955   (1) MENSKIMKDIEVANLLEEVQEDELNEVLGAKK-KSGVIPTVSHDCHMNTF
                    Consensus   (1)           EL NAIQEVSE ELD VIGG     GVI TISHDC  NSW

                                      51           65
Butyrivibriocin OR79 AAC19355  (41) QFIFTCCS-------
        Ruminococcin-A P83674  (41) QFLFTCC--------
              Variacin A58711  (40) QFVFTCCS-------
          Lacticin 481 P36499  (44) QFVFTCCS-------
     Bacteriocin J46 CAA61674  (44) NFVFTCCS-------
           Macedocin ABI30226  (44) AFLATCCS-------
 Streptococcin A-M49 AAA62362  (44) AFLATCCS-------
Streptococcin A-FF22 AAB92600  (44) AFLATCCS-------
          Mutacin II AAC38144  (47) QHVFTCC--------
               MukA1 ABK59354  (45) AFLATCCS-------
               MukA2 ABK59355  (44) MQVLFTC--------
               MukA3 ABK59356  (44) AFLATCCS-------
               MukA' ABK59357  (47) QAIFSCC--------
        Salivaricin A P36500  (46) VCC------------
     Salivaricin A1 ABB05512  (46) VCC------------
     Salivaricin A2 ABB05514  (46) VCC------------
     Salivaricin A3 ABB05511  (46) VCC------------
     Salivaricin A4 ABB05509  (46) VCC------------
     Salivaricin A5 ABB05510  (46) VCC------------
       Bovicin HJ-50 AAP23217  (44) SICAGTIITACKNCA
   Thermophilin 1277 BAF75720  (44) SICAGTIITACKNCA
       Nukacin ISK-1 BAD01007  (50) QFVFTCCS-------
     Nukacin KQU-131 BAG70955  (50) QFMFTCCS-------
                    Consensus  (51)  FL TCC
```



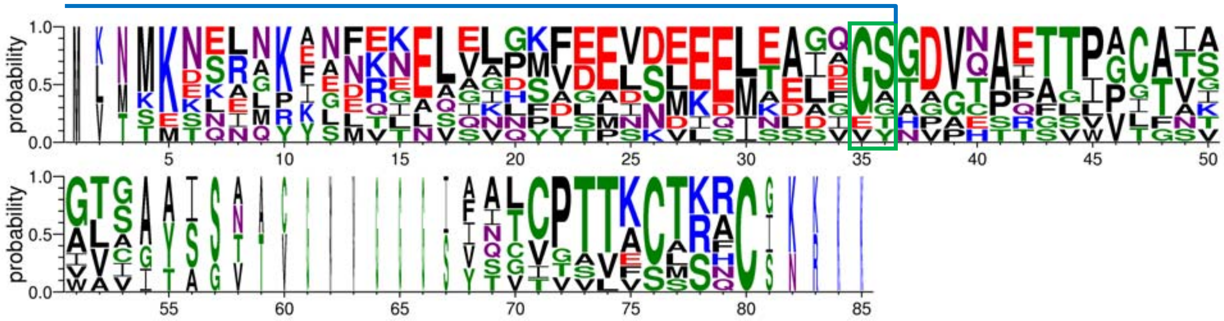
b. Two-peptide lantibiotics: α peptides

```
                                        |-------------leader peptide-------------|
                                        1                                               50
       Lacticin 3147 alpha O87236   (1) --------------MNKNEIETQPVTWLEEVSDQNFDEDVFGACSTN-TF
     Plantaracin W alpha AAG02567   (1) -------------MKISKIEAQARKDFFKKIDTNSNLLNVNGAKCK--WW
               BHT alpha AAZ76603   (1) -----------MKEIQKAGLQEELSILMDDANNLEQLTAGIGTTVVN--S
Staphylococcin C55 alpha BAB78438   (1) -----------MKSSFLEKDIEEQVTWFEEVSEQEFDDDIFGACSTN-TF
               Smb alpha BAD72776   (1) ---------MKSNLLKINNVTEMEKNMVTLIKDEDMLAGGSTPACAIG--
               Hal alpha BAB04173   (1) --MTNLLKEWKMPLERTHNNSNPAGDIFQELEDQDILAGVNGACAW---Y
     Lichenicidin alpha YP_093633   (1) MSKKEMILSWKNPMYRTESSYHPAGNILKELQEEEQHSIAGGTITLSTCA
                        Consensus   (1)              L            IL EI DQE L  V GA      F

                                       51                     74
       Lacticin 3147 alpha O87236  (36) SLSDYWGNNGAWCTLTHECMAWCK
     Plantaracin W alpha AAG02567  (36) NISCDLGNNGHVCTLSHECQVSCN
               BHT alpha AAZ76603  (38) TFSIVLGNKGYICTVTVECMRNCQ
Staphylococcin C55 alpha BAB78438  (39) SLSDYWGNKGNWCTATHECMSWCK
               Smb alpha BAD72776  (40) VVGITVAVTGISTACTSRCINK--
               Hal alpha BAB04173  (46) NISCRLGNKGAYCTLTVECMPSCN
     Lichenicidin alpha YP_093633  (51) ILSKPLGNNGYLCTVTKECMPSCN
                        Consensus  (51)  LS  LGN G  CTLT ECM  CN
```



c. Two-peptide lantibiotics: β peptides

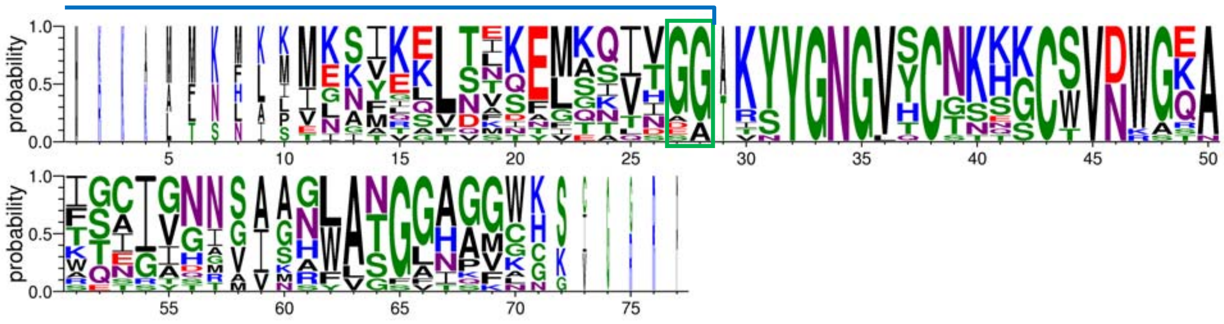
```
                                       |----------leader peptide----------|
                                       1                                               50
       Lacticin 3147 beta O87237   (1) ---MKEKNMKKNDTIELQLGKYLEDDMIELAEGDESHGGTTPATP--AIS
     Plantaricin W beta AAG02566   (1) --MTKTSR-----RKNAIANYLEPVDEKSINESFGAGDPEARSGIPCTIG
               BHT beta AAZ76602   (1) ---MKSNLLKINNVTEVEKDMVTLIKDEDMELAGGSTPACAIGVVGITVA
Staphylococcin C55 beta BAB78439   (1) ---MKNELGKFLEENELELGKFSESDMLEITDDEVYAAGTPLALLGGAAT
               Smb beta BAD72777   (1) ----MKEIQKAGLQEELSILMDDANNLEQLTAGIGTTVVNSTFSIVLGNK
               Hal beta BAB04172   (1) MVNSKDLRNP-EFRKAQGLQFVDEVNEKELSSLAGSGDVHAQTTWPCATV
               Lichenicidin beta   (1) MKTMKNSAAREAFKG--ANHPAGMVSEEELKALVGGNDVNPETTPATTSS
         Cytolysin L-AL AAK67264   (1) ------------MENLSVVPSFEELSVEEMEAIQGSGDVQAETTPVCAVA
         Cytolysin L-AS AAK67265   (1) MLNKENQENYYSNKLELVGPSFEELSLEEMEAIQGSGDVQAETTPACFTI
                       Consensus   (1)     K    K     EL L   EEV LEEL A  GSGDV A TT   A

                                       51                                85
       Lacticin 3147 beta O87237  (46) ILSAYIST----------NTCPTTKCTRAC-----
     Plantaricin W beta AAG02566  (44) AAVAAS-----------IAVCPTTKCSKRCGKRKK
               BHT beta AAZ76602  (48) VTG-----------------I-STACTSRCINK--
Staphylococcin C55 beta BAB78439  (48) GVIGYISN----------QTCPTTACTRAC-----
               Smb beta BAD72777  (47) G----------------YICTVTVECMRNCSK---
               Hal beta BAB04172  (50) GVS--------------VALCPTTKCTSQC-----
               Lichenicidin beta  (49) WTCITAGVTV------SASLCPTTKCTSRC-----
         Cytolysin L-AL AAK67264  (39) ATAAASSAACGWVGGGIFTGVTVVVSLKHC-----
         Cytolysin L-AS AAK67265  (51) GLG-----------------VGALFSAKFC-----
                       Consensus  (51) G                   CPTT CTK C
```



Supplementary Figure 4. Sequence alignments of the precursor peptides of selected bacteriocins that do not undergo posttranslational modifications. The logos below the alignments show the probability of each amino acid (height of the letter) and are scaled (width of the letter) according to how many sequences contributed to that position. Blue lines above the sequence logo indicate the leader peptide. The double Gly motif is boxed in green. While not as strongly conserved as in the class II lantibiotic precursor peptides (**Supplementary Fig. 3**), an E(B/M)XXBXG motif (with the G the first residue of the double Gly motif) is found in many leader peptides (see logo and individual peptides, e.g. curvacin for class IIa; lactacin F β for class IIb, and carnebacteriocin for class IId).

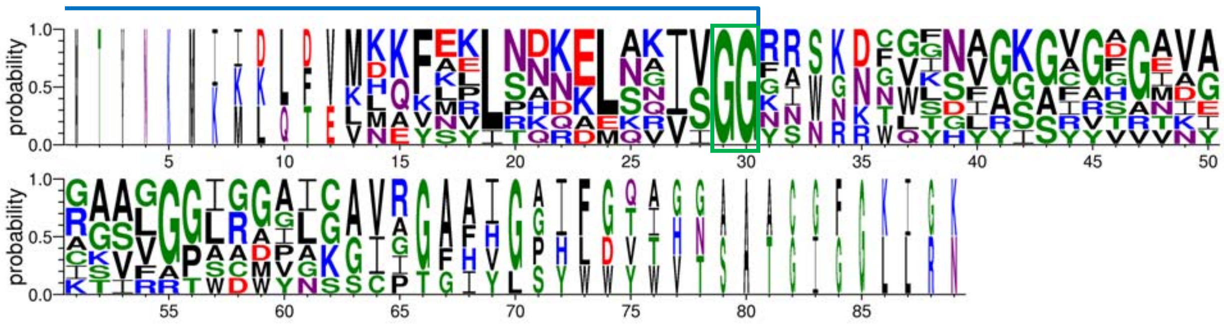
a. Class IIa bacteriocins

```
                                 |------leader peptide------|
                                 1                                                50
      mundticin KS BAB88211   (1) ---------------------MSQVVGG-KYYGNGVSCNKKGCSVDWGKA
         mundticin AAR26473   (1) -------------MKKLTAKEMSQVVGG-KYYGNGVSCNKKGCSVDWGKA
       mundticin L ACQ77507   (1) -------------MKKLTSKEMAQVVGG-KYYGNGLSCNKKGCSVDWGKA
carnobacteriocin B1 P38579   (1) ----------MKSVKELNKKEMQQINGGAISYGNGVYCNKEKCWVNKAEN
carnobacteriocin B2 P38580   (1) ----------MNSVKELNVKEMKQLHGG-VNYGNGVSCSKTKCSVNWGQA
         curvacin A P0A311   (1) ----------MNNVKELSMTELQTITGGARSYGNGVYCNNKKCWVNRGEA
         sakacin P AAM88858   (1) ----------MEKFIELSLKEVTAITGG-KYYGNGVHCGKHSCTVDWGTA
   sakacin G skgA1 AAM73713   (1) ----------MKNTRSLTIQEIKSITGG-KYYGNGVSCNSHGCSVNWGQA
   sakacin G skgA2 AAM73712   (1) ----------MKNAKSLTIQEMKSITGG-KYYGNGVSCNSHGCSVNWGQA
       pediocin PA-1 P29430   (1) ----------MKKIEKLTEKEMANIIGG-KYYGNGVTCGKHSCSVDWGKA
   plantaricin 423 AAL09346   (1) ---------MMKKIEKLTEKEMANIIGG-KYYGNGVTCGKHSCSVNWGQA
   enterocin SE-K4 EEI61656   (1) ---------MILGIVLLSVSTLGITVDAATYYGNGVYCNTQKCWVDWSRA
       enterocin A BAA92138   (1) -----MKHLKILSIKETQLIYGGTTHSG-KYYGNGVYCTKNKCTVDWAKA
          leucocin A P34034   (1) ----MMNMKPTESYEQLDNSALEQVVGG-KYYGNGVHCTKSGCSVNWGEA
mesentericin Y105 AAP37395   (1) ----MTNMKSVEAYQQLDNQNLKKVVGG-KYYGNGVHCTKSGCSVNWGEA
      divercin V41 CAA11804   (1) -----MKNLKEGSYTAVNTDELKSINGGTKYYGNGVYCNSKKCWVDWGQA
       enterocin P AAC45870   (1) MRKKLFSLALIGIFGLVVTNFGTKVDAATRSYGNGVYCNNSKCWVNWGEA
  listeriocin 743A AAK19401   (1) MKKAALKFIIVIAILGFSFSFFSIQSEA-KSYGNGVQCNKKKCWVDWGSA
                  Consensus   (1)           M  I  LT  EM  I GG KYYGNGVYCNK  CSVNWG A

                                 51                       77
      mundticin KS BAB88211  (29) IGIIGNNSAANLATGGAAGWKS-----
         mundticin AAR26473  (37) IGIIGNNSAANLATGGAAGWKS-----
       mundticin L ACQ77507  (37) IGIIGNNSAANLATGGAAGWKS-----
carnobacteriocin B1 P38579  (41) KQAITGIVIGGWASSLAGMGH------
carnobacteriocin B2 P38580  (40) FQERYTAGINSFVSGVASGAGSIGRRP
         curvacin A P0A311  (41) TQSIIGGMISGWASGLAGM--------
         sakacin P AAM88858  (40) IGNIGNNAAANWATGGNAGWNK-----
   sakacin G skgA1 AAM73713  (40) WTCGVN----HLANGGHGVC-------
   sakacin G skgA2 AAM73712  (40) WTCGVN----HLANGGHGVC-------
       pediocin PA-1 P29430  (40) TTCIINNGAMAWATGGHQGNHKC----
   plantaricin 423 AAL09346  (41) FSCSVS----HLANFGHGKC-------
   enterocin SE-K4 EEI61656  (42) RSEIVDRGVKAYVNGFTKVLGGVGGR-
       enterocin A BAA92138  (45) TTCIAGMSIGGFLGGAIPGKC------
          leucocin A P34034  (46) FSAGVH----RLANGGNGFW-------
mesentericin Y105 AAP37395  (46) ASAGIH----RLANGGNGFW-------
      divercin V41 CAA11804  (46) SGCIGQTVVGGWLGGAIPGKC------
       enterocin P AAC45870  (51) KENIAGIVISGWASGLAGMGH------
  listeriocin 743A AAK19401  (50) ISTIGNNSAANWATGGAAGWKS-----
                  Consensus  (51)  S I N     WATGG GG
```



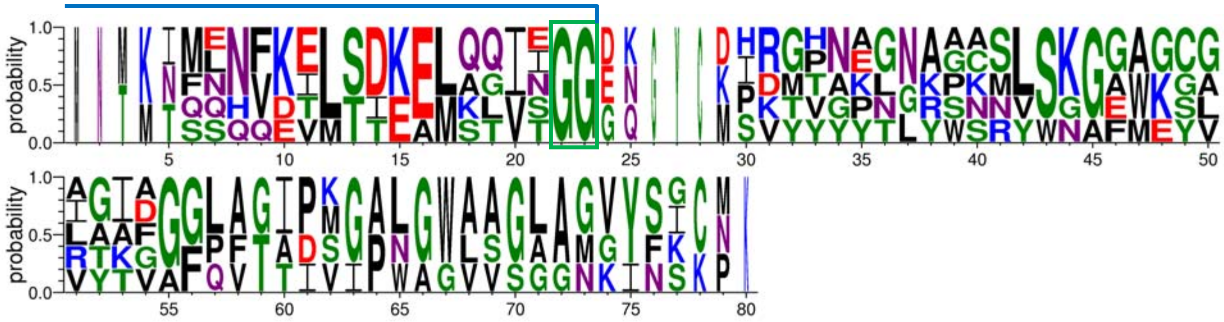
b. Class IIb bacteriocins (two-peptide)

```
                                 |-------leader peptide-------|
                                 1                                               50
Lactacin F alpha NP_964622   (1) ------------MKQFNYLSHKDLAVVVGGR--NNWQTNVGGAVGSAMIG
 Lactacin F beta NP_964623   (1) ----------------MKLNDKELSKIVGGNRWGDTVLSAASGAGTGIKA
    brochocin C A AAC95138   (1) ------------MHKVKKLNNQELQQIVGGYSSKDCLKDIGKGIGAGTVA
    brochocin C B AAC95139   (1) -------------MKKELLNKNEMSRIIGGKIN---WGNVGGSCVGGAVI
    plantaricin E CAA64188   (1) -------MLQFEKLQYSRLPQKKLAKISGGFN--RGGYNFGKSVRHVVDA
    plantaricin F CAA64187   (1) ------------MKKFLVLRDRELNAISGG------VFHAYSARGVRNNY
    plantaricin J CAA64198   (1) MTVNKMIKDLDVVDAFAPISNNKLNGVVGGGAWKNFWSSLRKGFYDGEAG
    plantaricin K CAA64197   (1) -----MKIKLTVLNEFEELTADAEKNISGGRRSRKNGIGYAIGYAFGAVE
                 Consensus   (1)             M  F  L  KELA IVGG         VG G G G VA

                                 51                                   89
Lactacin F alpha NP_964622  (37) ATVGGTICGPACAVAGAHYLPILWTAVTAATGGFGKIRK
 Lactacin F beta NP_964623  (35) CKSFGPWGMAICGVGGAAIGGYFGYTHN-----------
    brochocin C A AAC95138  (39) GAAGGGLAAGLGAIPGAFVGAHFGVIGGSAACIGGLLGN
    brochocin C B AAC95139  (35) GGALGGLGGAGGGCITGAIGSIWDQW-------------
    plantaricin E CAA64188  (42) IGSVAGIRGILKSIR------------------------
    plantaricin F CAA64187  (33) KSAVGPADWVISAVRGFIHG-------------------
    plantaricin J CAA64198  (51) RAIRR----------------------------------
    plantaricin K CAA64197  (46) RAVLGGSRDYNK---------------------------
                 Consensus  (51)  AALGGI G I AV GA  G
```



c. Class IId bacteriocins

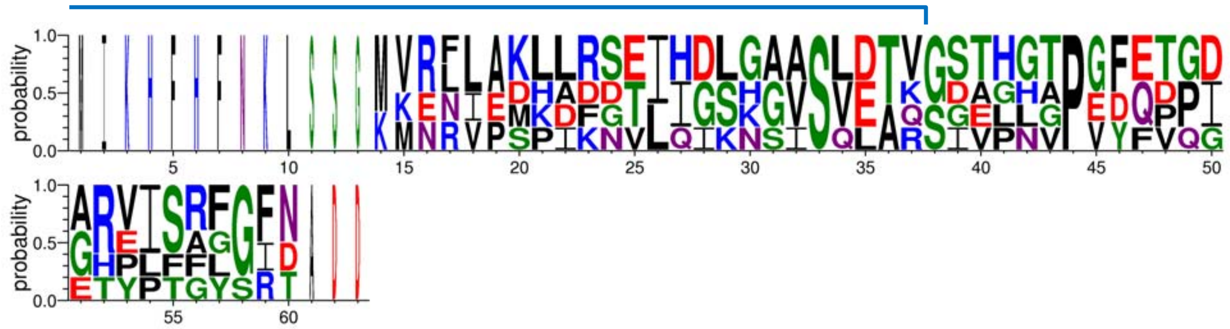
```
                                |---leader peptide----|
                                1                                               50
carnobacteriocin A P38578   (1) -----MNNVKELSIKEMQQVTGGDQ---MSDGVNYG-KGSSLSKGGAKCG
      enterocin B AAD28234   (1) -----MQNVKELSTKEMKQIIGGEN---DHRMPNELNRPNNLSKGGAKCG
     acidocin 8912 Q48501   (1) ---MISSHQKTLTDKELALISGG-------KTHYPTNAWKSLWKGFWESL
    lactococcin A1 D49786   (1) --MKNQLNFEILSDEELQGINGG------IRGTGKGLAAAMVSGAAMGGA
     bovicin 255 AAG29818   (1) MNTKTFEQFDVMTDEALSTVEGGGKGYCKPVYYAANGYSCRYSNGEWGYV
                Consensus   (1)        N K LSDKEL  I GG       R          LSKGG   G

                                51                          80
carnobacteriocin A P38578  (42) LGIVGGLATIPSGPLGWLAGAAGVINSCMK
      enterocin B AAD28234  (43) AAIAGGLFGIPKGPLAWAAGLANVYSKCN-
     acidocin 8912 Q48501  (41) RYTDGF------------------------
    lactococcin A1 D49786  (43) IGAFGGPVGAIMGAWGGAVGGAMKYSI---
     bovicin 255 AAG29818  (51) VTKGAFQATTDVIANGWVSSLGGGYFGKP-
                Consensus  (51) IG  GG      G  GW AG A  Y
```



Supplementary Figure 5a. Sequence alignment of putative precursor peptides of lasso peptide microcins. The logo below the alignments shows the probability of each amino acid (height of the letter) and is scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters). Blue line above the sequence logo indicates leader peptide. For BURTA, BURPS/BURMA, SPHAL, see [18].

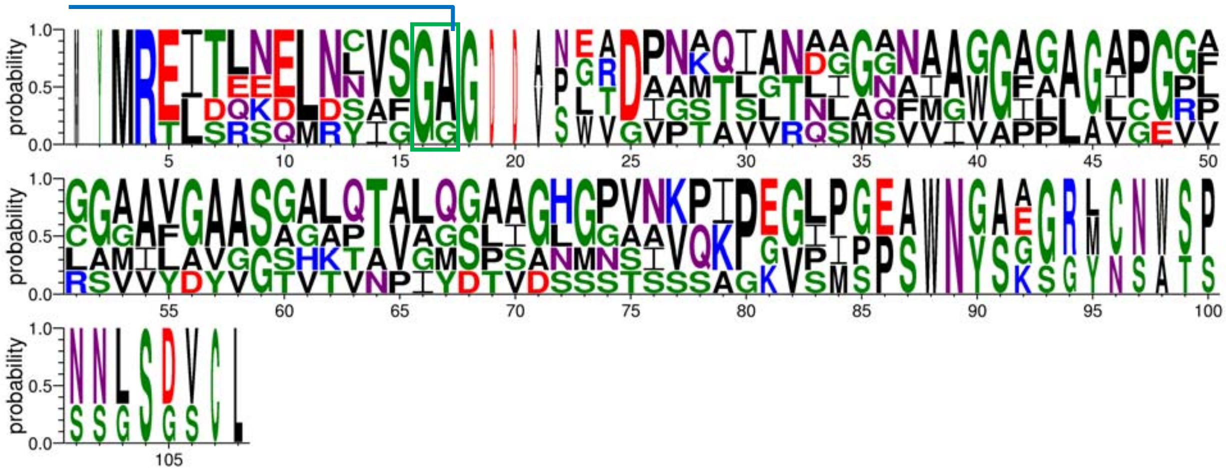
Lasso peptide microcins

```
                               |------leader peptide--(McjA)-------|
                               1                                               50
McjA Microcin J25 Q9X2V7   (1) MIKHFHFNKLSSGKKNNVPSPAKGVIQIKKSASQLTKGGAGHVPEYFVGI
                   BURTA   (1) -------------MVRLLAKLLRSTIHGSNGVSLDAVSSTHGTPGFQTPD
             BURPS/BURMA   (1) -------------MVRFLAKLLRSTIHGSHGVSLDAVSSTHGTPGFQTPD
            CAULO Q0LUL4   (1) --------------MERIEDHIDDELIDLGAASVETQGDVLNAP--EPGI
                   SPHAL   (1) -------------------MKDFNELIDLGAISVETRGIEPLGPVDEDQG
               Consensus   (1)               V  L   LR  I    AVSLDT G     P F

                               51         63
McjA Microcin J25 Q9X2V7  (51) GTPISFYG-----
                   BURTA  (38) ARVISRFGFN---
             BURPS/BURMA  (38) ARVISRFGFN---
            CAULO Q0LUL4  (35) GREPTGLSRD---
                   SPHAL  (32) EHYLFAGGITADD
               Consensus  (51) AR IS FG
```

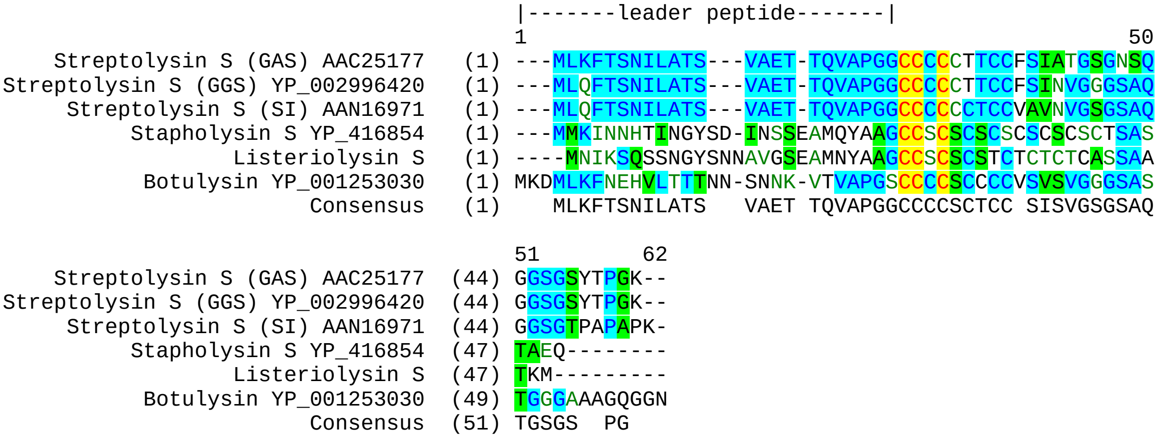


Supplementary Figure 5b. Sequence alignment of the short precursor peptides of selected class II microcins. The logo below the alignments shows the probability of each amino acid (height of the letter) and is scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters). Blue line above the sequence logo indicates leader peptide. The double Gly motif is boxed in green.

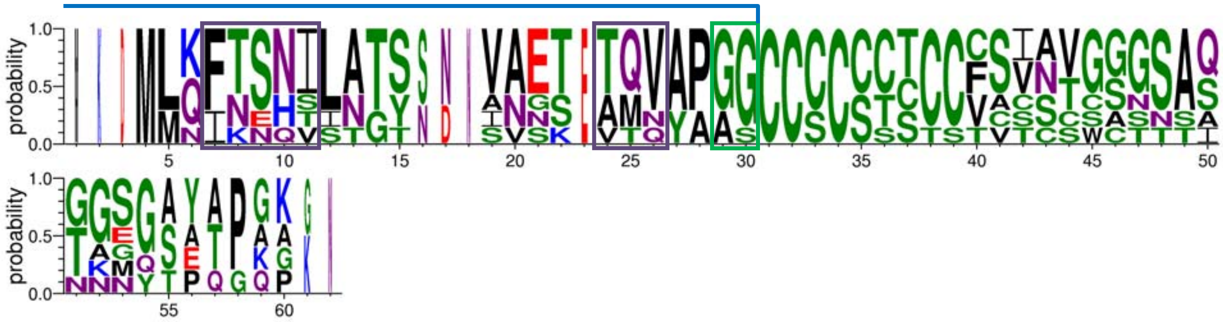
```
                             |leader peptide-|
                             1                                                50
 Microcin H47 P62530     (1) --MREITESQLRYISGAG------GAPATSANAAGAAAIVGALAGIPGGP
Microcin E492 Q9Z4N4     (1) --MREISQKDLNLAFGAG----ETDPNTQLLNDLGNNMAWGAALGAPGGL
  Microcin L AAP03989    (1) --MREITLNEMNNVSGAG-DVNWVDVGKTVATNGAGVIGGAFGAGLCGPV
  Microcin V CAQ87179    (1) --MRTLTLNELDSVSGG---ASGRDIAMAIGTLSGQFVAGGIGAAAGGVA
Microcin 24 AAA88772     (1) MYMRELDREELNCVGGAGD--PLADPNSQIVRQIMSNAAWGPPL-VPERF
            Consensus    (1)   MREIT  ELN VSGAG      D    IA   G  IA G GAGIPG

                             51                                              100
 Microcin H47 P62530    (43) LGVVVGAVSAGLTTAIGSTVGSGSASSSAGGGS-----------------
Microcin E492 Q9Z4N4    (45) GSAALGAAGGALQTVGQGLIDHGPVNVPIP-VLIGPSWNGSGSGYNSATS
  Microcin L AAP03989   (48) CAGAFAVGSSAAVAALYDAAGNSNSAKQKPEGLPPEAWNYAEGRMCNWSP
  Microcin V CAQ87179   (46) GGAIYDYASTHKPNPAMSPSGLGGTIKQKPEGIPSEAWNYAAGRLCNWSP
Microcin 24 AAA88772    (48) RGMAVGAAGGVTQTVLQGAAAHMPVNVPIPKVPMGPSWNGSKG-------
            Consensus   (51)  GAAVGAASAA  T L   AG G      P GL   AWN A G     S

                             101
 Microcin H47 P62530    (76) --------
Microcin E492 Q9Z4N4    (94) SSGSGS--
  Microcin L AAP03989   (98) NNLSDVCL
  Microcin V CAQ87179   (96) NNLSDVCL
Microcin 24 AAA88772    (91) --------
            Consensus  (101)    S
```



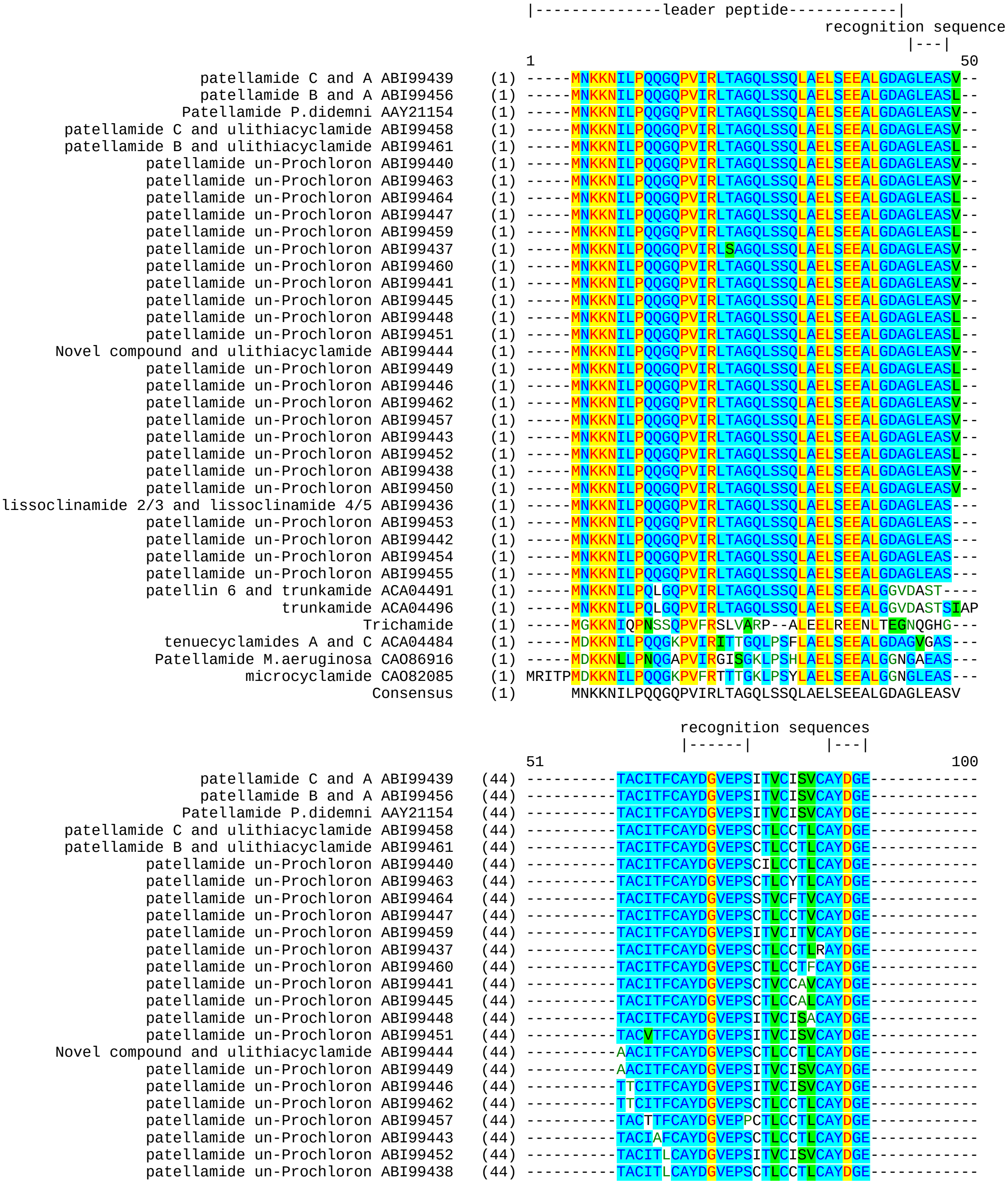
Supplementary Figure 6. Sequence alignment of the precursor peptides of selected cytolysins. The logo below the alignments shows the probability of each amino acid (height of the letter) and is scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters). Blue line above sequence logo indicates the leader peptide. The FXXXB and TQV motifs are boxed in grey, and the double Gly motif is boxed in green.



```
                                        |-------leader peptide-------|
                                        1                                               50
    Streptolysin S (GAS) AAC25177   (1) ---MLKFTSNILATS---VAET-TQVAPGGCCCCCTTCCFSIATGSGNSQ
Streptolysin S (GGS) YP_002996420   (1) ---MLQFTSNILATS---VAET-TQVAPGGCCCCCTTCCFSINVGGGSAQ
     Streptolysin S (SI) AAN16971   (1) ---MLQFTSNILATS---VAET-TQVAPGGCCCCCCTCCVAVNVGSGSAQ
          Stapholysin S YP_416854   (1) ---MMKINNHTINGYSD-INSSEAMQYAAGCCSCSCSCSCSCSCTSAS
                  Listeriolysin S   (1) ----MNIKSQSSNGYSNNAVGSEAMNYAAGCCSCSCSTCTCTCTCASSAA
           Botulysin YP_001253030   (1) MKDMLKFNEHVLTTTNN-SNNK-VTVAPGSCCCCSCCCCVSVSVGGGSAS
                        Consensus   (1)    MLKFTSNILATS   VAET TQVAPGGCCCCSCTCC SISVGSGSAQ

                                        51        62
    Streptolysin S (GAS) AAC25177  (44) GGSGSYTPGK--
Streptolysin S (GGS) YP_002996420  (44) GGSGSYTPGK--
     Streptolysin S (SI) AAN16971  (44) GGSGTPAPAPK-
          Stapholysin S YP_416854  (47) TAEQ--------
                  Listeriolysin S  (47) TKM---------
           Botulysin YP_001253030  (49) TGGGAAAGQGGN
                        Consensus  (51) TGSGS  PG
```



Supplementary Figure 7. Sequence alignment of selected cyanobactin precursor peptides. The logo below the alignments shows the probability of each amino acid (height of the letter) and is scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters). Blue line above sequence logo indicates leader peptide. Red line indicates recognition sequences. The hypervariable core peptides are located between the recognition sequences.

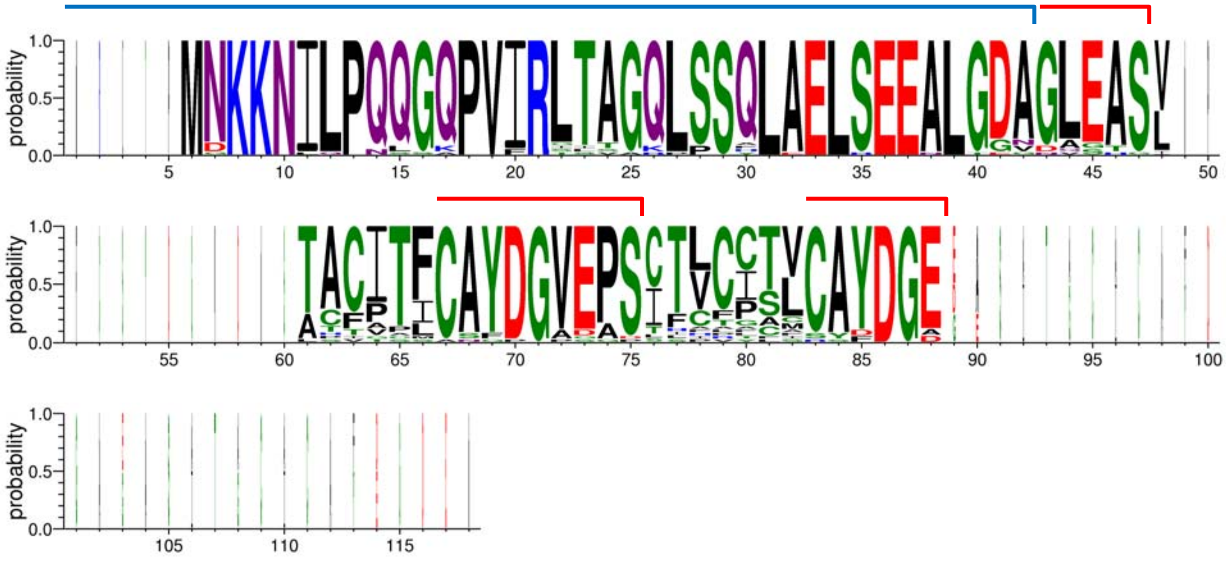


```
                                                          |--------------leader peptide------------|
                                                                                    recognition sequence
                                                                                            |---|
                                                         1                                                  50
                      patellamide C and A ABI99439   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                      patellamide B and A ABI99456   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASL--
                    Patellamide P.didemni AAY21154   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
         patellamide C and ulithiacyclamide ABI99458   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
         patellamide B and ulithiacyclamide ABI99461   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASL--
                  patellamide un-Prochloron ABI99440   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99463   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99464   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASL--
                  patellamide un-Prochloron ABI99447   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99459   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASL--
                  patellamide un-Prochloron ABI99437   (1) -----MNKKNILPQQGQPVIRLSAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99460   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99441   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99445   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99448   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASL--
                  patellamide un-Prochloron ABI99451   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASL--
        Novel compound and ulithiacyclamide ABI99444   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99449   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASL--
                  patellamide un-Prochloron ABI99446   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASL--
                  patellamide un-Prochloron ABI99462   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99457   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99443   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99452   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASL--
                  patellamide un-Prochloron ABI99438   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
                  patellamide un-Prochloron ABI99450   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV--
lissoclinamide 2/3 and lissoclinamide 4/5 ABI99436   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEAS---
                  patellamide un-Prochloron ABI99453   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEAS---
                  patellamide un-Prochloron ABI99442   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEAS---
                  patellamide un-Prochloron ABI99454   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEAS---
                  patellamide un-Prochloron ABI99455   (1) -----MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEAS---
                  patellin 6 and trunkamide ACA04491   (1) -----MNKKNILPQLGQPVIRLTAGQLSSQLAELSEEALGGVDAST----
                                trunkamide ACA04496   (1) -----MNKKNILPQLGQPVIRLTAGQLSSQLAELSEEALGGVDASTSIAP
                                        Trichamide   (1) -----MGKKNIQPNSSQPVFRSLVARP--ALEELREENLTEGNQGHG---
                      tenuecyclamides A and C ACA04484   (1) -----MDKKNILPQQGKPVIRITTGQLPSFLAELSEEALGDAGVGAS---
                    Patellamide M.aeruginosa CAO86916   (1) -----MDKKNLLPNQGAPVIRGISGKLPSHLAELSEEALGGNGAEAS---
                              microcyclamide CAO82085   (1) MRITPMDKKNILPQQGKPVFRTTTGKLPSYLAELSEEALGGNGLEAS---
                                         Consensus   (1)      MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASV

                                                                       recognition sequences
                                                                         |------|        |---|
                                                         51                                                 100
                      patellamide C and A ABI99439  (44) ----------TACITFCAYDGVEPSITVCISVCAYDGE------------
                      patellamide B and A ABI99456  (44) ----------TACITFCAYDGVEPSITVCISVCAYDGE------------
                    Patellamide P.didemni AAY21154  (44) ----------TACITFCAYDGVEPSITVCISVCAYDGE------------
         patellamide C and ulithiacyclamide ABI99458  (44) ----------TACITFCAYDGVEPSCTLCCTLCAYDGE------------
         patellamide B and ulithiacyclamide ABI99461  (44) ----------TACITFCAYDGVEPSCTLCCTLCAYDGE------------
                  patellamide un-Prochloron ABI99440  (44) ----------TACITFCAYDGVEPSCILCCTLCAYDGE------------
                  patellamide un-Prochloron ABI99463  (44) ----------TACITFCAYDGVEPSCTLCYTLCAYDGE------------
                  patellamide un-Prochloron ABI99464  (44) ----------TACITFCAYDGVEPSSTVCFTVCAYDGE------------
                  patellamide un-Prochloron ABI99447  (44) ----------TACITFCAYDGVEPSCTLCCTVCAYDGE------------
                  patellamide un-Prochloron ABI99459  (44) ----------TACITFCAYDGVEPSITVCITVCAYDGE------------
                  patellamide un-Prochloron ABI99437  (44) ----------TACITFCAYDGVEPSCTLCCTLRAYDGE------------
                  patellamide un-Prochloron ABI99460  (44) ----------TACITFCAYDGVEPSCTLCCTFCAYDGE------------
                  patellamide un-Prochloron ABI99441  (44) ----------TACITFCAYDGVEPSCTVCCAVCAYDGE------------
                  patellamide un-Prochloron ABI99445  (44) ----------TACITFCAYDGVEPSCTLCCALCAYDGE------------
                  patellamide un-Prochloron ABI99448  (44) ----------TACITFCAYDGVEPSITVCISACAYDGE------------
                  patellamide un-Prochloron ABI99451  (44) ----------TACVTFCAYDGVEPSITVCISVCAYDGE------------
        Novel compound and ulithiacyclamide ABI99444  (44) ----------AACITFCAYDGVEPSCTLCCTLCAYDGE------------
                  patellamide un-Prochloron ABI99449  (44) ----------AACITFCAYDGVEPSITVCISVCAYDGE------------
                  patellamide un-Prochloron ABI99446  (44) ----------TTCITFCAYDGVEPSITVCISVCAYDGE------------
                  patellamide un-Prochloron ABI99462  (44) ----------TTCITFCAYDGVEPSCTLCCTLCAYDGE------------
                  patellamide un-Prochloron ABI99457  (44) ----------TACTTFCAYDGVEPPCTLCCTLCAYDGE------------
                  patellamide un-Prochloron ABI99443  (44) ----------TACIAFCAYDGVEPSCTLCCTLCAYDGE------------
                  patellamide un-Prochloron ABI99452  (44) ----------TACITLCAYDGVEPSITVCISVCAYDGE------------
                  patellamide un-Prochloron ABI99438  (44) ----------TACITLCAYDGVEPSCTLCCTLCAYDGE------------
```

```
                  patellamide un-Prochloron ABI99450   (44) ----------TACITSCAYDGVEPSCTLCCTLCAYDGE------------
lissoclinamide 2/3 and lissoclinamide 4/5 ABI99436   (43) ----------ACFPTICAYDGVEPS-FCFPTVCAYDGE------------
                  patellamide un-Prochloron ABI99453   (43) ----------ACFPTICAYDGVEPS-FCFPAVCAYDGE------------
                  patellamide un-Prochloron ABI99442   (43) ----------ACFPTICAYDGVEPS-LCFPTVCAYDGE------------
                  patellamide un-Prochloron ABI99454   (43) ----------ACFPTICAYDGVEPS-FCVPTVCAYDGE------------
                  patellamide un-Prochloron ABI99455   (43) ----------ACFPTICAYDGVEPS-FCLPTVCAYDGE------------
                  patellin 6 and trunkamide ACA04491   (42) ----------LPVPTLCSYDGVDAST-VPTLCSYDD--------------
                                 trunkamide ACA04496   (46) FCSYDGVDASTSIAPFCSYDGVDASTSIAPFCSYDD--------------
                                         Trichamide   (41) --------------PLANGPGPSGDLHPRLCSCSYDGDDE----------
                    tenuecyclamides A and C ACA04484   (43) -----------ATGCMCAYDGAGAS--ATGCMCAYDGAGASATACACAYD *
                   Patellamide M.aeruginosa CAO86916   (43) -----------ATVSICAFDGAEAS--FTGCMCAFDGAEASITGCICAFD *
                           microcyclamide CAO82085   (48) -----------HCATICAFDGAEAS--HCATICAFDGDEA----------
                                          Consensus   (51)           TACITFCAYDGVEPS TLC TVCAYDGE
```

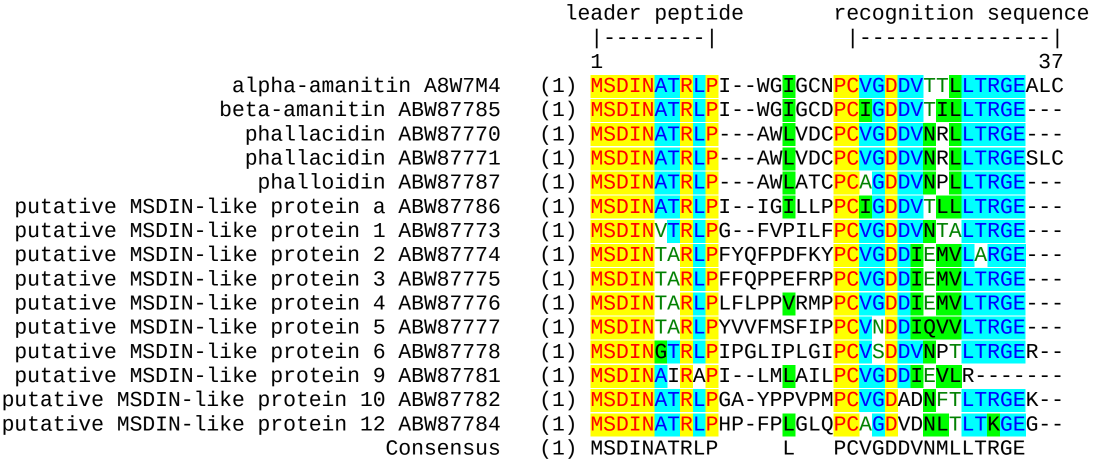
* Tenuecyclamides A and C and patellamide from M.aeruginosa CAO86916 have an additional C-terminal extension:

```
                    tenuecyclamides A and C ACA04484   (80) GAGASATACACAYE----
                   Patellamide M.aeruginosa CAO86916   (80) GAEASITGCICAFDGDEA
```

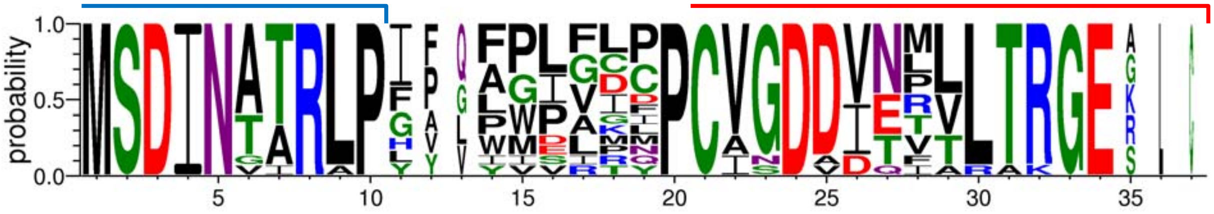


NOTE: The C-terminal stretches of narrow letters are due to the additional extensions in Tenuecyclamides A and C and a patellamide from *M. aeruginosa* CAO86916.

Supplementary Figure 8. Sequence alignment of selected precursor peptides of amatoxins and phallotoxins (mycotoxins). The logo below each alignment shows the probability of each amino acid (height of the letter) and is scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters). Blue line above sequence logo indicates leader peptide. Red line indicates recognition sequences. The core peptides are located between the leader peptides and the recognition sequences. The amatoxins and phallotoxins contain highly conserved Pro residues at the protease cleavage sites. Both the leader peptide and recognition sequences are highly conserved, but the core peptides are hypervariable.



```
                                                 leader peptide       recognition sequence
                                                   |--------|          |---------------|
                                                   1                                  37
                     alpha-amanitin A8W7M4    (1) MSDINATRLPI--WGIGCNPCVGDDVTTLLTRGEALC
                     beta-amanitin ABW87785    (1) MSDINATRLPI--WGIGCDPCIGDDVTILLTRGE---
                       phallacidin ABW87770    (1) MSDINATRLP---AWLVDCPCVGDDVNRLLTRGE---
                       phallacidin ABW87771    (1) MSDINATRLP---AWLVDCPCVGDDVNRLLTRGESLC
                        phalloidin ABW87787    (1) MSDINATRLP---AWLATCPCAGDDVNPLLTRGE---
   putative MSDIN-like protein a ABW87786    (1) MSDINATRLPI--IGILLPPCIGDDVTLLLTRGE---
   putative MSDIN-like protein 1 ABW87773    (1) MSDINVTRLPG--FVPILFPCVGDDVNTALTRGE---
   putative MSDIN-like protein 2 ABW87774    (1) MSDINTARLPFYQFPDFKYPCVGDDIEMVLARGE---
   putative MSDIN-like protein 3 ABW87775    (1) MSDINTARLPFFQPPEFRPPCVGDDIEMVLTRGE---
   putative MSDIN-like protein 4 ABW87776    (1) MSDINTARLPLFLPPVRMPPCVGDDIEMVLTRGE---
   putative MSDIN-like protein 5 ABW87777    (1) MSDINTARLPYVVFMSFIPPCVNDDIQVVLTRGE---
   putative MSDIN-like protein 6 ABW87778    (1) MSDINGTRLPIPGLIPLGIPCVSDDVNPTLTRGER--
   putative MSDIN-like protein 9 ABW87781    (1) MSDINAIRAPI--LMLAILPCVGDDIEVLR-------
  putative MSDIN-like protein 10 ABW87782    (1) MSDINATRLPGA-YPPVPMPCVGDADNFTLTRGEK--
  putative MSDIN-like protein 12 ABW87784    (1) MSDINATRLPHP-FPLGLQPCAGDVDNLTLTKGEG--
                              Consensus    (1) MSDINATRLP     L    PCVGDDVNMLLTRGE
```



Supplementary Figure 9. Sequence alignment of the precursor peptides of selected cyclotides. The logo below the alignment shows the probability of each amino acid (height of the letter) and is scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters).

```
                                          |---------signal---------|-----leader peptide-----
                                          1                                                50
        cyclotide vitri-A B1NRR3     (1) -----------------AAFALPAFA-SFEKDVITPAALEAVLNRKAPLS
         cyclotide vibi-E B1NRQ8     (1) -----------------AAFALPALASSFEKDVISFRAIQAVLEKRG---
         cyclotide vibi-I B1NRQ9     (1) -----------------AAFALPAFA-SFEKDVITPAALEAVLNRKAPLS
         cyclotide vibi-J B1NRR0     (1) -----------------AAFALPALATSFEKDFITHETVQEILKKVGSN-
         cyclotide vibi-K B1NRR1     (1) -----------------AAFALPAFA-SFEKDVITPSVLEAVLNRKAPLS
 cyclotide precursor Vbc1 ABW08090   (1) -----------------AAFALPALASSFEKDVISFRAIQAVLEKRG---
 cyclotide precursor Vbc2 ABW08091   (1) -----------------AAFALPAFA-SFEKDVITPAALEAVLNRKAPLS
 cyclotide precursor Vbc3 ABW08092   (1) -----------------AAFALPALATSFEKDFITHETVQEILKKVGSN-
 cyclotide precursor Vbc4 ABW08093   (1) -----------------AAFALPAFA-SFEKDVITPSVLEAVLNRKAPLS
 cyclotide precursor Vbc5 ABW08094   (1) -----------------AAFALPAFA-SFEKDVITPAALEAVLNRKAPLY
 cyclotide precursor Vbc6 ABW08095   (1) -----------------AAFALPAFA-SFEKDVITPAALEAVLNRKAPLS
cyclotide precursor Mra13 ACI29317   (1) ---------MLVGVVLIATFATAAFAA-LDKDAITPGAIHALSKSMK---
         Cycloviolacin O8 P58440     (1) ----MEMKNMVVGLFLIAAFALPALATNFEKDFITHETVQAILKKVGPS-
         Cycloviolacin O9 B1NRR2     (1) -----------------AAFALPAFA-SFEKDVITPAALEAVLNRKAPLY
        Cycloviolacin O11 P58443     (1) ----MEMKNMVVGLFLIAAFALPALATSFEKDFITHETVQAILKKVGPS-
        Cycloviolacin O13 Q5USN8     (1) ----MDAKKMFVALVLIATFALP-SLATFEKDFITPETIQAILKKSAPLS
   cyclotide precursor c1 ACI29323   (1) ----MEMKNMVVGLFLIAAFALPALATNFEKDFITHETVQAILKKVGPS-
    cyclotide precursor 1 ABI48956   (1) ----MDSKIVFVALVLIATFALPTLAT-FEKDFITTEAVRAILKKTN---
   cyclotide precursor 1a ACG69841   (1) ----MDSKIVFVALVLIATFALPTLAT-FEKDFITTEAVRAILKKTN---
   cyclotide precursor 1c ACG69843   (1) ----MDSKIVFVALVLIATFALPSLAT-FEKDFITPEAVRAILKKTN---
    cyclotide precursor 2 ABI48957   (1) ----MDSKIVFVALVLIATFALPSLAT-FEKDFITTEAVRSILQKTN---
   cyclotide precursor 2a ACG69844   (1) ----MDSKIVFVALVLIATFALPPLAT-FEKYFITTEAVRAILQKTN---
   cyclotide precursor 2b ACG69845   (1) ----MDSKIVFVALVLIATFALPSLAT-FEKDFITTEAVRSILQKTN---
   cyclotide precursor 2c ACG69846   (1) ----MDSKIVFVALVLIATFALPSLAT-FEKDFITTEAVRSILQKTN---
   cyclotide precursor 2d ACG69847   (1) ----MDSKIVFVALVLIATFALPSLAT-FEKDFITPEAVKAILKKTN---
   cyclotide precursor 2e ACG69848   (1) ----MDSKIVFVALVLIATFALPSLAT-FEKDFITTETVRAILKKTN---
    cyclotide precursor 3 ABI48958   (1) ----MDAKKMFLALVLIATFAVIPSFATFEKDFISPEAVQAILKKRAPLS
   cyclotide precursor 3a ACG69849   (1) ----MDAKKMFLALVLIATFAVIPSFATFEKDIISPETIQAVLEKTAPH-
   cyclotide precursor 3b ACG69850   (1) ----MDAKKMFLALVLIATFAVIPSLATFEKDFITREAIQAILKKSAPLS
   cyclotide precursor 3c ACG69851   (1) ----MDAKKMFLALVLIATFAVIPSFATFEKDFISPEAVQAILKKRAPLS
   cyclotide precursor 3d ACG69852   (1) ----MDAKKMFVALVLVATFVLP-SLATFEKDFITPETIQAILKKSAPLS
    cyclotide precursor 4 ABI48959   (1) --MESNKKMQLVGFVLLAAFALPALASSFEKDVISPRAIQAVLEKRG---
   cyclotide precursor 4a ACG69853   (1) --MESNKKMQLVGFVLLAAFALPALASSFEKDVISPRAIQAVLEKRG---
   cyclotide precursor 4b ACG69854   (1) --MESNKKMQLIGFVLLAAFALPALASSFEKDVISPHAIQAVLEKRG---
   cyclotide precursor 4c ACG69855   (1) --MESNKKMQLVGFVLLAAFALPALASSFEKDVISPRAIQAVLEKRG---
    cyclotide precursor 5 ACG69856   (1) -------MKMFVALVLFAAFALPAAFA-AQQDAIALQAYEELLKNGAAN-
    cyclotide precursor 5 ABI48960   (1) -------MKMFVALVLFAAFALPAAFA-AQQDAIALQAYEELLKNGAAN-
               Violacin A Q2HY54     (1) MDAQKMKMVIGLVLVATTAFALMIPAASAVDDFITRRAYDNLVKSGAIK-
                Kalata B1 P56254     (1) -MAKFTVCLLLCLLLAAFVGAFGSELSDSHKTTLVNEIAEKMLQRKILDG
                Kalata B2 P58454     (1) -MAKFTNCLVLSLLLAAFVGAFGAEFSEADKATLVNDIAENIQKEILGEV
             Kalata B3/B6 P58455     (1) -MAKFTKSLVLCLLLAAFVGAFGAELSEADKANVVNEIAANIQREILKGV
                Kalata B7 P58457     (1) -MAKFTNCLALCLLLAAVVGAFGVELSEADKSAVVNEIAEKMALQEMLDG
           Pro-tricyclons P0C589     (1) -----MKTMASFVLVLVAAFALPAAFASLETKDVITRAAYEKLVESGAIQ
                       Consensus     (1)        K M VALVLIAAFALPA ATSFEKD IT EAV AIL K A

                                         ------------------leader peptide------------------
                                         51                                              102
        cyclotide vitri-A B1NRR3    (33) NIM-----MENDAIVNVIANV-------------------------------
         cyclotide vibi-E B1NRQ8    (31) LSK-----LEDDPVLSALAHT-------------------------------
         cyclotide vibi-I B1NRQ9    (33) NIM-----MENDAILNVIANV-------------------------------
         cyclotide vibi-J B1NRR0    (33) SNG-----MLDEQTISALTG--------------------------------
         cyclotide vibi-K B1NRR1    (33) NIM-----MENDAILNVIANV-------------------------------
 cyclotide precursor Vbc1 ABW08090  (31) LSK-----LEDDPVLSALAHT-------------------------------
 cyclotide precursor Vbc2 ABW08091  (33) NIM-----MENDAILNVIANV-------------------------------
 cyclotide precursor Vbc3 ABW08092  (33) SNG-----MLDEQTISALTG--------------------------------
 cyclotide precursor Vbc4 ABW08093  (33) NIM-----MENDAILNVIANV-------------------------------
 cyclotide precursor Vbc5 ABW08094  (33) NIM-----MENDAILNVIANV-------------------------------
 cyclotide precursor Vbc6 ABW08095  (33) NIM-----MENDAIVNVIANV-------------------------------
cyclotide precursor Mra13 ACI29317  (38) ---------LPTDAVNVLTKS-------------------------------
         Cycloviolacin O8 P58440    (46) SNG-----MLDEQTISALTG--------------------------------
         Cycloviolacin O9 B1NRR2    (33) NIM-----MENDAILNVIANV-------------------------------
        Cycloviolacin O11 P58443    (46) SNG-----MLDEQAISALTG--------------------------------
```

```
          Cycloviolacin O13 Q5USN8   (46) NIM-----LEEDVINALLK-S-------------------------------
   cyclotide precursor c1 ACI29323   (46) SNG-----MLDEQTISALTG--------------------------------
    cyclotide precursor 1 ABI48956   (43) ANV-----MPSEDVINALTG--------------------------------
   cyclotide precursor 1a ACG69841   (43) ANV-----MPSEDVINALTG--------------------------------
   cyclotide precursor 1c ACG69843   (43) ANV-----MPSEDAINALTG--------------------------------
    cyclotide precursor 2 ABI48957   (43) SNA-----MPSEDAINALTG--------------------------------
   cyclotide precursor 2a ACG69844   (43) SNA-----MPSENAINALTG--------------------------------
   cyclotide precursor 2b ACG69845   (43) SNA-----MPSEDAINALTG--------------------------------
   cyclotide precursor 2c ACG69846   (43) SNA-----MPSEDAINALTG--------------------------------
   cyclotide precursor 2d ACG69847   (43) ANV-----MPSEDAINALTG--------------------------------
   cyclotide precursor 2e ACG69848   (43) SNT-----MPSEDAINALTG--------------------------------
    cyclotide precursor 3 ABI48958   (47) NIM-----LEEDAMSALIK-S-------------------------------
   cyclotide precursor 3a ACG69849   (46) ---------LKDAMNAIIK-S-------------------------------
   cyclotide precursor 3b ACG69850   (47) NIM-----LEEDVMNALIK-S-------------------------------
   cyclotide precursor 3c ACG69851   (47) NIM-----LEEDAMSALIK-S-------------------------------
   cyclotide precursor 3d ACG69852   (46) NIM-----LEEDVINALLK-S-------------------------------
    cyclotide precursor 4 ABI48959   (46) LSK-----LEDDPVLSALART-------------------------------
   cyclotide precursor 4a ACG69853   (46) LSK-----LEDDPVLSALART-------------------------------
   cyclotide precursor 4b ACG69854   (46) FSK-----LEDDPVVSALART-------------------------------
   cyclotide precursor 4c ACG69855   (46) LSK-----LEDDPVLSALART-------------------------------
    cyclotide precursor 5 ACG69856   (42) ------------------GMT-------------------------------
    cyclotide precursor 5 ABI48960   (42) ------------------GMT-------------------------------
                 Violacin A Q2HY54   (50) -----------D--IPVMA---------------------------------
                  Kalata B1 P56254   (50) VEATLVTDVAEKMFLRKMKAE-------------------------------
                  Kalata B2 P58454   (50) KTS----ETVLTMFLKEMQLKGLPVCGETCFGGTCNTPGCSCT-WPICTRDS
               Kalata B3/B6 P58455   (50) KSS----ETTLTMFLKEMQLKGLPTCGETCFGGTCNTPGCSCSSWPICTRNG
                  Kalata B7 P58457   (50) VD---------KLFLRKMK---------------------------------
             Pro-tricyclons P0C589   (46) GITMTKTIISNPILEEALVAHFNRKLGGGTIFDCGESCFLGTCYTKGCSC*
                         Consensus   (51)         M  D II  L
```

*The procyclons POC589 have an unusually long leader peptide that contains the following additional intervening sequence appended where the asterisk is shown:

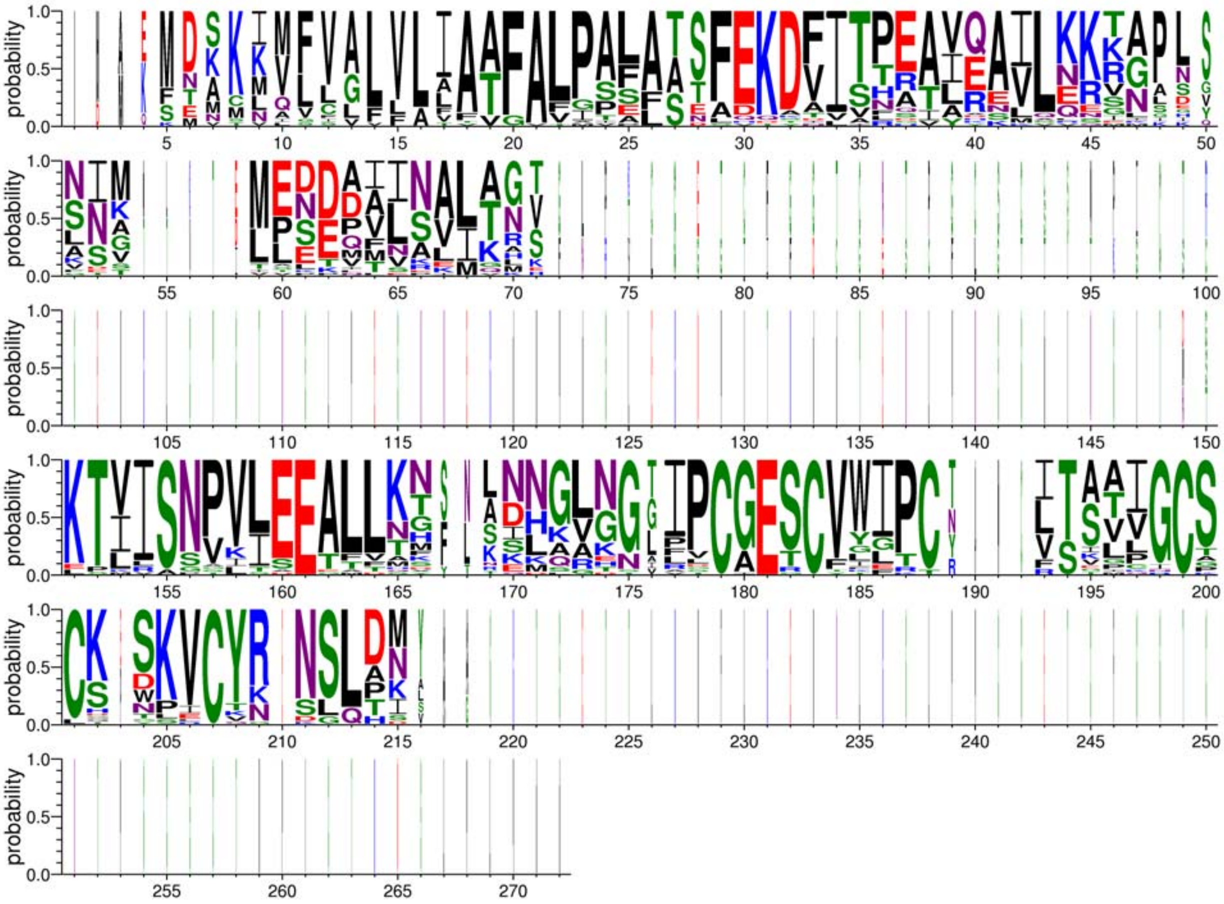
```
             Pro-tricyclons P0C589   (96) GEWKLCYGTNSLPESNNEKAMVASLEKDVITRAAYENLVNSGAIQGITMT
```

```
                                          ----leader peptide------|
                                          101                                            150
          cyclotide vitri-A B1NRR3   (49) KTVISNPVLEEALLKT---NHGVNG-IPCGESCVWIPC----ITSAIGCS
           cyclotide vibi-E B1NRQ8   (47) KTIISNPVIEEALLNG--ANLKAGNGIPCAESCVWIPCT---VTALIGCG
           cyclotide vibi-I B1NRQ9   (49) KTVISNPVLEEALLKT---NHGVNG-IPCGESCVWIPC----LTSTVGCS
           cyclotide vibi-J B1NRR0   (48) KTIISNPLLEEALFKSSNSINALGGTFPCGESCVWIPC----ISKVIGCA
           cyclotide vibi-K B1NRR1   (49) KTVISNPVLEEALLKT---NHGVNG-IPCGESCVWIPC----LTSAVGCP
 cyclotide precursor Vbc1 ABW08090   (47) KTIISNPVIEEALLNG--ANLKAGNGIPCAESCVWIPCT---VTALIGCG
 cyclotide precursor Vbc2 ABW08091   (49) KTVISNPVLEEALLKT---NHGVNG-IPCGESCVWIPC----LTSTVGCS
 cyclotide precursor Vbc3 ABW08092   (48) KTIISNPLLEEALFKSSNSINALGGTFPCGESCVWIPC----ISKVIGCA
 cyclotide precursor Vbc4 ABW08093   (49) KTVISNPVLEEALLKT---NHGVNG-IPCGESCVWIPC----LTSAVGCP
 cyclotide precursor Vbc5 ABW08094   (49) KTVISNPVLEEALLKT---NHGVNG-IPCGESCVWIPC----LTSAVGCS
 cyclotide precursor Vbc6 ABW08095   (49) KTVISNPVLEEALLKT---NHGVNG-IPCGESCVWIPC----ITSAIGCS
cyclotide precursor Mra13 ACI29317   (50) KTLVSPTKLAEEFLND--ANDGVNG-IPCGESCVYLPC----FTTIIGCK
           Cycloviolacin O8 P58440   (61) KTIISNPVLEEALLTHSNSINALGGTLPCGESCVWIPC----ISSVVGCS
           Cycloviolacin O9 B1NRR2   (49) KTVISNPVLEEALLKT---NHGVNG-IPCGESCVWIPC----LTSAVGCS
          Cycloviolacin O11 P58443   (61) KTIISNPVLEEALLTHSNSINALGGTLPCGESCVWIPC----ISAVVGCS
          Cycloviolacin O13 Q5USN8   (61) KTVISNPIIEEAFLKN---SNGLNG-IPCGESCVWIPC----ISAAIGCS
   cyclotide precursor c1 ACI29323   (61) KTIISNPVLEEALLTHSNSINALGGTLPCGESCVWIPC----ISSVVGCS
    cyclotide precursor 1 ABI48956   (58) KTVISNVVLEETLVMK--LDNGLTG-IPCGESCVYLPC----FTAPLGCS
   cyclotide precursor 1a ACG69841   (58) KTVISNVVLEETLVMK--LDNGLTG-IPCGEGCVYLPC----FTAPLGCS
   cyclotide precursor 1c ACG69843   (58) KTVISNVLLEEALLKN--LDNGLNG-IPCGESCVWIPC----LTAAIGCS
    cyclotide precursor 2 ABI48957   (58) KTLLSSVVLEEALLKN--LDNGRNG-IPCGESCVWIPC----LTATIGCS
   cyclotide precursor 2a ACG69844   (58) KTLISSVVLEEALLKN--LDHGRNG-VPCGESCVWIPC----LTSAIGCS
   cyclotide precursor 2b ACG69845   (58) KTLLSSVVLEEALLKN--LDNGRNG-IPCGESCVWIPC----LTATIGCS
   cyclotide precursor 2c ACG69846   (58) KTLISSVVLEEALLKN--LDNGRNG-IPCGESCVLIPC----ISSVIGCS
   cyclotide precursor 2d ACG69847   (58) KTVISNVVLEEALLKN--LDNGLNG-IPCGESCVWIPC----LTAAIGCS
   cyclotide precursor 2e ACG69848   (58) KTLISSVVLDEALLKN--LDNGLNGVIPCGESCVFIPC----ISSVIGCS
    cyclotide precursor 3 ABI48958   (62) KTVISNPVIEEALLKN---SNSLHG-IPCAESCVYLPC----VTIVIGCS
   cyclotide precursor 3a ACG69849   (57) KTVISNPVIEEALLKN---SNGLKG-AGCIETCYTFPC----ISEMINCS
   cyclotide precursor 3b ACG69850   (62) KSVISNPVIEEALLKN---NNGLNG-IPCGESCVWIPC----ISSAIGCS
   cyclotide precursor 3c ACG69851   (62) KTVISNPVIEEALLKN---SNSLHG-IPCAESCVYLPC----VTIVIGCS
   cyclotide precursor 3d ACG69852   (61) KTVISNPIIEEAFLKN---SNGLNG-IPCGESCVWIPC----ISGAIGCS
    cyclotide precursor 4 ABI48959   (62) KTIISNPVIEEALLNG--ANLKAGNGIPCAESCVWIPCT---VTALLGCS
   cyclotide precursor 4a ACG69853   (62) KTIISNPVIEEALLNG--ANLKAGNGIPCAESCVWIPCT---VTALLGCS
   cyclotide precursor 4b ACG69854   (62) KTIISNPVIEEALLNG--TNLKAGNGIPCAESCVWIPCT---VTALVGCS
   cyclotide precursor 4c ACG69855   (62) KTIISNPVIEEALLNG--ANLKAGNTIPCAESCVWIPCT---VTALLGCS
```

```
    cyclotide precursor 5 ACG69856   (45) ETVISNPVLEEALTAFS--KKGLEG-RLCGERCVIERTRAWCRTVGCICS
    cyclotide precursor 5 ABI48960   (45) ETVISNPVLEEALTAFS--KKGLEG-RLCGERCVIERTRAWCRTVGCICS
                  Violacin A Q2HY54   (56) KTIISNPVLEEGMLTYYTNKKLGDSAISCGETCFKFKCY------TPRCS
                   Kalata B1 P56254   (71) ------AKTSETADQVFLKQLQLKGLPVCGETCVGGTCN------TPGCT
                   Kalata B2 P58454   (97) LPMRAGGKTSETTLHMFLKEMQLKGLPVCGETCFGGTCN------TPGCS
                Kalata B3/B6 P58455   (98) LPKRAGVKSSETTLTMFLKEMQLKGLPTCGETCFGGTCN------TPGCT
                   Kalata B7 P58457   (60) ------S--SETTLTMFLKEMQLKGLPVCGETCTLGTCY------TQGCT
              Pro-tricyclons P0C589  (146) KTIISNPILEEALVSHFNRKLGGGTIFDCGESCFLGTCY------TKGCS
                          Consensus  (151) KTVISNPVLEEALLK      GL G IPCGESCVWIPC    ITA IGCS

                                               recognition sequence
                                                      |----|
                                           151                                            200
          cyclotide vitri-A B1NRR3   (91) CK-SKVCYR-NSLDN-----------------------------------
           cyclotide vibi-E B1NRQ8   (92) CS-NKVCYN-SLQTKY----------------------------------
           cyclotide vibi-I B1NRQ9   (91) CK-SKVCYR-NSLDN-----------------------------------
           cyclotide vibi-J B1NRR0   (94) CK-SKVCYK-NSLA------------------------------------
           cyclotide vibi-K B1NRR1   (91) CK-SKVCYR-NSLDN-----------------------------------
 cyclotide precursor Vbc1 ABW08090   (92) CS-NKVCYN-SLQTKY----------------------------------
 cyclotide precursor Vbc2 ABW08091   (91) CK-SKVCYR-NSLDN-----------------------------------
 cyclotide precursor Vbc3 ABW08092   (94) CK-SKVCYK-NSLA------------------------------------
 cyclotide precursor Vbc4 ABW08093   (91) CK-SKVCYR-NSLDN-----------------------------------
 cyclotide precursor Vbc5 ABW08094   (91) CK-SKVCYR-NSLDN-----------------------------------
 cyclotide precursor Vbc6 ABW08095   (91) CK-SKVCYR-NSLDN-----------------------------------
cyclotide precursor Mra13 ACI29317   (93) CQ-GKVCY------------------------------------------
           Cycloviolacin O8 P58440  (107) CK-SKVCYK-NSLA------------------------------------
           Cycloviolacin O9 B1NRR2   (91) CK-SKVCYR-NSLDN-----------------------------------
          Cycloviolacin O11 P58443  (107) CK-SKVCYK-NSLA------------------------------------
          Cycloviolacin O13 Q5USN8  (103) CK-SKVCYR-NSLDN-----------------------------------
   cyclotide precursor c1 ACI29323  (107) CK-SKVCYK-NSLA------------------------------------
    cyclotide precursor 1 ABI48956  (101) CS-SKVCYR-NSLHM-----------------------------------
   cyclotide precursor 1a ACG69841  (101) CS-SKVCYR-NSLHM-----------------------------------
   cyclotide precursor 1c ACG69843  (101) CS-SKVCYR-NSLHM-----------------------------------
    cyclotide precursor 2 ABI48957  (101) CK-SKVCYR-NSLDM-----------------------------------
   cyclotide precursor 2a ACG69844  (101) CK-SSVCYR-NSLDM-----------------------------------
   cyclotide precursor 2b ACG69845  (101) CK-SKVCYR-NSLDM-----------------------------------
   cyclotide precursor 2c ACG69846  (101) CK-SKVCYR-NSLDM-----------------------------------
   cyclotide precursor 2d ACG69847  (101) CS-SKVCYR-NSLDM-----------------------------------
   cyclotide precursor 2e ACG69848  (102) CK-SKVCYR-NSLDM-----------------------------------
    cyclotide precursor 3 ABI48958  (104) CK-DKVCY--NSLDI-----------------------------------
   cyclotide precursor 3a ACG69849   (99) CK-NSRCQK-NSLDI-----------------------------------
   cyclotide precursor 3b ACG69850  (104) CK-NKVCYR-KSLDI-----------------------------------
   cyclotide precursor 3c ACG69851  (104) CK-DKVCY--NSLDI-----------------------------------
   cyclotide precursor 3d ACG69852  (103) CK-SKVCYR-NSLDN-----------------------------------
    cyclotide precursor 4 ABI48959  (107) CK-DKVCYN-SLQTKY----------------------------------
   cyclotide precursor 4a ACG69853  (107) CK-DKVCYN-SLQTKY----------------------------------
   cyclotide precursor 4b ACG69854  (107) CS-DKVCYN-SLQTKY----------------------------------
   cyclotide precursor 4c ACG69855  (107) CK-DKVCYN-SLQTKY----------------------------------
    cyclotide precursor 5 ACG69856   (92) LH-TLECVR-NSLPM-----------------------------------
    cyclotide precursor 5 ABI48960   (92) LH-TLECVR-NSLPM-----------------------------------
                  Violacin A Q2HY54  (100) CS-YPVCK------------------------------------------
                   Kalata B1 P56254  (109) CS-WPVCTR-NGLPSLAA--------------------------------
                   Kalata B2 P58454  (141) CT-WPICTR-DSLPMSAGGKTSETTLHMFLKEMQLKGLPVCGETCFGGTC*
                Kalata B3/B6 P58455  (142) CDPWPICTR-DGLPSAAA--------------------------------
                   Kalata B7 P58457   (96) CS-WPICKR-NGLPDVAA--------------------------------
              Pro-tricyclons P0C589  (190) CGEWKLCYGENSLAI-----------------------------------
                          Consensus  (201) CK SKVCYR NSL
```

*Kalata B2 P58454 has an unusually long precursor peptide. The additional C-terminal extension is shown below.

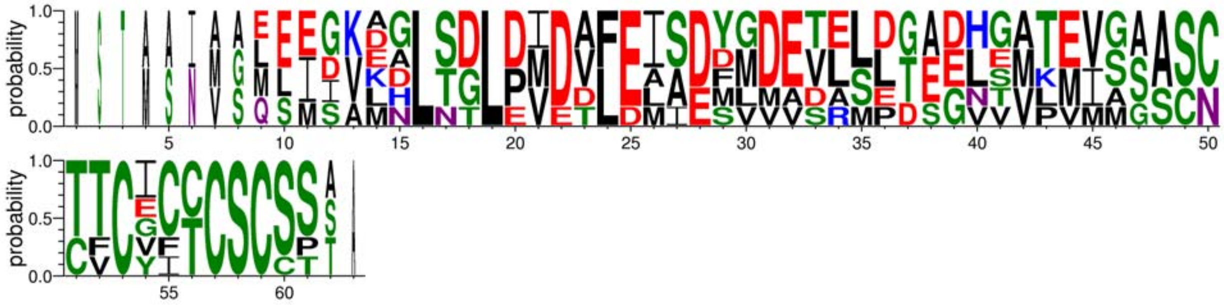
```
                                           251                272
                   Kalata B2 P58454  (189) NTPGCSCTWPICTRDSLPLVAA
```



NOTE: The stretches of narrow letters are due to the unusually long precursor peptides for Pro-tricyclons P0C589 and Kalata B2 P58454.

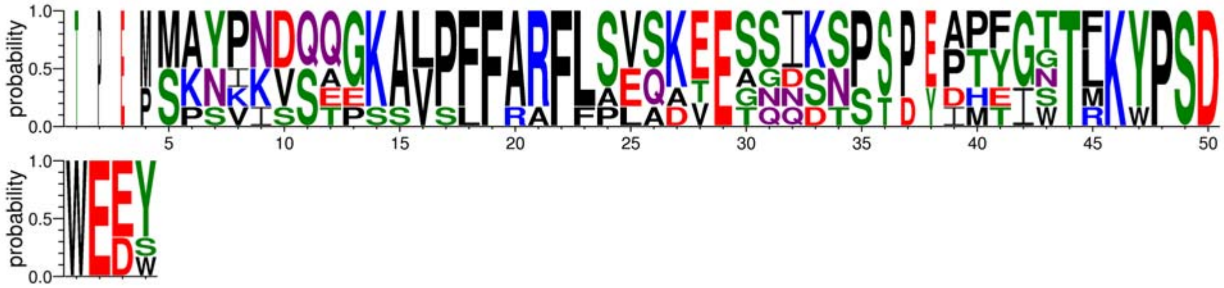
Supplementary Figure 10. Sequence alignment of the precursor peptides of selected thiopeptides. The logo below the alignments shows the probability of each amino acid (height of the letter) and is scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters). Some of the leader peptides are of the double Gly type (e.g. thiamuracin and the thiocillins). Two highly conserved Leu residues are found in the leader as well as a preponderance of Asp and Glu residues.

```
                                                |-----------------leader peptide---------------|
                                                1                                             48
                     Nosiheptide ACR48342   (1) ----------MDAAHLSDLDIDALEISEFLDESRLEDSEVVAKVMS-A
Micrococcin P1/P2 Thiocillins I-IV Q812G9   (1) --------MSEIKKALNTLEIEDFDAIEMVDVDAMP--ENEALEIMGA
                   Thiomuracin A ACS83781   (1) -------------MDLSDLPMDVFELADDGVAVESLTAGHGMTEVGA-
                          GE2270 ACS83766   (1) ------MSELESKLNLSDLPMDVFEMADSGMEVESLTAGHGMPEVGA-
                     Thiostrepton ACN80670  (1) ---MSNAALEIGVEGLTGLDVDTLEISDYMDETLLDGEDLT-------
                        Siomycin ACN80645   (1) MSTAAIVGQEIGVDGLTGLDVDALEISDYMDETLLDGEDLS-------
                                Consensus   (1)           I   GLSDLDID LEISDYMDET LD ADL MTEVSAA

                                                |----core peptide----|
                                                42                  63
                     Nosiheptide ACR48342  (37) -------SCTTCECCCSCSS--
Micrococcin P1/P2 Thiocillins I-IV Q812G9  (38) -------SCTTCVCTCSCCTT-
                   Thiomuracin A ACS83781  (34) ------SCNCFCYICCSCSSA-
                          GE2270 ACS83766  (41) ------SCNCVCGFCCSCSPSA
                    Thiostrepton ACN80670  (38) VTMIASASCTTCICTCSCSS--
                        Siomycin ACN80645  (41) VTMVSSASCTTCICTCSCSS--
                                Consensus  (41) MTEVSAASCTTCICTCSCSS
```



Supplementary Figure 11. Sequence alignment of selected precursor peptides of microviridins. The logo below the alignments shows the probability of each amino acid (height of the letter) and is scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters). Two highly conserved Phe residues are found in the leader peptide. The precursor peptide from *Anabaena* that contains multiple core peptides is shown at the bottom.

```
                                        |-------------leader peptide------------|
                                        1                                      41
Microviridin B (NIES298) CAQ16116   (1) ----MAYPNDQQGKALPFFARFLSVSKEESSIKSPSPEPT-
Microviridin B (NIES843) BAG02233   (1) ----MAYPNDQQGKALPFFARFLSVSKEESSIKSPSPEPT-
 Microviridin B (UOWOCC) CAQ16121   (1) ----MAYPNDQQGKALPFFARFLSVSKEESSIKSPSP-DHE
          Microviridin 2 ACC54552   (1) ---MSKNIKVSTGSAVPFFARFLSEQDTETGDSTSTD-IPT
          Microviridin K ACC54551   (1) ---MSKNVKVSAPKAVPFFARFLAEQAVEANNSNS---AP-
                   Microviridin J   (1) TPEPSPSKISSEEKSVSLFRAFFPLAK--GQQDNP--YAM-
                        Consensus   (1)     SAYPNDSQGKALPFFARFLSVSKEESSIKSPSP

                                         core peptide
                                        |------------|
                                        42          55
Microviridin B (NIES298) CAQ16116  (36) FGTTLKYPSDWEEY
Microviridin B (NIES843) BAG02233  (36) YGGTFKYPSDWEDY
 Microviridin B (UOWOCC) CAQ16121  (36) -ISTRKYPSDWEEW
          Microviridin 2 ACC54552  (37) -IWTFKWPSDWEDS
          Microviridin K ACC54551  (35) YGNTMKYPSDWEEY
                   Microviridin J  (37) FGTTLKYPSDWEEY
                        Consensus  (41) FGTTLKYPSDWEEY
```



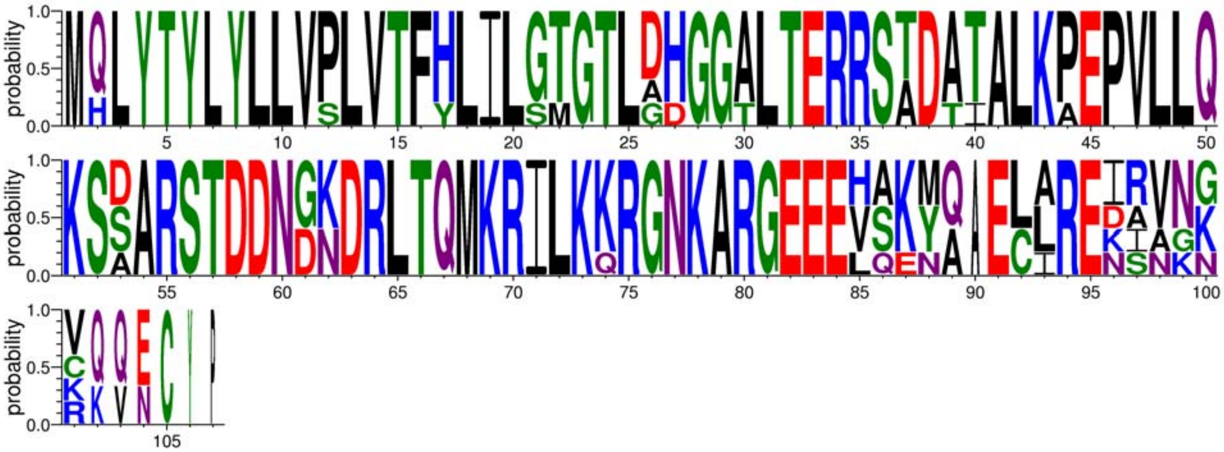
```
                                                                      core peptide
                               |------------leader peptide------------|----------
                               1                                                50
       Anabaena PCC7120    (1) MPENRQEDLNAQAVPFFARFLEGQNCEDLTDEESEAVSGGKPGQTRKYPS

                                               core peptide       core peptide
                               ---|           |------------|    |---------------|
                               51                                              100
       Anabaena PCC7120   (51) DCEDGNGVTGKLRDEDIAVTLKYPSDNEDNGGGEIVTLKFPSDDDDQPVG
```

Supplementary Figure 12. Sequence alignment of the precursor peptides for selected conopeptides. The logos below the alignments show the probability of each amino acid (height of the letter) and are scaled (width of the letter) according to how many sequences contributed to that position (i.e. narrow letters were generated from a smaller number of sequences than wider letters).

Conopeptides lacking disulfide crosslinks

a. Conantokins

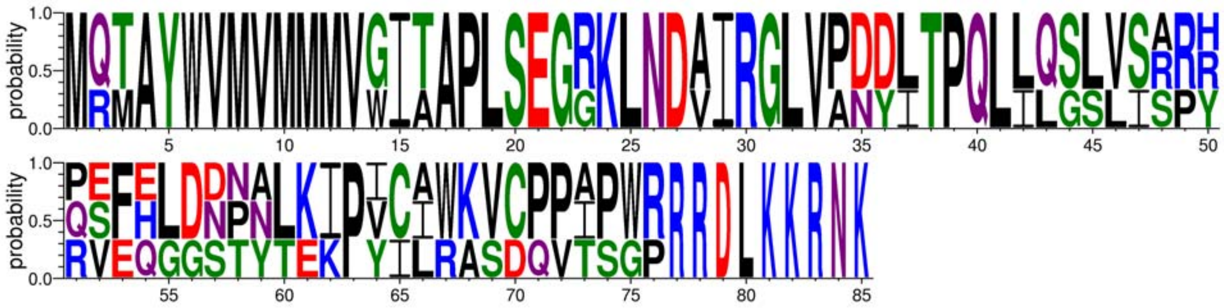
```
                                 |------------signal-----------|-----------------
                                 1                                              50
Conantokin-E P0C8D9    (1) --------LLVPLVTFHLILGMGTLDHGGALTERRSADATALKPEPVLLQ
Conantokin-P P0C8E3    (1) MQLYTYLYLLVPLVTFHLILSTGTLAHGGTLTERRSTDTTALKPEPVLLQ
Conantokin-G P07231    (1) MHLYTYLYLLVPLVTFHLILGTGTLDDGGALTERRSADATALKAEPVLLQ
Conantokin-L P69745    (1) MQLYTYLYLLVPLVTFHLILGTGTLDHGGALTERRSTDAIALKPEPVLLQ
Conantokin-R P58806    (1) MQLYTYLYLLVSLVTFYLILGTGTLGHGGALTERRSTDATALKPEPVLLQ
          Consensus    (1) MQLYTYLYLLVPLVTFHLILGTGTLDHGGALTERRSTDATALKPEPVLLQ

                           ---------leader peptide------|
                           51                                              100
Conantokin-E P0C8D9   (43) KSDARSTDDNDKDRLTQMKRILKKRGNKARGEEEHSKYQ-ECLREIRVNN
Conantokin-P P0C8E3   (51) KSDARSTDDNDKDRLTQMKRILKKRGNKARGEEEHSKYQ-ECLREIRVNK
Conantokin-G P07231   (51) KSAARSTDDNGKDRLTQMKRILKQRGNKARGEEELQENQ-ELIREKSNGK
Conantokin-L P69745   (51) KSSARSTDDNGNDRLTQMKRILKKRGNKARGEEEVAKMAAELAREDAVNG
Conantokin-R P58806   (51) KSSARSTDDNGNDRLTQMKRILKKRGNKARGEEEVAKMAAELARENIAKG
          Consensus   (51) KSSARSTDDNGKDRLTQMKRILKKRGNKARGEEEVAK Q ELLRE  VN

                           101
Conantokin-E P0C8D9   (92) VQQEC--
Conantokin-P P0C8E3  (100) VQQEC--
Conantokin-G P07231  (100) R------
Conantokin-L P69745  (101) K------
Conantokin-R P58806  (101) CKVNCYP
          Consensus  (101)     C
```



This family has highly conserved leader peptides including an AlaArg protease cleavage site to remove the leader peptide**.** One highly conserved Pro at position 46 is unusual in leader peptides.

b. Contulakins

```
                           |-------signal-------|------leader peptide—(G)---|
                           |-------signal-------|---------leader peptide-----
                           1                                               50
  Contulakin-G Q9XYR5   (1) MQTAYWVMVMMMVWIAAPLSEGGKLNDVIRGLVPDDITPQLILGSLISRR
Contulakin-Lt1 Q2I2P1   (1) MRTAYWVMVMMMVGITAPLSEGRKLNDAIRGLVADYLTPQLLQSLVSAPY
Contulakin-Lt2 Q2I2P2   (1) MQMAYWVMVMMMVGITAPLSEGRKLNDAIRGLVPNDLTPQLLQSLVSRRH
            Consensus   (1) MQTAYWVMVMMMVGITAPLSEGRKLNDAIRGLVPDDLTPQLLQSLVSARH

                                          recognition sequences
                                           |------(G)--------|
                           ---(Lt)----|             |--(Lt)--|
                           51                               85
  Contulakin-G Q9XYR5  (51) QSEEGGSNATKKPYILRASDQVASGP---------
Contulakin-Lt1 Q2I2P1  (51) PEFQLDDPNLEIPVCIWKVCPPIPWRRRDLKKRNK
Contulakin-Lt2 Q2I2P2  (51) RVFHLDNTYLKIPICAWKVCPPTPWRRRDLKKRNK
            Consensus  (51)   F LD   LKIPICIWKVCPP PWRRRDLKKRNK
```

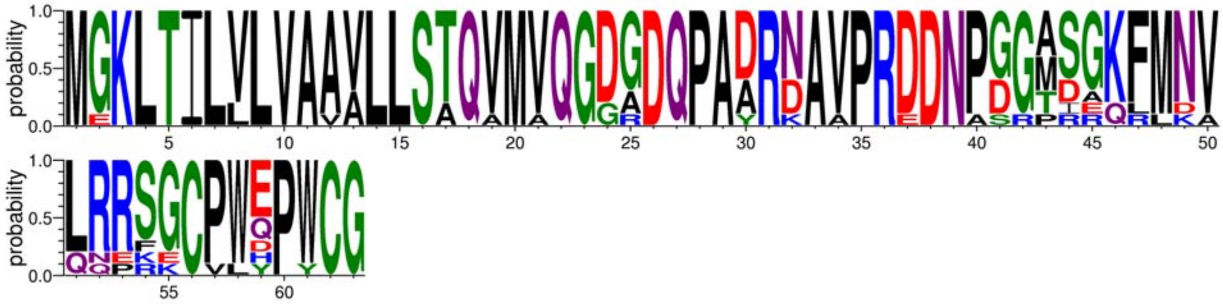


This presently small family (in terms of full sequence information) has highly conserved leader peptides including a conserved Pro at position 39.

Conopeptides with one disulfide

c. Contryphans

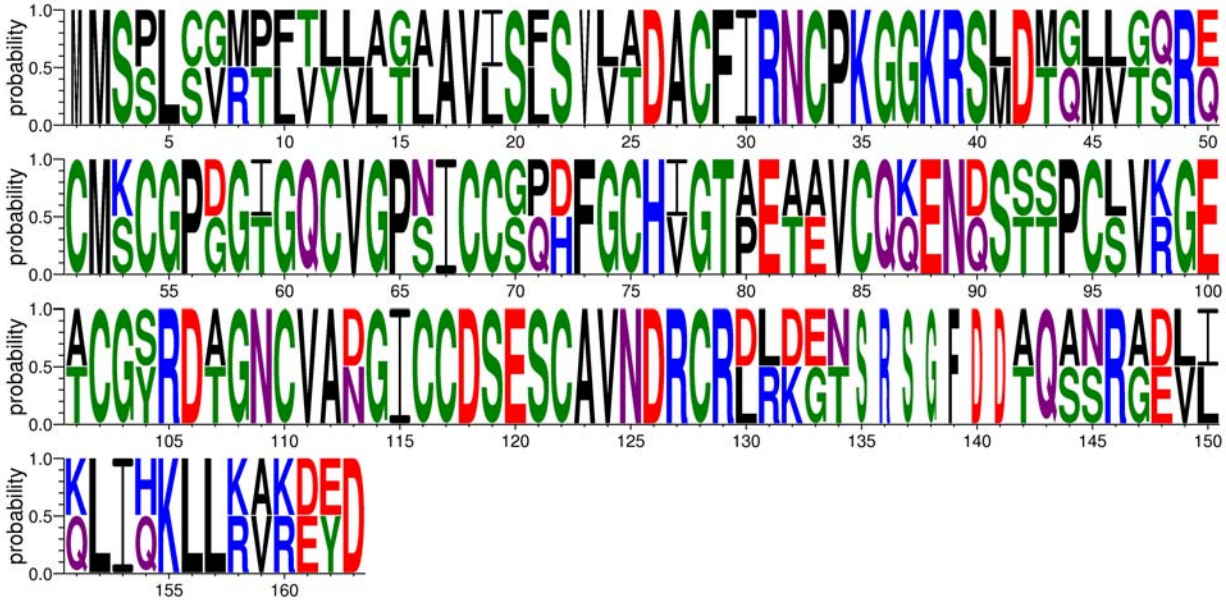
```
                                         |-------signal--------|--------leader peptide-----
                                         1                                               50
Contryphan-R/bromocontryphan P58786   (1) MGKLTILVLVAAVLLSAQVMVQGDGDQPADRNAVPRDDNPGGASGKFMNV
             Contryphan-Lt ABC74997   (1) MGKLTILLLVAAALLSTQVMVQGGGDQPAARNAVPRDDNPDGMSGQFMNV
               Contryphan-Lt Q2I2P3   (1) MGKLTILLLVAAALLSTQVMVQGGGDQPAARNAVPRDDNPDGMSGQFMNV
             Contryphan-R/Tx Q9NDA5   (1) MGKLTILVLVAAVLLSTQAMAQGDGDQPAARNAVPRDDNPDGPSAKFMNV
                Contryphan-P P58784   (1) MGKLTILVLVAAVLLSTQVMVQGDGDQPAYRNAAPRDDNPGGAIGKFMNV
                Contryphan-S P0C837   (1) MEKLTILVLVAAVLLSTQVMVQGDADQPADRDAVPRDDNAGGTDGKFMNV
               Contryphan-Sm P58787   (1) MGKLTILVLVAAVLLSTQVMVQGDADQPADRDAVPRDDNPSGTDGKFMNV
               Contryphan-Tx Q9NDA7   (1) MGKLTILVLVAVALLSTQVMVQGDGDQPADRDAVPRDDNPGGMSEKFLNA
             Glacontryphan-M P62903   (1) MGKLTILVLVAAVLLSTQVMVQGDRDQPADRNAVPRDDNPGRARRKRMKV
         D-Leu-contryphan-Tx Q9NDA6   (1) MGKLTILVLVAAVLLSAQVMVQGDGDQPADRKAVPREDNPGGASGKLMDV
                          Consensus   (1) MGKLTILVLVAAVLLSTQVMVQGDGDQPADRNAVPRDDNPGG SGKFMNV

                                         ---|        |
                                         51         63
Contryphan-R/bromocontryphan P58786  (51) LRRSGCPWEPWCG
             Contryphan-Lt ABC74997  (51) LRRSGCPWEPWCG
               Contryphan-Lt Q2I2P3  (51) LRRSGCPWEPWCG
             Contryphan-R/Tx Q9NDA5  (51) QRRSGCPWEPWCG
                Contryphan-P P58784  (51) LRRSGCPWDPWCG
                Contryphan-S P0C837  (51) QRRSGCPWEPWCG
               Contryphan-Sm P58787  (51) LRRFGCPWQPWCG
               Contryphan-Tx Q9NDA7  (51) LQRRGCPWQPYCG
             Glacontryphan-M P62903  (51) LNESECPWHPWCG
         D-Leu-contryphan-Tx Q9NDA6  (51) LRPKKCVLYPWCG
                          Consensus  (51) LRRSGCPWEPWCG
```



This family has highly conserved leader peptides including three highly conserved Pro residues at positions 28, 35, and 40, which is unusual in leader peptides.

d.Conopressins

```
                                     1                                                50
Conopressin/neurophysin Q00945    (1) MMSSLCGMPLTYLLTAAVLSLS-LTDACFIRNCPKGGKRSLDTGMVTSRE
      Lys-conopressin BAB40371    (1) -MSPLSVRTFVLVAGLAVISFSVVADACFIRNCPKGGKRSMDMQLLGQRQ
                     Consensus    (1)  MS L       L   AVIS S L DACFIRNCPKGGKRSLD  LL  R
                                      51                                               100
Conopressin/neurophysin Q00945   (50) CMKCGPGGTGQCVGPSICCGQDFGCHVGTAEAAVCQQENDSSTPCLVKGE
      Lys-conopressin BAB40371   (50) CMSCGPDGIGQCVGPNICCSPHFGCHIGTPETEVCQKENQSTSPCSVRGE
                     Consensus   (51) CM CGP G GQCVGP ICC   FGCHIGT E  VCQ EN SSSPC VKGE
                                      101                                              150
Conopressin/neurophysin Q00945  (100) ACGSRDAGNCVADGICCDSESCAVNDRCRDLDGN-------AQANRGDLI
      Lys-conopressin BAB40371  (100) TCGYRDTGNCVANGICCDSESCAVNDRCRLRKETSRSGFDDTQSSRAEVL
                     Consensus  (101)  CG RD GNCVA GICCDSESCAVNDRCR             QA RADLI
                                      151        163
Conopressin/neurophysin Q00945  (143) QLIHKLLKVRDYD
      Lys-conopressin BAB40371  (150) KLIQKLLRAKEED
                     Consensus  (151)  LI KLLK KD D
```

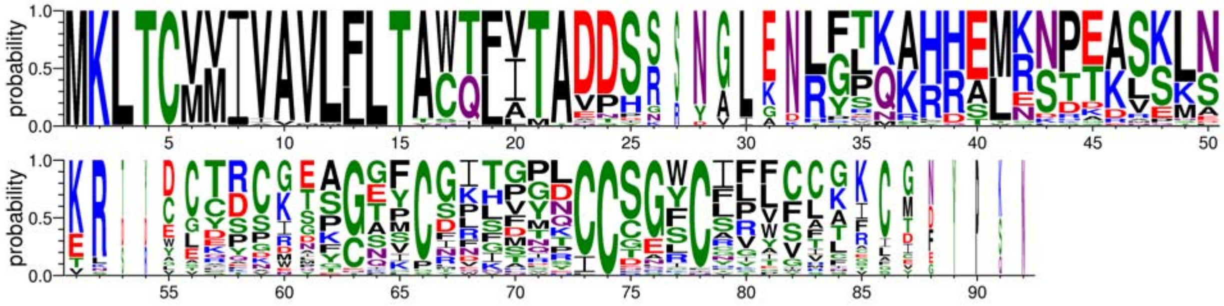


Disulfide Rich Conopeptides

e. O-Superfamily

```
                                          |-------signal-------|------leader peptide--------
                                          1                                               50
       Delta-conotoxin Ac6.1 P0C8V5   (1) MKLTCVVIVAVLFLTAWTFVMADDSR-YGLKDLFPKARHEMKNPEASKLN
       Delta-conotoxin Ac6.2 P0C8V6   (1) MKLTCVVIVAVLFLTAWTFVTADDSR-YGLKNLFPKARHEMKNPEASKLN
        Delta-conotoxin S6.8 B2KJ30   (1) MKLTCMMIVAVLFLTAWTFVTADDSR-NGLKNLFPKARHEMKNPDASKLN
      delta-conotoxin SVIF ABF69254   (1) MKLTCMMIVAVLFLTAWTFVTADDSR-NGLKNLFPKARHEMKNPDASKLN
        Delta-conotoxin SVIE Q9XZK5   (1) MKLTCVMIVAVLFLTTWTFVTADDSR-YGLKNLFPKARHEMKNPEASKLN
           Omega-conotoxin 9 Q5K0D2   (1) MKLTCMMIAAVLFLTTWTFVTADDSR-YGLKNLFPKARHEMKNPEASKLN
        Delta-conotoxin PVIA P58913   (1) MKLTCVMIVAVLFLTAWTFVTADDSK-NGLENHFWKARDEMKNREASKLD
       Delta-conotoxin Ac6.3 P0C8V7   (1) MKLTCVMIVAVLFLTAWTFVTADDSR-NGLENLSPKARHEMKNPEASKSN
       Delta-conotoxin Ai6.1 P0CB09   (1) MKLTCVMIVAVLFLTAWTFATADDPR-NGLGNLFSNAHHEMKNPEASKLN
       Delta-conotoxin TxVIA P18511   (1) MKLTCMMIVAVLFLTAWTFATADDPR-NGLGNLFSNAHHEMKNPEASKLN
       Omega-conotoxin TeA61 Q3YEF3   (1) MKLTCMMIVAVLFLTAWTFATADDSS-NGLGNLFLKAHHEMKNPEASKLN
Omega-conotoxin TxMKLT1-0141 Q9U651   (1) MKLTCMMIVAVLFLTAWTFATADDSS-NGLENLFPKAHHEMKNPEASKLN
Omega-conotoxin TxMKLT1-0142 Q9U650   (1) MKLTCMMIVAVLFLTAWTFATADDSS-NGLENLFLKAHHEMN-PEASKLN
        Omega-conotoxin TxO1 Q9XZK8   (1) MKLTCVVIVAVLFLTVWTFATADDSG-NGLEKLFSNAHHEMKNPEASKLN
        Omega-conotoxin TxO5 Q9XZL2   (1) MKLTCMVIVAVLFLTAWTFVTAITS--NGLENLFPNAHHEMKNPEASKLN
         Omega-conotoxin 2/7 Q5K0B9   (1) MKLTCMMIVAVMFLTASIFITADN-SRNGIENLPRMRRHEMKKPKASKLN
           Omega-conotoxin 3 Q5K0C0   (1) MKLTCMMIVAVLFLTASIFITADN-SRNGIENLPRMRRHEMKKPKASKLN
          Omega-conotoxin 12 Q5K0D6   (1) MKLTCVVIVAVLLLTACQLITADDS-------RGTQKHRSLRSTTKVSKA
        Omega-conotoxin SO-4 Q9XZK3   (1) MKLTCMVIVAVLLLTACQLITADDS-------RGTQKHRSLRSTTKVSKA
           Omega-conotoxin 2 Q5K0D8   (1) MKLTCVVIVAVLLLTACQLITADDS-------RGTQKHRSLRSTTKVSKA
       Omega-conotoxin Ac6.5 P0C8V9   (1) MKLTCVVIVAVLLLTACQLLTADDS-------RGTQKHRSLRSTTKVSKA
           Omega-conotoxin 6 Q5K0D3   (1) MKLTCVVIIAVLLLTACQLITADDS-------RGVQKHRSLRSTTKVSKS
        Omega-conotoxin SO-5 Q9XZK4   (1) MKLTCVMIVAVLLLTACQLITADDS-------RGTQKHRSLRSTTKVSKS
          Omega-conotoxin 15 Q5K0D5   (1) MKLTCVVIVAVLLLTACQLITAEDS-------RGTQKHRTLRSTARRSKS
        Omega-conotoxin SVIA P28880   (1) MKLTCVMIVAVLLLTACQLITAEDS-------RGTQKHRTLRSTARRSKS
       Omega-conotoxin Ac6.4 P0C8V8   (1) MKLTCVVIVAVLLLTACQLLTADDS-------RGTQKHRALRSDTKLSMS
        Omega-conotoxin S6.6 P0C831   (1) MKLTCVVIVAVLLLTACQLLTADDS-------RGTQKHRALRSDTKLSMS
        Omega-conotoxin CVIA P58917   (1) MKLTCVVIVAMLLLTACQLITANDS-------RGTQKHRALRSDTKLSMS
        Omega-conotoxin SVIB P28881   (1) MKLTCVVIVAVLLLTACQLITADDS-------RGTQKHRALRSDTKLPMS
        Omega-conotoxin CVID P58920   (1) MKLTCVVIVAVLLLTACQLITADDS-------RGTQKHRALRSDTKLSMS
        Omega-conotoxin SO-3 Q9XZK2   (1) MKLTCMVIVAVLLLTACQLITADDS-------RGTQKHRTLRSKTKLSMS
        Omega-conotoxin GVIA P01522   (1) MKLTCVVIVAVLLLTACQLITADDS-------RGTQKHRALGSTTELSLS
       Omega-conotoxin PuIIA Q9XYZ1   (1) MKLTCVVIVAVLFLTACQLITAETYS------RGEQKHRALSSTDKNSKL
          Omega-conotoxin 10 Q5K0C4   (1) MKLTCVLIVVVLFLTACQLITTDDS-------TGKQRYQAWKLRSKMQNS
       Omega-conotoxin MVIIA P05484   (1) MKLTCVVIVAVLLLTACQLITADDS-------RGTQKHRALRSTTKLSTS
       Kappa-conotoxin PVIIA P56633   (1) MKLTCVVIVVVLFLTACQLITADDS-------RRTQKHRALRSTTKLSLS
       Omega-conotoxin Ai6.2 P0CB10   (1) MKLTCVMIIAVLFLTAWTFATADDSG-NGLENLFSKAHHEMKNPKASKLN
        Omega-conotoxin TxO6 Q9XZL3   (1) MKLTCVVIVAVLFLTAWTLVMADDSN-NGLANLFSKSRDEMEDPEAAKLE
       Omega-conotoxin Ai6.3 P0CB11   (1) MKLTCLMIVAVLFLTAWTFVTAVPDSSNALENLYLKAHHEMNNPEDSELN
      Omega-conotoxin TeAr94 Q9U647   (1) MKLTCMMIVAVLFLTAWTFVTAVPHSSNALENLYLKAHHEMNNPEDSELN
        Omega-conotoxin TxO2 Q9XZK9   (1) MKLTCVVIVAVLFLTAWTFVTAAPHSSNALENLYLKAHHEMNNPEDSELN
Omega-conotoxin TxMKLT1-0211 Q9U648   (1) MKLTCMMIVAVLFLTAWTFVTAVPHSSNALENLYLKAHHEMNNPEDSELN
        Omega-conotoxin TxO3 Q9XZL0   (1) MKLTCVVIVAVLFLTAWTFVTAVPHSSNALENLYLKAHHEMNNPEASELN
     Omega-conotoxin MaIr137 Q3YED6   (1) MKLTCMMIVAVLFLTAWTFVTAVPHSSDVLENLYLKALHETENHEASKLN
        Omega-conotoxin TxO4 Q9XZL1   (1) MKLTCVVIVAVLFLTAWTFVTAVPHSSNALENLYLKARHEMENPEASKLN
       Omega-conotoxin Vc6.4 P69761   (1) MKLTCVMIVAVLFLTANTFVTAVPHSSNVLENLYLKARHEMENPEASKLN
        Omega-conotoxin PuIA Q9XYZ0   (1) MKLTCVMIVAVLFLTAWTFVTADSIR--ALEDLFAKAPDEMENSGASPLN
        Mu-O-conotoxin MrVIB Q26443   (1) MKLTCMMIVAVLFLTAWTLVMADDSN-NGLANHFLKSRDEMEDPEASKLE
                          Consensus   (1) MKLTCVVIVAVLFLTAWTF TADDS    L NL   AHHEM NP AS LN

                                          -|
                                          51                                      92
       Delta-conotoxin Ac6.1 P0C8V5  (50) KR--DECFS--PGTFCGIKPGLCCSAWCYSFFCLTLTF----
       Delta-conotoxin Ac6.2 P0C8V6  (50) KR--DECYP--PGTFCGIKPGLCCSERCFPFVCLSLEF----
        Delta-conotoxin S6.8 B2KJ30  (50) KR--DGCSN--AGGFCGIHPGLCCSEICLVWCT---------
      delta-conotoxin SVIF ABF69254  (50) KR--DGCSN--AGGFCGIHPGLCCSEICLVWCT---------
        Delta-conotoxin SVIE Q9XZK5  (50) KR--DGCSS--GGTFCGIHPGLCCSEFCFLWCITFID-----
           Omega-conotoxin 9 Q5K0D2  (50) KR--EGCSS--GGTFCGIHPGLCCSEFCFLWCITFID-----
        Delta-conotoxin PVIA P58913  (50) KK--EACYA--PGTFCGIKPGLCCSEFCLPGVCFGG------
       Delta-conotoxin Ac6.3 P0C8V7  (50) KR--YECYS--TGTFCGVNGGLCCSNLCLFFVCLFS------
       Delta-conotoxin Ai6.1 P0CB09  (50) KR--WCKQ---SGEMCNLLDQNCCEGYCIVLVCT--------
       Delta-conotoxin TxVIA P18511  (50) KR--WCKQ---SGEMCNLLDQNCCDGYCIVLVCT--------
       Omega-conotoxin TeA61 Q3YEF3  (50) ER--CLD----AGEVCDIFFPTCCG-YCILLFCA--------
Omega-conotoxin TxMKLT1-0141 Q9U651  (50) ER--CLD----AGEICDFFFPTCCG-YCILLFCA--------
```

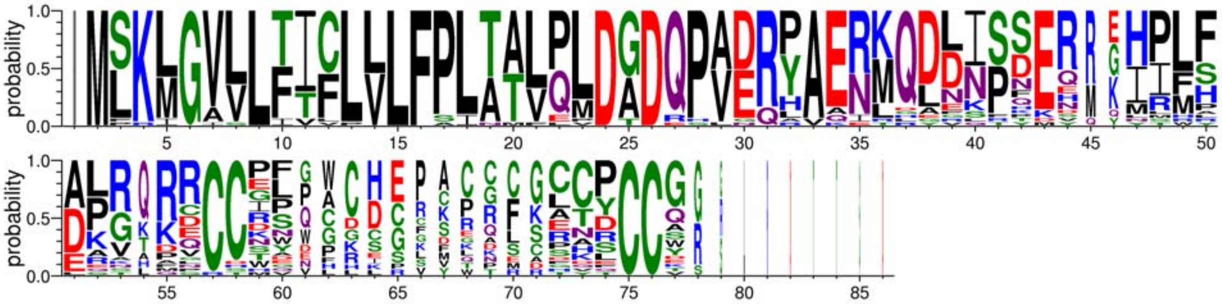
```
Omega-conotoxin TxMKLT1-0142 Q9U650  (49) ER--CLD----AGEICDFFFPTCCG-YCILLFCA--------
        Omega-conotoxin TxO1 Q9XZK8  (50) ER--CLD----AGEVCDIFFPTCCG-YCILLFCA--------
        Omega-conotoxin TxO5 Q9XZL2  (49) KR--CVP----YEGPCNWLTQNCCDATCVVFWCL--------
         Omega-conotoxin 2/7 Q5K0B9  (50) KR--VCIADDMPCGFGLFGGPLCCSGWCLFVCL---------
           Omega-conotoxin 3 Q5K0C0  (50) KR--GCLPDEYFCGFSMIGALLCCSGWCLGICMT--------
          Omega-conotoxin 12 Q5K0D6  (44) ------TDCIEAGNYCGPTVMKICCGFCSPYS-KICMNYPKN
        Omega-conotoxin SO-4 Q9XZK3  (44) ------TDCIEAGNYCGPTVMKICCGFCSPYS-KICMNYPKN
           Omega-conotoxin 2 Q5K0D8  (44) ------ADCIEAGNYCGPTVMKLCCGFCSPYS-KICMNYPKN
       Omega-conotoxin Ac6.5 P0C8V9  (44) ------TDCIEAGNYCGPTVMKICCGFCSPFS-KICMNYPQN
           Omega-conotoxin 6 Q5K0D3  (44) ------TSCMEAGSYCG-STTRICCGYCAYFG-KKCIDYPSN
        Omega-conotoxin SO-5 Q9XZK4  (44) ------TSCMEAGSYCG-STTRICCGYCAYFG-KKCIDYPSN
          Omega-conotoxin 15 Q5K0D5  (44) ELT---TRCRPSGSPCG--VTSICCGRCSR---GKCT-----
        Omega-conotoxin SVIA P28880  (44) EST---TRCRSSGSPCG--VTSICCGRCYR---GKCT-----
       Omega-conotoxin Ac6.4 P0C8V8  (44) ------TRCKGKGASCSRTMYNCCTGSCNR---GKCG-----
        Omega-conotoxin S6.6 P0C831  (44) ------TRCKGKGAPCRKTMYDCCSGSCGR--RGKCG-----
        Omega-conotoxin CVIA P58917  (44) ------TRCKSTGASCRRTSYDCCTGSCR--S-GRCG-----
        Omega-conotoxin SVIB P28881  (44) ------TRCKLKGQSCRKTSYDCCSGSCGR--SGKCG-----
        Omega-conotoxin CVID P58920  (44) ------TRCKSKGAKCSKLMYDCCSGSCSGTV-GRCG-----
        Omega-conotoxin SO-3 Q9XZK2  (44) ------TRCKAAGKPCSRIAYNCCTGSCR--S-GKCG-----
        Omega-conotoxin GVIA P01522  (44) ------TRCKSPGSSCSPTSYNCCRSCNP--YTKRCYG----
       Omega-conotoxin PuIIA Q9XYZ1  (45) T-----RTCNTPTQYCT-LHRHCCSLYCHKTIHACA------
          Omega-conotoxin 10 Q5K0C4  (44) VLSRLSKRCDEEGTGCS-SDSECCSGRCTPEGLFEFCE----
       Omega-conotoxin MVIIA P05484  (44) ------TRCKGKGAKCSRLMYDCCTGSCR--S-GKCG-----
       Kappa-conotoxin PVIIA P56633  (44) ------TRCRIPNQKCFQHLDDCCSRKCNR--FNKCV-----
       Omega-conotoxin Ai6.2 P0CB10  (50) KR--CTQ----SGELCDVIDPDCCNNFCIIFFCI--------
        Omega-conotoxin TxO6 Q9XZL3  (50) KN--YCQEKWDYCPVPFLGSRYCCDGLFCTLFFCA-------
       Omega-conotoxin Ai6.3 P0CB11  (51) KR---CYDGGTSC----NTGNQCCSGWCIFLCL---------
      Omega-conotoxin TeAr94 Q9U647  (51) KR---CYDSGTSC----NTGNQCCSGWCIFVCL---------
        Omega-conotoxin TxO2 Q9XZK9  (51) KR---CYDSGTSC----NTGNQCCSGWCIFVCL---------
Omega-conotoxin TxMKLT1-0211 Q9U648  (51) KR---CYDGGTSC----DSGIQCCSGWCIFVCL---------
        Omega-conotoxin TxO3 Q9XZL0  (51) KR---CYDGGTSC----DSGIQCCSGWCIFVCF---------
     Omega-conotoxin MaIr137 Q3YED6  (51) VRDDECEPPGDFCGF-FKIGPPCCSGWCFLWCA---------
        Omega-conotoxin TxO4 Q9XZL1  (51) TR-YDCEPPGNFCGM-IKIGPPCCSGWCFFACA---------
       Omega-conotoxin Vc6.4 P69761  (51) TR-YDCEPPGNFCGM-IKVGPPCCSGWCFFACA---------
        Omega-conotoxin PuIA Q9XYZ0  (49) ER--DCRPVGQYCGIPYEHNWRCCSQLCAIICVS--------
        Mu-O-conotoxin MrVIB Q26443  (50) KR--ACSKKWEYCIVPILGFVYCCPGLICGPFVCV-------
                          Consensus  (51)  R          G  C      CCSG C
```



The leader peptides of this family are less conserved, possible because a large number of members are known, or reflecting the hypervariable nature of the core peptide. The Arg in the P1 position with respect to leader peptide removal (position 52 in the logo) is highly but not completely conserved.

f. M-Superfamily

```
                                                  |--------signal---------|-----leader peptide------
                                                  1                                               50
             Kappa-M-conotoxin RIIIK P69769   (1) -MSKLGVLLTICLLLFPLTALPMDGDQPVDRLAERMQDNISSEQ--HTFF
                  Mu-conotoxin GIIIA P01523   (1) -MSKLGVLLTICLLLFPLTALPMDGDEPANRPVERMQDNISSEQ--YPLF
                  Mu-conotoxin PIIIA P58925   (1) -MSKLGVLLTICLLLFPITALPMDGDQPADRLAERMQDNISSEE--HP-F
                 Psi-conotoxin PIIIE P56529   (1) -MSKLGALLTICLLLFPITALLMDGDQPADRPAERMDYDISSEV--HRLL
                Psi-conotoxin PrIIIE P0C7I1   (1) -MSKLGVLLTICLLLFPITALPVDGDQPADRPVERMQDNISSEQ--HPFF
         M-2-superfamily conotoxin 3 Q9BPJ7   (1) -MSKLGALLIICLLLFPLTAVPMDGDQPADRPAERMQDDISFEQ--HPMF
      M-2-superfamily conotoxin Tx3b P0C1N8   (1) -MSKLGALLTICLLLFSLTAVPLDGDQHADQPAQRLQDRIPTED--HPLF
 M-2-superfamily conotoxin TxMMSK-01 Q9BPJ1   (1) MMSKLGVLLITCLLLFPLTAVPLDGDQPADQPAERLQDDISSEN--HPFF
     M-2-superfamily conotoxin Lt3.5 Q2I2Q1   (1) MMSKLGALLTICLLLFPLTAVPLDGDQPLDRHAERMHDGISPKR--HPWF
     M-2-superfamily conotoxin Mr3.3 P0C1N2   (1) -MSRLGVLLTICLLLFPLTAVPLDGDQPADRPAERLQDDISSEH--HPHF
     M-2-superfamily conotoxin Mr3.4 P0C1N3   (1) -MSKLGVLLTICLLLFPLTAVPLDGDQPADRPAERMQDDISSER--HPFF
      M-2-superfamily conotoxin Mr3d P0C1N1   (1) -MSKLGILLTICLLLFPLTAVPLDGDQPADRPAERMQDDISSEH--HPFF
      M-2-superfamily conotoxin Mr3g P0C1N5   (1) -MSKLGVLLTICLLLFALTAVPLDGDQPADRPAERMQDDISSER--HPMF
                  Mu-conotoxin SIIIA Q86DU6   (1) MMSKLGVLLTVCPLLFPLTALPPDGDQPADRPAERMQDDISSDE--HPLF
   M-3-superfamily conotoxin Tx3.5-a P0C1N7   (1) -MSKLGVLLTICLLLFPLTALPLDGDQPADQAAERMQAEQ------HPLF
     M-3-superfamily conotoxin Mr3.5 P0C1N4   (1) -MSKLGVLLTICLLLFPLTALPLDGDQPADQRAERTQAEK------HSLP
      M-2-superfamily conotoxin Tx3c P58846   (1) -MFKLGVLLTICLLLFSLNAVPLDGDQPADQPAERLLDDISFEN--NPFY
         M-1-superfamily conotoxin 2 Q9BPI0   (1) -MLKMGVVLFIFLVLFPLATLQLDADQPVERYAENKQLLNTDERREIILS
     M-1-superfamily conotoxin Lt3.1 Q2I2Q5   (1) -MLKMGVLLFTFLVLFPLTTLELDTDRPVERHAAIKQDLKPQERRGIRLH
     M-1-superfamily conotoxin Lt3.4 Q2I2Q2   (1) -MLKMGVLLFTFLVLFPLAMFQLDADQPVERYAENKQDLNRDERMKIMLS
    M-1-superfamily conotoxin Lt16.1 Q2I2P8   (1) -MPKLGVSLFIFLVLFPLATLQLDGDQSAGRHAQERGEDLFKMYQYLRRA
     M-1-superfamily conotoxin Lp3.1 Q5I2P0   (1) -MLKMGVLLFIFLVLFPLTTLELDTDRPVERHAAIKQDLKPQERRGIRLH
     M-1-superfamily conotoxin Lp3.2 Q5EHP4   (1) -MLKMGVVLFTFLVLFPLATLQLDADQPVERYAENKQDLNPNERMKMIMS
 M-3-superfamily conotoxin TsMLKM-02 Q9BPH4   (1) -MLKMGVVLFTFLVLFPLATLQLDADQPVERYAENKQGLNPDESREIILS
M-3-superfamily conotoxin TsMLKM-011 Q9BH51   (1) -MLKMGVVLFVFLVLFPLATLQLDADQPVERYAENKQLVSPYERRQIILH
     M-1-superfamily conotoxin Tx3.1 P0C1N6   (1) -MLKMGVVLFIFLVLFPLATLQLDADQPVERYAENKQLLNPDERREILLP
    M-1-superfamily conotoxin Tx3a-a Q9BH73   (1) -MLKMGVVLFIFLVLFPLATLQLDADQPVERYAENKQLLSPDERREIILH
      M-1-superfamily conotoxin Mr3e Q5EHP3   (1) -MLKMGVVLFIVLVLFPLATLQLDADQPVERYAENKRLLNPDERRGIILH
     M-1-superfamily conotoxin Mr3.8 Q5EHP2   (1) -MLKMGVVLFIFLVLFPLATLQLDADQPVERYAKNKQLFNPHKRRGIILR
     M-1-superfamily conotoxin Lt3.2 Q2I2Q4   (1) -MLKIGVVLFTFLVLFPLATLQLDADQPVERYAENKQDLNPNERMKMIMS
     M-1-superfamily conotoxin Lt3.3 Q2I2Q3   (1) -MLKIGVVLFTLLVLFPLATLQLDADQPVERYAENKQDLNPNERMKMIMS
                                  Consensus   (1)  MSKLGVLL ICLLLFPLTALPLDGDQPADR AERMQD I  E   H L

                                                  -----|
                                                 51                               86
             Kappa-M-conotoxin RIIIK P69769  (48) EKRLPSCCSLNLRLCPVPACKRNPCCTG--------
                  Mu-conotoxin GIIIA P01523  (48) EKR-RDCCTP-PKKCKDRQCKPQRCCAGR-------
                  Mu-conotoxin PIIIA P58925  (47) EKRQRLCCGF-PKSCRSRQCKPHRCCGR--------
                 Psi-conotoxin PIIIE P56529  (48) ERRHPPCCMY-GRCRRYPGCSSASCCQGG-------
                Psi-conotoxin PrIIIE P0C7I1  (48) EKRAARCCTY-HGSCLKEKCRRKYCCGR--------
         M-2-superfamily conotoxin 3 Q9BPJ7  (48) DAT-RRCCNA-----GFCRFGCTPCCY---------
      M-2-superfamily conotoxin Tx3b P0C1N8  (48) DPN-KRCCPP-----VACNMGCKPCCG---------
 M-2-superfamily conotoxin TxMMSK-01 Q9BPJ1  (49) DPV-KRCCRL-------LCLSCNPCCG---------
     M-2-superfamily conotoxin Lt3.5 Q2I2Q1  (49) DPV-KRCCKV-------QCESCTPCC----------
     M-2-superfamily conotoxin Mr3.3 P0C1N2  (48) DSG-RECCGS-----FACRFGCVPCCV---------
     M-2-superfamily conotoxin Mr3.4 P0C1N3  (48) DRS-KQCCHL-----PACRFGCTPCCW---------
      M-2-superfamily conotoxin Mr3d P0C1N1  (48) DPV-KRCCRL------SCGLGCHPCCG---------
      M-2-superfamily conotoxin Mr3g P0C1N5  (48) DAV-RDCCPL-----PACPFGCNPCCG---------
                  Mu-conotoxin SIIIA Q86DU6  (49) DKR-QNCCNG--GCSSKWCRDHARCCGR--------
   M-3-superfamily conotoxin Tx3.5-a P0C1N7  (44) DQK-RRCCKF-----PCPDSCRYLCCG---------
     M-3-superfamily conotoxin Mr3.5 P0C1N4  (44) DPR-MGCCPF-----PCKTSCTTLCCG---------
      M-2-superfamily conotoxin Tx3c P58846  (48) DPA-KRCCR--------TCFGCTPCCG---------
         M-1-superfamily conotoxin 2 Q9BPI0  (50) ALRTRVCCPFGGCHE------LCQCCEG--------
     M-1-superfamily conotoxin Lt3.1 Q2I2Q5  (50) APR-DECCEPQWCDG------ACDCCS---------
     M-1-superfamily conotoxin Lt3.4 Q2I2Q2  (50) ALRQRQCCDWEWCDE------LCSCCW---------
    M-1-superfamily conotoxin Lt16.1 Q2I2P8  (50) LER--RRTGEDFLEECMGGCAFDFCCKRSLRDTTSD
     M-1-superfamily conotoxin Lp3.1 Q5I2P0  (50) APR-DECCEPQWCDG------ACDCCS---------
     M-1-superfamily conotoxin Lp3.2 Q5EHP4  (50) ALGQRRCCISPACHE------ECYCCQ---------
 M-3-superfamily conotoxin TsMLKM-02 Q9BPH4  (50) ALRQRDCCEQGWCDG------GCDCCQ---------
M-3-superfamily conotoxin TsMLKM-011 Q9BH51  (50) ALGQRQCCDWQWCDG------ACDCCA---------
     M-1-superfamily conotoxin Tx3.1 P0C1N6  (50) ALRKFCCDSNWCHIS------DCECCYG--------
    M-1-superfamily conotoxin Tx3a-a Q9BH73  (50) ALGTRCCSWDVCDHP------SCTCCG---------
      M-1-superfamily conotoxin Mr3e Q5EHP3  (50) ALGQRVCCPFGGCHE------LCYCCDG--------
     M-1-superfamily conotoxin Mr3.8 Q5EHP2  (50) APGKRCCHWNWCDH-------LCSCCGS--------
     M-1-superfamily conotoxin Lt3.2 Q2I2Q4  (50) ALGQRRCCISPACHE------ECYCCQ---------
     M-1-superfamily conotoxin Lt3.3 Q2I2Q3  (50) ALGQRRCCISPACHE------ECYCC----------
                                  Consensus  (51) D   R CC                CC
```

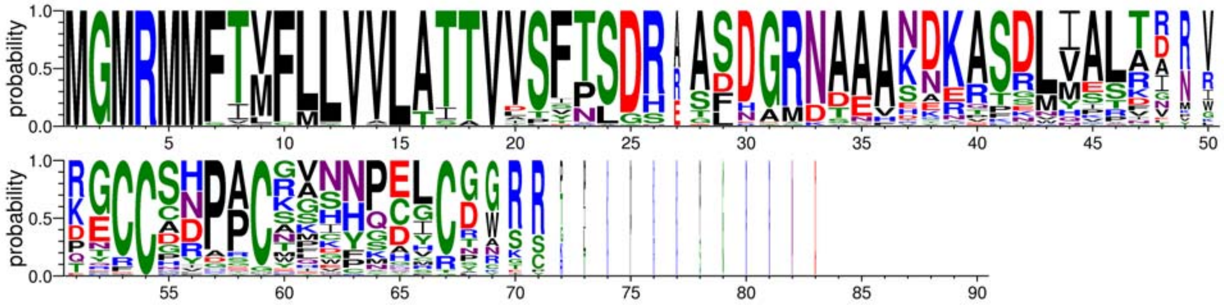


This family has moderately conserved leader peptides including a highly conserved Pro at position 28. Several Asp/Glu residues are also highly conserved. A subset of family members also contain a (Lys/Arg)Arg protease cleavage site to remove the leader peptide (position 55-56, see individual alignments) as well as several prolines at other positions (see alignments).

g. A-Superfamily

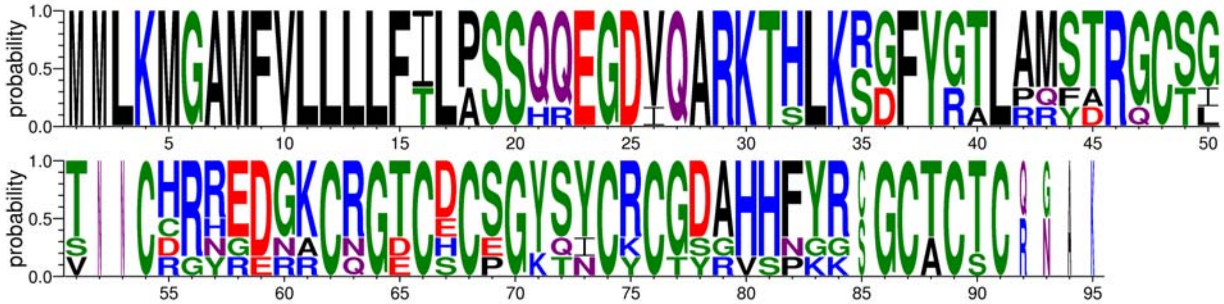
```
                                           |-------signal------|--------leader peptide------|
                                           1                                               50
      alpha-A-conotoxin Tx1 AAD31912   (1) MGMRMMFVVFLLVVLASTVVSSTSGRRAFHGRNAAA--KASGLVSLTDRR
      alpha-A-conotoxin Tx2 AAD31913   (1) MGMRMMFTVFLLVVLATTVVSFTSGRRTFHGRNAAA--KASGLVSLTDRR
     alpha-conotoxin Ac1.1a ABD33847   (1) MGMRMMFTVFLLVVLTTTVVSYPSDSASD-GRDDEAKDERSDMYESKRN-
     alpha-conotoxin Ac1.1a ABD48789   (1) MGMRMMFTLFLLVVLTTTVVSYPSDSASD-GRDDEAKDERSDMYELKRN-
       Alpha-conotoxin Ac1.1a P0CAQ4   (1) MGMRMMFTLFLLVVLTTTVVSYPSDSASD-GRDDEAKDERSDMYELKRN-
       Alpha-conotoxin Ac1.1b P0CAQ5   (1) MGMRMMFTLFLLVVLTTTVVSFPSDSASD-GRDDEAKDERSDMYKSKRN-
  Alpha-conotoxin Bn1.1/Bn1.2 P0C1Y1   (1) MGMRMMFTMFLLVVLATTVVSFASDR-ASDGRNAAAKDKASDLVALT--V
        Alpha-conotoxin Bn1.3 P0C1Y2   (1) MGMRMMFTVFLLVVLATAVLPVTLDR-ASDGRNAAANAKTPRLIAPFIR-
        Alpha-conotoxin Ca1.1 P0C1V9   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-ASDGRNAAA--NAFDLIALIAR-
        Alpha-conotoxin Ca1.2 P0C1W0   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-ASEGRNAAAKDKASDLVALT--V
        Alpha-conotoxin ImIIA Q9U619   (1) MGMRMMFTVFLLVVLATAVLPVTLDR-ASDGRNAAANAKTPRLIAPFIR-
      alpha-conotoxin Lp1.1 ABC39769   (1) MGMRMMFIMFMLVVLATTVDTFTSDH-ALDAMNAAASNKASRLIALA--V
        Alpha-conotoxin Lp1.1 Q6PTD5   (1) MGMRMMFIMFMLVVLATTVVTFTSDR-ALDAMNAAASNKASRLIALA--V
      alpha-conotoxin lp1.2 AAS93425   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-AFDGRNAAASDKASDLISLA--V
      alpha-conotoxin lp1.3 AAS93426   (1) MGMRMMFIMFMLVVLATTVVSFTSDR-ASDGRNAEA--KALDLIAATAR-
        Alpha-conotoxin Lp1.4 A1X8B6   (1) MGMRMMSIMFMLVVLATTVVSFTSDR-ALDAMNAAASKKASRLIALA--V
      alpha-conotoxin lp1.4 AAS93427   (1) MGMRMMFIMFMLVVLATTVVSFTSDR-ALDAMNAAASKKASRLIALA--V
      alpha-conotoxin Lp1.4 ABD33848   (1) MGMRMMSIMFMLVVLATTVVSFTSDR-ALDAMNAAASKKASRLIALA--V
      alpha-conotoxin Lp1.5 AAS99934   (1) MGMRMMFTVFLLVILASTVTLDRASN----GMNAAAIRKASALVAQIAYR
       Alpha-conotoxin Lp1.6a P0CAQ6   (1) MGMRMMFIIFLFVVLATTVVSFTSGR-ASDGRNAPANNKVSDLIR-----
        Alpha-conotoxin Lp1.7 A1X8C2   (1) MGMRMMFTMFLLVVLTTTVVSFNSDRE------SNHENRRTSNQITRGVW
        Alpha-conotoxin Lp1.8 A1X8C3   (1) MGMRMMFTMFLLVVLTTTVVSFNSDRE------SNHENRRTSNQITRGMW
        Alpha-conotoxin Lt1.1 Q2I2R8   (1) MGMRMMFIMFMLVVLATTVVTFTSDR-ALDAMNAAASNKASRLIALA--V
        Alpha-conotoxin Lt1.2 Q2I2R7   (1) MGMRMMFIMFMLVVLATTVDTFTSDR-ALDAMNAAASNKASRLIALA--V
        Alpha-conotoxin Lt1.3 Q2I2R6   (1) MGMRMMFTMFLLVVLTTTVVSFNLDRE------SNHENRRTSNQITRGMW
      alpha-conotoxin Mi1.1 AAS99935   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-GSDGRNAAAKDKASDLVALT--V
          Alpha-conotoxin MII P56636   (1) MGMRMMFTVFLLVVLATTVVSFPSDR-ASDGRNAAANDKASDVITLA--L
      alpha-conotoxin Mr1.1 AAS93428   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-ASDGRKAAAKDKASDLVALT--V
        Alpha-conotoxin Mr1.1 Q6PTD1   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-ASDGRKAAAKDKASDLVALT--V
        Alpha-conotoxin Mr1.2 A6M934   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-GSDGRNAAAKDKASDLVALT--V
      alpha-conotoxin Mr1.2 ABC39767   (1) MGMRMMFTMCLLVVLATTVISFTSDR-ASNGRNAAAKDKASDLNALN--V
        Alpha-conotoxin Mr1.3 A0SE59   (1) MGMRMMFTMCLLVVLATTVISFTSDR-ASNGRNAAAKDKASDLNALN--V
         Alpha-conotoxin PnIB P50985   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-ASDDG----NAAASDLIALT--I
        Alpha-conotoxin Pu1.1 A0SE60   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-TSDGRNAAF--NAFDLIALTAR-
     Alpha-conotoxin Pu14.1 ACL13206   (1) MGMRMMFAVFLLVVLATTVVSFNSDR-ASDGRNAAANVKASDLMARVL--
     Alpha-conotoxin Pu14.2 ACL13207   (1) MGMRMMFTVFLLVVLATTVDSFNSDRASD-GRDAEVVSTESDVIVTCEP-
  alpha-conotoxin Qc alphaL-2 A1X8B9   (1) MGMRMMFTMFLLVVLATTVVSFNLDH-AFDGRNAAANNKATDLMARTVR-
      alpha-conotoxin qc1.1 AAS93422   (1) MGMRMMFTMFLLVVLAITVVSFTSDH-ASDGRNTAANDKASNLMALRD--
     alpha-conotoxin Qc1.1a ABD33852   (1) MGMRMMFTMFLLVVLAITVVSFTSDH-ASDGRNTAANDKASNLMALRD--
       Alpha-conotoxin Qc1.1a Q6PTD7   (1) MGMRMMFTMFLLVVLAITVVSFTSDH-ASDGRNTAANDKASNLMALRD--
     alpha-conotoxin Qc1.1b ABD33853   (1) MGMRMMFTMFLLVVLAITVVSFTSDH-ASDGRNTAANDKASKLMALRN--
       Alpha-conotoxin Qc1.1b Q6PPB3   (1) MGMRMMFTMFLLVVLAITVVSFTSDH-ASDGRNTAANDKASKLMALRN--
      alpha-conotoxin qc1.2 AAS93423   (1) MGMRMMFTVFLLVALATTVASFTLDR-ASNGRNAAADDKPSDWIALA--I
        Alpha-conotoxin Qc1.2 Q6PTD6   (1) MGMRMMFTVFLLVALATTVASFTLDR-ASNGRNAAADDKPSDWIALA--I
      alpha-conotoxin Qc1.3 AAS99932   (1) MGMRMMFTMFLLVVLAITVVSFTSDH-ASDGRNTAANDKASKLMALRN--
     alpha-conotoxin QcaL-1 ABD33850   (1) MGMRMMFTMFLLVVLATTVVSINLDH-AFDGRNAAANNKATDLMARTVR-
     alpha-conotoxin Qcal-2 ABD33851   (1) MGMRMMFTMFLLVVLATTVVSFNLDH-AFDGRNAAANNKATDLMARTVR-
           Alpha-conotoxin S1 P15471   (1) MGMRMMFTVFLLVVLATTVVSFPSDRASD-GRDDEAKDERSDMHESDR--
         Alpha-conotoxin S1.1 P0C1W4   (1) MGMRMMFTVFLLVVLAITVVSFPLDRESD-GANAEARTHDHEKHALDRN-
           Alpha-conotoxin S2 P28879   (1) MGMRMMFTVFLLVVLATTVVSFPSDRASD-GRDDEAKDERSDMHESDRNG
        alpha-conotoxin SII AAN77902   (1) MGMRMMFTVFLLVVLATTVVSFPSDRASD-GRDDEAKDERSDMHESDRNG
    Alpha-conotoxin SrIA/SrIB P85886   (1) MGMRMMFTVFLLVVLATTVVSFTSDS-AFDSRNVAANDKVSDMIALTAR-
          Alpha-conotoxin Tx1 Q9XZK6   (1) MGMRMMFVVFLLVVLASTVVSSTSGRRAFHGRNAAA--KASGLVSLTDRR
          Alpha-conotoxin Tx2 Q9XZK7   (1) MGMRMMFTVFLLVVLATTVVSFTSGRRTFHGRNAAA--KASGLVSLTDRR
         Alpha-conotoxin Vc1A P69747   (1) MGMRMMFTVFLLVVLATTVVSSTSGRREFRGRNAAA--KASDLVSLTDKK
         conotoxin  Act1.1b ABD48790   (1) MGMRMMFTLFLLVVLTTTVVSFPSDSASD-GRDDEAKDERSDMYKSKRN-
            conotoxin ImIIA AAF12824   (1) MGMRMMFTVFLLVVLATAVLPVTLDR-ASDGRNAAANAKTPRLIAPFIR-
            conotoxin Mr1.2 ABD48791   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-GSDGRNAAAKDKASDLVALT--V
       Kappa-A-conotoxin MIVA P0C1X1   (1) MGMRMMFTVFLLVVLATTVVSIPSDR-ASDGRNAVVHERAPELVVTAT--
       Kappa-A-conotoxin S4.2 P0C829   (1) MGMRMMFTVFLSVVLATTVVSTPSDR-ASDGRNAAVHERQKELVPSVI--
       Kappa-A-conotoxin SIVA P0C828   (1) MGMRMMFTVFLLVVLATTVVSTPSDR-ASDGRNAAVHERQKSLVPSVI--
Kappa-A-conotoxin SmIVA/SmIVB P0C1Y3   (1) MGMRMMFTVFLLVVLATTVVSIPSDR-ASDGRNAAVNERQTWLVPSTI--
                      Pu1.1 ABC39768   (1) MGMRMMFTVFLLVVLATTVVSFTSDR-TSDGRNAAF--NAFDLIALTAR-
                           Consensus   (1) MGMRMMFTVFLLVVLATTVVSFTSDR ASDGRNAAA  KASDL AL
```

```
                                           51                             83
      alpha-A-conotoxin Tx1 AAD31912  (49) PECCSDPRCNSSHPELCGGRR------------
      alpha-A-conotoxin Tx2 AAD31913  (49) PECCSHPACNVDHPEICR---------------
     alpha-conotoxin Ac1.1a ABD33847  (49) -GRCCHPACGKHFSCGR----------------
     alpha-conotoxin Ac1.1a ABD48789  (49) -GRCCHPACGKHFNCGR----------------
       Alpha-conotoxin Ac1.1a P0CAQ4  (49) -GRCCHPACGKHFNCGR----------------
       Alpha-conotoxin Ac1.1b P0CAQ5  (49) -GRCCHPACGKHFSCGR----------------
  Alpha-conotoxin Bn1.1/Bn1.2 P0C1Y1  (48) KGCCSHPACSVNNPDICG---------------
        Alpha-conotoxin Bn1.3 P0C1Y2  (49) DYCCHRGPCMVWCG-------------------
        Alpha-conotoxin Ca1.1 P0C1V9  (47) QNCCSIPSCWEKYKCS-----------------
        Alpha-conotoxin Ca1.2 P0C1W0  (48) RGCCAIRECRLQNAAYCGGIY------------
        Alpha-conotoxin ImIIA Q9U619  (49) DYCCHRGPCMVWCG-------------------
      alpha-conotoxin Lp1.1 ABC39769  (48) RGCCARAACAGIHQELCGGRR------------
        Alpha-conotoxin Lp1.1 Q6PTD5  (48) RGCCARAACAGIHQELCGGGR------------
      alpha-conotoxin lp1.2 AAS93425  (48) RGCCSHPACSVNNPYFCGGKR------------
      alpha-conotoxin lp1.3 AAS93426  (47) PRCCSNPACGAGHPEICAGRR------------
        Alpha-conotoxin Lp1.4 A1X8B6  (48) RGCCSHPACSGNHQELCDGRR------------
      alpha-conotoxin lp1.4 AAS93427  (48) RGCCSHPACSGNHQELCDGRR------------
      alpha-conotoxin Lp1.4 ABD33848  (48) RGCCSHPACSGNHQELCDGRR------------
      alpha-conotoxin Lp1.5 AAS99934  (47) -DCCDDPACTVNNPGLCT---------------
       Alpha-conotoxin Lp1.6a P0CAQ6  (45) QFCCGHYDCDFIPNVCG----------------
        Alpha-conotoxin Lp1.7 A1X8C2  (45) DECCKDPQCRQNHMQHCPAR-------------
        Alpha-conotoxin Lp1.8 A1X8C3  (45) DECCDDPPCRQNNMEHCPAS-------------
        Alpha-conotoxin Lt1.1 Q2I2R8  (48) RGCCARAACAGIHQELCGGGR------------
        Alpha-conotoxin Lt1.2 Q2I2R7  (48) RGCCARAACAGIHQELCGGRR------------
        Alpha-conotoxin Lt1.3 Q2I2R6  (45) DECCDDPPCRQNNMEHCPAS-------------
      alpha-conotoxin Mi1.1 AAS99935  (48) KGCCSNPPCYANNQAYCNGRR------------
          Alpha-conotoxin MII P56636  (48) KGCCSNPVCHLEHSNLCGRRR------------
      alpha-conotoxin Mr1.1 AAS93428  (48) KGCCSHPACSVNNPDICG---------------
        Alpha-conotoxin Mr1.1 Q6PTD1  (48) KGCCSHPACSVNNPDICG---------------
        Alpha-conotoxin Mr1.2 A6M934  (48) KGCCSNPPCYANNQAYCNGRR------------
      alpha-conotoxin Mr1.2 ABC39767  (48) RGCCSHPACRVHYPHVCYGRR------------
        Alpha-conotoxin Mr1.3 A0SE59  (48) RGCCSHPACRVHYPHVCYGRR------------
         Alpha-conotoxin PnIB P50985  (44) KGCCSLPPCALSNPDYCG---------------
        Alpha-conotoxin Pu1.1 A0SE60  (47) QNCCNVPGCWAKYKHLCGRKR------------
     Alpha-conotoxin Pu14.1 ACL13206  (48) EKDCPPHPVPGMHKCVCLKTCR-----------
     Alpha-conotoxin Pu14.2 ACL13207  (49) ---CMNPACGPNYGKCR----------------
  alpha-conotoxin Qc alphaL-2 A1X8B9  (49) -GFCSDPSCRFGNPELCDWRR------------
      alpha-conotoxin qc1.1 AAS93422  (48) -ECCPDPPCKASNPDLCDWRS------------
     alpha-conotoxin Qc1.1a ABD33852  (48) -ECCPDPPCKASNPDLCDWRS------------
       Alpha-conotoxin Qc1.1a Q6PTD7  (48) -ECCPDPPCKASNPDLCDWRS------------
     alpha-conotoxin Qc1.1b ABD33853  (48) -ECCDNPPCKSSNPDLCDWRS------------
       Alpha-conotoxin Qc1.1b Q6PPB3  (48) -ECCDNPPCKSSNPDLCDWRS------------
      alpha-conotoxin qc1.2 AAS93423  (48) KQCCANPPCKHVN---CR---------------
        Alpha-conotoxin Qc1.2 Q6PTD6  (48) KQCCANPPCKHVN---CR---------------
      alpha-conotoxin Qc1.3 AAS99932  (48) -ECCDNPPCKSSNPDLCDWRS------------
     alpha-conotoxin QcaL-1 ABD33850  (49) -RFCSDPPCRISNPESCGWEP------------
     alpha-conotoxin Qcal-2 ABD33851  (49) -GFCSDPSCRFGNPELCDWRR------------
           Alpha-conotoxin S1 P15471  (48) KEICCNPACGPKYSCGR----------------
         Alpha-conotoxin S1.1 P0C1W4  (49) -GCCRNPACESHRCG------------------
           Alpha-conotoxin S2 P28879  (50) RGCCCNPACGPNYGCGTSCSRTL----------
        alpha-conotoxin SII AAN77902  (50) RGCCCNPACGPNYGCGTSCSRTI----------
    Alpha-conotoxin SrIA/SrIB P85886  (49) RTCCSRPTCRMEYPELCGGRR------------
          Alpha-conotoxin Tx1 Q9XZK6  (49) PECCSDPRCNSSHPELCGGRR------------
          Alpha-conotoxin Tx2 Q9XZK7  (49) PECCSHPACNVDHPEICR---------------
         Alpha-conotoxin Vc1A P69747  (49) RGCCSDPRCNYDHPEICG---------------
         conotoxin  Act1.1b ABD48790  (49) -GRCCHPACGKHFSCGR----------------
            conotoxin ImIIA AAF12824  (49) DYCCHRGPCMVWCG-------------------
            conotoxin Mr1.2 ABD48791  (48) KGCCSNPPCYANNQAYCNGRR------------
       Kappa-A-conotoxin MIVA P0C1X1  (48) TNCCGYNPMTICPPCMCTYSCPPK-RKPGRRND
       Kappa-A-conotoxin S4.2 P0C829  (48) TTCCGYDPGTMCPPCRCTNSCPTKPKKPGRRND
       Kappa-A-conotoxin SIVA P0C828  (48) TTCCGYDPGTMCPPCRCTNSCG-----------
Kappa-A-conotoxin SmIVA/SmIVB P0C1Y3  (48) TTCCGYDPGTMCPTCMCDNTCKPKPKKSGRRND
                      Pu1.1 ABC39768  (47) QNCCNVPGCWAKYKHLCGRKR------------
                           Consensus  (51)   CC  P C       C
```



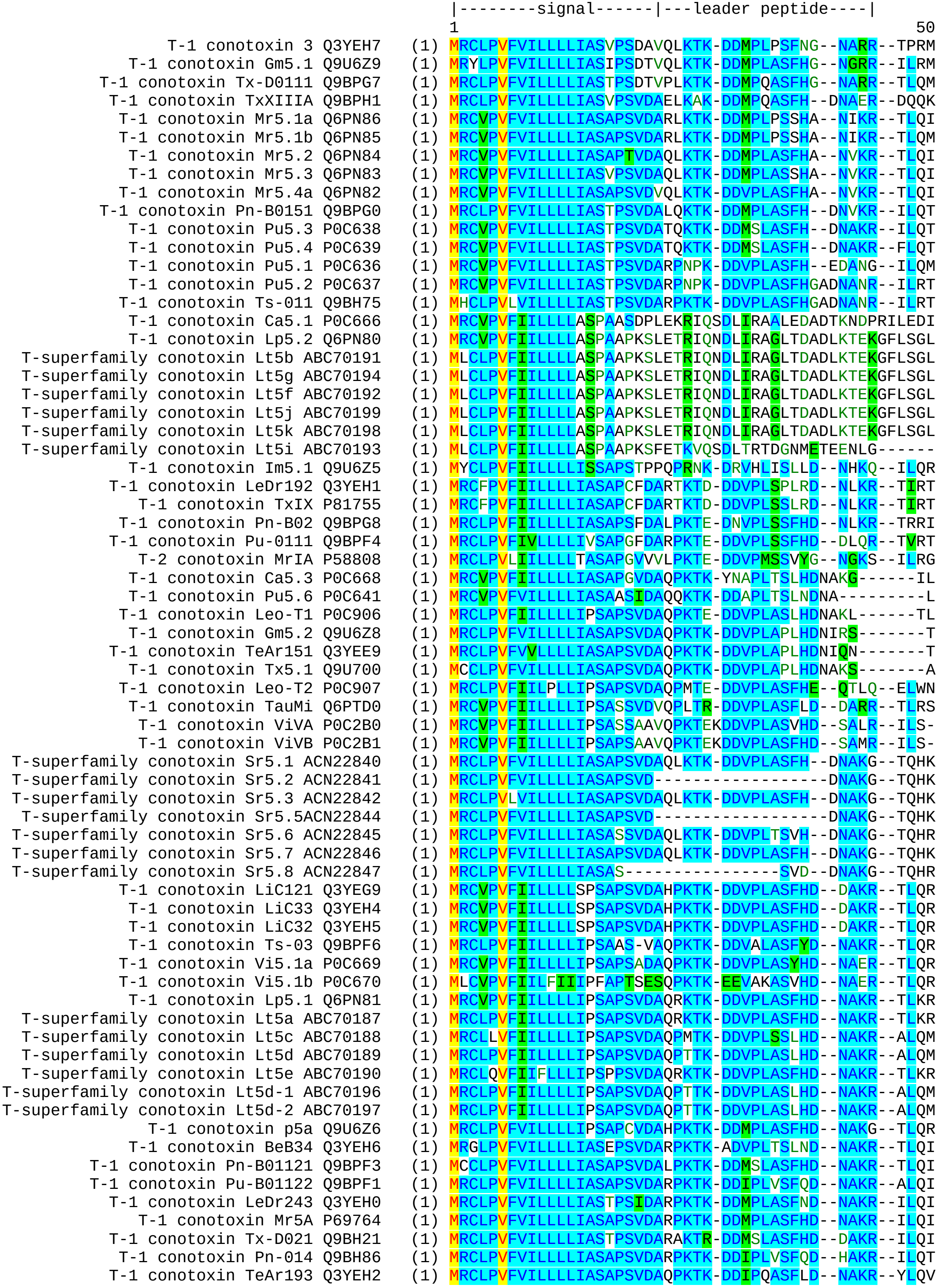
h. S-Superfamily

```
                                              |-------signal------|-----leader peptide-----|
                                              1                                               50
S-superfamily conotoxin Tx8.1 B2CI27   (1) --LKMGAMFVLLLLFTLASSHREGDIQARKTHLKSDFYRTLPRFARGCTI
S-superfamily conotoxin Ac8.1 B2CI28   (1) --LKMGAMFVLLLLFTLASSQQEGDVQARKTSLKSDFYRALRQYDRQCTL
S-superfamily conotoxin Ca8.1 B0L0Y5   (1) MMLKMGAMFVLLLLFILPSSQQEGDVQARKTHLKRGFYGTLAMSTRGCSG
S-superfamily conotoxin Ca8.2 B1PZN6   (1) MMLKMGAMFVLLLLFILPSSQQEGDVQARKTHLKRGFYGTLAMSTRGCSG
S-superfamily conotoxin Ca8.3 B2CJ87   (1) MMLKMGAMFVLLLLFILPSSQQEGDVQARKTHLKSGFYGTLAMSTRGCSG
    S-superfamily conotoxin ABQ00572   (1) MMLKMGAMFVLLLLFILPSSQQEGDVQARKTHLKRGFYGTLAMSTRGCSG
                           Consensus   (1) MMLKMGAMFVLLLLFILPSSQQEGDVQARKTHLKSGFYGTLAMSTRGCSG

                                              51                                          95
S-superfamily conotoxin Tx8.1 B2CI27  (49) S---CGYEDNRCQGECHCPGKTNCYCTSGHHNKGCGCAC------
S-superfamily conotoxin Ac8.1 B2CI28  (49) VNNCDRNGERACNGDCSCEG-QICKCGYRVSPGKSGCACTCRNAK
S-superfamily conotoxin Ca8.1 B0L0Y5  (51) T--CHRREDGKCRGTCDCSGYSYCRCGDAHHFYR-GCTCTC----
S-superfamily conotoxin Ca8.2 B1PZN6  (51) T--CHRREDGKCRGTCDCSGYSYCRCGDAHHFYR-GCTCSCQG--
S-superfamily conotoxin Ca8.3 B2CJ87  (51) T--CRRHRDGKCRGTCECSGYSYCRCGDAHHFYR-GCTCTC----
    S-superfamily conotoxin ABQ00572  (51) T--CHRREDGKCRGTCDCSGYSYCRCGDAHHFYR-GCTCTC----
                           Consensus  (51) T  CHRREDGKCRGTCDCSGYSYCRCGDAHHFYR GCTCTC
```



The leader peptides for this presently small family (in terms of full sequence information) have several conserved charged amino acids.

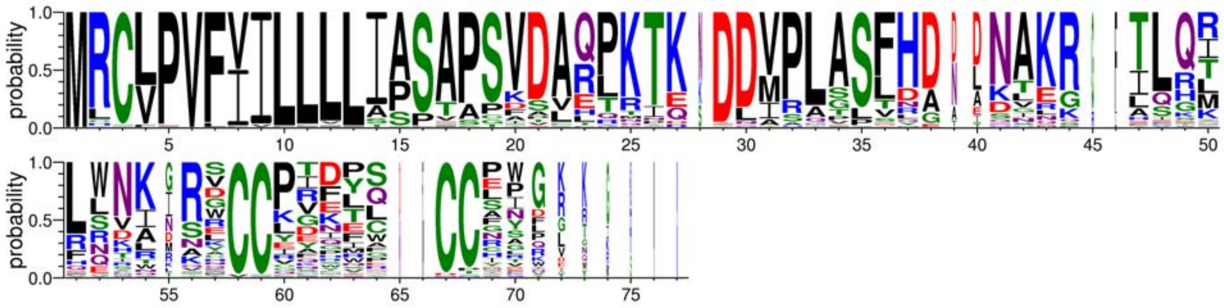
i. T-Superfamily



```
                                              |--------signal------|---leader peptide----|
                                              1                                               50
                 T-1 conotoxin 3 Q3YEH7   (1) MRCLPVFVILLLLIASVPSDAVQLKTK-DDMPLPSFNG--NARR--TPRM
             T-1 conotoxin Gm5.1 Q9U6Z9   (1) MRYLPVFVILLLLIASIPSDTVQLKTK-DDMPLASFHG--NGRR--ILRM
          T-1 conotoxin Tx-D0111 Q9BPG7   (1) MRCLPVFVILLLLIASTPSDTVPLKTK-DDMPQASFHG--NARR--TLQM
           T-1 conotoxin TxXIIIA Q9BPH1   (1) MRCLPVFVILLLLIASVPSVDAELKAK-DDMPQASFH--DNAER--DQQK
            T-1 conotoxin Mr5.1a Q6PN86   (1) MRCVPVFVILLLLIASAPSVDARLKTK-DDMPLPSSHA--NIKR--TLQI
            T-1 conotoxin Mr5.1b Q6PN85   (1) MRCVPVFVILLLLIASAPSVDARLKTK-DDMPLPSSHA--NIKR--TLQM
             T-1 conotoxin Mr5.2 Q6PN84   (1) MRCVPVFVILLLLIASAPTVDAQLKTK-DDMPLASFHA--NVKR--TLQI
             T-1 conotoxin Mr5.3 Q6PN83   (1) MRCVPVFVILLLLIASVPSVDAQLKTK-DDMPLASSHA--NVKR--TLQI
            T-1 conotoxin Mr5.4a Q6PN82   (1) MRCVPVFVILLLLIASAPSVDVQLKTK-DDVPLASFHA--NVKR--TLQI
          T-1 conotoxin Pn-B0151 Q9BPG0   (1) MRCLPVFVILLLLIASTPSVDALQKTK-DDMPLASFH--DNVKR--ILQT
             T-1 conotoxin Pu5.3 P0C638   (1) MRCVPVFVILLLLIASTPSVDATQKTK-DDMSLASFH--DNAKR--ILQT
             T-1 conotoxin Pu5.4 P0C639   (1) MRCVPVFVILLLLIASTPSVDATQKTK-DDMSLASFH--DNAKR--FLQT
             T-1 conotoxin Pu5.1 P0C636   (1) MRCVPVFVILLLLIASTPSVDARPNPK-DDVPLASFH--EDANG--ILQM
             T-1 conotoxin Pu5.2 P0C637   (1) MRCVPVFVILLLLIASTPSVDARPNPK-DDVPLASFHGADNANR--ILRT
            T-1 conotoxin Ts-011 Q9BH75   (1) MHCLPVLVILLLLIASTPSVDARPKTK-DDVPLASFHGADNANR--ILRT
             T-1 conotoxin Ca5.1 P0C666   (1) MRCVPVFIILLLLASPAASDPLEKRIQSDLIRAALEDADTKNDPRILEDI
             T-1 conotoxin Lp5.2 Q6PN80   (1) MRCVPVFIILLLLASPAAPKSLETRIQNDLIRAGLTDADLKTEKGFLSGL
  T-superfamily conotoxin Lt5b ABC70191   (1) MLCLPVFIILLLLASPAAPKSLETRIQNDLIRAGLTDADLKTEKGFLSGL
  T-superfamily conotoxin Lt5g ABC70194   (1) MLCLPVFIILLLLASPAAPKSLETRIQNDLIRAGLTDADLKTEKGFLSGL
  T-superfamily conotoxin Lt5f ABC70192   (1) MLCLPVFIILLLLASPAAPKSLETRIQNDLIRAGLTDADLKTEKGFLSGL
  T-superfamily conotoxin Lt5j ABC70199   (1) MLCLPVFIILLLLASPAAPKSLETRIQNDLIRAGLTDADLKTEKGFLSGL
  T-superfamily conotoxin Lt5k ABC70198   (1) MLCLPVFIILLLLASPAAPKSLETRIQNDLIRAGLTDADLKTEKGFLSGL
  T-superfamily conotoxin Lt5i ABC70193   (1) MLCLPVFIILLLLASPAAPKSFETKVQSDLTRTDGNMETEENLG------
             T-1 conotoxin Im5.1 Q9U6Z5   (1) MYCLPVFIILLLLISSAPSTPPQPRNK-DRVHLISLLD--NHKQ--ILQR
           T-1 conotoxin LeDr192 Q3YEH1   (1) MRCFPVFIILLLLIASAPCFDARTKTD-DDVPLSPLRD--NLKR--TIRT
              T-1 conotoxin TxIX P81755   (1) MRCFPVFIILLLLIASAPCFDARTKTD-DDVPLSSLRD--NLKR--TIRT
            T-1 conotoxin Pn-B02 Q9BPG8   (1) MRCLPVFIILLLLIASAPSFDALPKTE-DNVPLSSFHD--NLKR--TRRI
           T-1 conotoxin Pu-0111 Q9BPF4   (1) MRCLPVFIVLLLLIVSAPGFDARPKTE-DDVPLSSFHD--DLQR--TVRT
              T-2 conotoxin MrIA P58808   (1) MRCLPVLIILLLLTASAPGVVVLPKTE-DDVPMSSVYG--NGKS--ILRG
             T-1 conotoxin Ca5.3 P0C668   (1) MRCVPVFIILLLLIASAPGVDAQPKTK-YNAPLTSLHDNAKG------IL
             T-1 conotoxin Pu5.6 P0C641   (1) MRCVPVFVILLLLIASAASIDAQQKTK-DDAPLTSLNDNA---------L
            T-1 conotoxin Leo-T1 P0C906   (1) MRCLPVFIILLLLIPSAPSVDAQPKTE-DDVPLASLHDNAKL------TL
             T-1 conotoxin Gm5.2 Q9U6Z8   (1) MRCLPVFVILLLLIASAPSVDAQPKTK-DDVPLAPLHDNIRS-------T
           T-1 conotoxin TeAr151 Q3YEE9   (1) MRCLPVFVVLLLLIASAPSVDAQPKTK-DDVPLAPLHDNIQN-------T
             T-1 conotoxin Tx5.1 Q9U700   (1) MCCLPVFVILLLLIASAPSVDAQPKTK-DDVPLAPLHDNAKS-------A
            T-1 conotoxin Leo-T2 P0C907   (1) MRCLPVFIILPLLIPSAPSVDAQPMTE-DDVPLASFHE--QTLQ--ELWN
             T-1 conotoxin TauMi Q6PTD0   (1) MRCVPVFIILLLLIPSASSVDVQPLTR-DDVPLASFLD--DARR--TLRS
              T-1 conotoxin ViVA P0C2B0   (1) MRCVPVFIILLLLIPSASSAAVQPKTEKDDVPLASVHD--SALR--ILS-
              T-1 conotoxin ViVB P0C2B1   (1) MRCVPVFIILLLLIPSAPSAAVQPKTEKDDVPLASFHD--SAMR--ILS-
 T-superfamily conotoxin Sr5.1 ACN22840   (1) MRCLPVFVILLLLIASAPSVDAQLKTK-DDVPLASFH--DNAKG--TQHK
 T-superfamily conotoxin Sr5.2 ACN22841   (1) MRCLPVFVILLLLIASAPSVD------------------DNAKG--TQHK
 T-superfamily conotoxin Sr5.3 ACN22842   (1) MRCLPVLVILLLLIASAPSVDAQLKTK-DDVPLASFH--DNAKG--TQHK
  T-superfamily conotoxin Sr5.5ACN22844   (1) MRCLPVFVILLLLIASAPSVD------------------DNAKG--TQHK
 T-superfamily conotoxin Sr5.6 ACN22845   (1) MRCLPVFVILLLLIASASSVDAQLKTK-DDVPLTSVH--DNAKG--TQHR
 T-superfamily conotoxin Sr5.7 ACN22846   (1) MRCLPVFVILLLLIASAPSVDAQLKTK-DDVPLASFH--DNAKG--TQHK
 T-superfamily conotoxin Sr5.8 ACN22847   (1) MRCLPVFVILLLLIASAS----------------SVD--DNAKG--TQHR
            T-1 conotoxin LiC121 Q3YEG9   (1) MRCVPVFIILLLLSPSAPSVDAHPKTK-DDVPLASFHD--DAKR--TLQR
             T-1 conotoxin LiC33 Q3YEH4   (1) MRCVPVFIILLLLSPSAPSVDAHPKTK-DDVPLASFHD--DAKR--TLQR
             T-1 conotoxin LiC32 Q3YEH5   (1) MRCVPVFIILLLLSPSAPSVDAHPKTK-DDVPLASFHD--DAKR--TLQR
             T-1 conotoxin Ts-03 Q9BPF6   (1) MRCLPVFIILLLLIPSAAS-VAQPKTK-DDVALASFYD--NAKR--TLQR
            T-1 conotoxin Vi5.1a P0C669   (1) MRCVPVFIILLLLIPSAPSADAQPKTK-DDVPLASYHD--NAER--TLQR
            T-1 conotoxin Vi5.1b P0C670   (1) MLCVPVFIILFIIIPFAPTSESQPKTK-EEVAKASVHD--NAER--TLQR
             T-1 conotoxin Lp5.1 Q6PN81   (1) MRCVPVFIILLLLIPSAPSVDAQRKTK-DDVPLASFHD--NAKR--TLKR
  T-superfamily conotoxin Lt5a ABC70187   (1) MRCLPVFIILLLLIPSAPSVDAQRKTK-DDVPLASFHD--NAKR--TLKR
  T-superfamily conotoxin Lt5c ABC70188   (1) MRCLLVFIILLLLIPSAPSVDAQPMTK-DDVPLSSLHD--NAKR--ALQM
  T-superfamily conotoxin Lt5d ABC70189   (1) MRCLPVFIILLLLIPSAPSVDAQPTTK-DDVPLASLHD--NAKR--ALQM
  T-superfamily conotoxin Lt5e ABC70190   (1) MRCLQVFIIFLLLIPSPPSVDAQRKTK-DDVPLASFHD--NAKR--TLKR
T-superfamily conotoxin Lt5d-1 ABC70196   (1) MRCLPVFIILLLLIPSAPSVDAQPTTK-DDVPLASLHD--NAKR--ALQM
T-superfamily conotoxin Lt5d-2 ABC70197   (1) MRCLPVFIILLLLIPSAPSVDAQPTTK-DDVPLASLHD--NAKR--ALQM
               T-1 conotoxin p5a Q9U6Z6   (1) MRCLPVFVILLLLIPSAPCVDAHPKTK-DDMPLASFHD--NAKG--TLQR
             T-1 conotoxin BeB34 Q3YEH6   (1) MRGLPVFVILLLLIASEPSVDARPKTK-ADVPLTSLND--NAKR--TLQI
         T-1 conotoxin Pn-B01121 Q9BPF3   (1) MCCLPVFVILLLLIASAPSVDALPKTK-DDMSLASFHD--NAKR--TLQI
         T-1 conotoxin Pu-B01122 Q9BPF1   (1) MRCLPVFVILLLLIASAPSVDARPKTK-DDIPLVSFQD--NAKR--ALQI
           T-1 conotoxin LeDr243 Q3YEH0   (1) MRCLPVFVILLLLIASTPSIDARPKTK-DDMPLASFND--NAKR--ILQI
              T-1 conotoxin Mr5A P69764   (1) MRCLPVFVILLLLIASAPSVDARPKTK-DDMPLASFHD--NAKR--ILQI
           T-1 conotoxin Tx-D021 Q9BH21   (1) MRCLPVFVILLLLIASTPSVDARAKTR-DDMSLASFHD--DAKR--ILQI
            T-1 conotoxin Pn-014 Q9BH86   (1) MRCLPVFVILLLLIASAPSVDARPKTK-DDIPLVSFQD--HAKR--ILQT
           T-1 conotoxin TeAr193 Q3YEH2   (1) MRCLPVFVILLLLIASAPSVDAQPKTK-DDIPQASFLD--NAKR--YLQV
```

```
             T-1 conotoxin Ca5.2 P0C667   (1) MRCVPVFLILLGLIASAPSVDARPQTK-DDA-LASFHD--SAKR--HLQR
             T-1 conotoxin Pu5.5 P0C640   (1) MRCVPVFIILLVLIASAPSVDARPQTK-DDA-LASFRD--SIKR--HLQT
             T-1 conotoxin Vn-05 Q9BPF2   (1) MRCLPVFIILLVLSASAPSVDARPKNE-DDVSLASFHD--NAQR--TLQR
           T-1 conotoxin TeAr154 Q3YEE8   (1) MHCLPVFVILLLLTASGLSVDARPKTE-DDVPLSSFRD--NTKS--TLQR
                              Consensus   (1) MRCLPVFVILLLLIASAPSVDA PKTK DDVPLASFHD  NAKR   L

                                              51                       77
                 T-1 conotoxin 3 Q3YEH7  (46) LSNK-RICCYPNVW--CCD--------
             T-1 conotoxin Gm5.1 Q9U6Z9  (46) LSNK-RLCCVTEDW--CCEWW------
          T-1 conotoxin Tx-D0111 Q9BPG7  (46) LSKK-QCCWYFDIS--CCITV------
           T-1 conotoxin TxXIIIA Q9BPH1  (46) ---KTSDCCFYHNC--CC---------
            T-1 conotoxin Mr5.1a Q6PN86  (46) HRNK-R-CCPGWEL--CCEWDEW----
            T-1 conotoxin Mr5.1b Q6PN85  (46) LRNK-R-CCPGWEL--CCEWDDGW---
             T-1 conotoxin Mr5.2 Q6PN84  (46) LRDK-RFCCRTQEV--CCEAIKNG---
             T-1 conotoxin Mr5.3 Q6PN83  (46) LRNK-R-CCITFES--CCEFDLK----
            T-1 conotoxin Mr5.4a Q6PN82  (46) LLNK-R-CCQVMPQ--CCEWN------
          T-1 conotoxin Pn-B0151 Q9BPG0  (46) LSNK-RSCCPTILS--CCFV-------
             T-1 conotoxin Pu5.3 P0C638  (46) LRNT-RSCCPEEP---CCFWR------
             T-1 conotoxin Pu5.4 P0C639  (46) LRNT-RSCCPEEIT--CCPWG------
             T-1 conotoxin Pu5.1 P0C636  (46) LWKKGRSCCPSPTS--CCPWGKRK---
             T-1 conotoxin Pu5.2 P0C637  (48) LWNL-RGCCEDKT---CCFIG------
            T-1 conotoxin Ts-011 Q9BH75  (48) LWNL-RGCCEDKT---CCFIG------
             T-1 conotoxin Ca5.1 P0C666  (51) VSTALATCCKFQFLNFCCNEK------
             T-1 conotoxin Lp5.2 Q6PN80  (51) LNVAGSVCCKVDTS--CCSN-------
  T-superfamily conotoxin Lt5b ABC70191  (51) LNVAGSVCCKVDTS--CCSN-------
  T-superfamily conotoxin Lt5g ABC70194  (51) LNVAGSVCCKVDTS--CCSNQ------
  T-superfamily conotoxin Lt5f ABC70192  (51) LNVAGSVCCKVDTS--CC---------
  T-superfamily conotoxin Lt5j ABC70199  (51) LNVAGSVCCKVDTS--CCSSQ------
  T-superfamily conotoxin Lt5k ABC70198  (51) LNVADSVCCKVDTS--CCSN-------
  T-superfamily conotoxin Lt5i ABC70193  (45) -EVRKVYCCLGVRDDWCCAGQIQI---
             T-1 conotoxin Im5.1 Q9U6Z5  (46) DWNS---CCGKNPG--CCPWGK-----
           T-1 conotoxin LeDr192 Q3YEH1  (46) RLNI-RECCEDGWC--CTAAPLTGR--
              T-1 conotoxin TxIX P81755  (46) RLNI-RECCEDGWC--CTAAPLTGR--
            T-1 conotoxin Pn-B02 Q9BPG8  (46) HLNI-RECCSDGWC--CPAG-------
           T-1 conotoxin Pu-0111 Q9BPF4  (46) LLDI-RMCCLGTSG--CCPWG------
              T-2 conotoxin MrIA P58808  (46) ILRN-GVCCGYKLC--HPC--------
             T-1 conotoxin Ca5.3 P0C668  (44) QEHWNKRCCPRRLA--CCIIGRK----
             T-1 conotoxin Pu5.6 P0C641  (41) QQHWNKRCCPRKIW--CCMIPR-----
            T-1 conotoxin Leo-T1 P0C906  (44) QGLWDKRCCPNLFY--CCPDRRK----
             T-1 conotoxin Gm5.2 Q9U6Z8  (43) LQTLRKKVCCRPVQ--DCCSGK-----
           T-1 conotoxin TeAr151 Q3YEE9  (43) LQTLRKKVCCRPMQ--DCCSGK-----
             T-1 conotoxin Tx5.1 Q9U700  (43) LQHLNQRCCQTFYW--CCVQGK-----
            T-1 conotoxin Leo-T2 P0C907  (46) ---K-RPCCPLIPG--CCR--------
             T-1 conotoxin TauMi Q6PTD0  (46) PWMT-RRCCPGNFA--CCGK-------
              T-1 conotoxin ViVA P0C2B0  (46) -----RQCCITIPE--CCRIG------
              T-1 conotoxin ViVB P0C2B1  (46) -----RQCCPTIPE--CCRVG------
 T-superfamily conotoxin Sr5.1 ACN22840  (46) R--IINWCCLIFYQ--CC---------
 T-superfamily conotoxin Sr5.2 ACN22841  (31) R--IINWCCLIFYQ--CC---------
 T-superfamily conotoxin Sr5.3 ACN22842  (46) R--IINWCCLIFYQ--CC---------
  T-superfamily conotoxin Sr5.5ACN22844  (31) R--IINWCCLTFYQ--CC---------
 T-superfamily conotoxin Sr5.6 ACN22845  (46) R--IMAGCCPRFYQ--CCYPG------
 T-superfamily conotoxin Sr5.7 ACN22846  (46) R--IINWCCLIFYQ--CCLRR------
 T-superfamily conotoxin Sr5.8 ACN22847  (31) R--IMAGCCPRFYQ--CCYPG------
            T-1 conotoxin LiC121 Q3YEG9  (46) LWIK-ALCCYGYRF--CCPNFR-----
             T-1 conotoxin LiC33 Q3YEH4  (46) LWIK-ALCCYGYRF--CCPIFGKGK--
             T-1 conotoxin LiC32 Q3YEH5  (46) LWQN-TWCCRDHLR--CCG--------
             T-1 conotoxin Ts-03 Q9BPF6  (45) HWAK-SLCCPEDAW--CC---------
            T-1 conotoxin Vi5.1a P0C669  (46) LWNQ-RHCCPIDLP--CCPPG------
            T-1 conotoxin Vi5.1b P0C670  (46) LWNQ-SHCCPIDLQ--CCPPG------
             T-1 conotoxin Lp5.1 Q6PN81  (46) LWNK-RSCCPQEFL--CCLYLVK----
  T-superfamily conotoxin Lt5a ABC70187  (46) LWNK-RSCCPQEFL--CCLYLVK----
  T-superfamily conotoxin Lt5c ABC70188  (46) FWNK-RDCCPAKMF--CCQW-------
  T-superfamily conotoxin Lt5d ABC70189  (46) FWNK-RDCCPAKLL--CCNP-------
  T-superfamily conotoxin Lt5e ABC70190  (46) LWNK-RSCCPREFL--CCLR-------
T-superfamily conotoxin Lt5d-1 ABC70196  (46) FWNK-RDCCPSKLL--CCNP-------
T-superfamily conotoxin Lt5d-2 ABC70197  (46) FWNK-RDCCPANLL--CCNP-------
               T-1 conotoxin p5a Q9U6Z6  (46) FWKK-RGCCPKQMR--CCTLG------
             T-1 conotoxin BeB34 Q3YEH6  (46) LRNK-RACCPYEPS--CCI--------
         T-1 conotoxin Pn-B01121 Q9BPF3  (46) LSNK-RYCCVYDYS--CCLSWG-----
         T-1 conotoxin Pu-B01122 Q9BPF1  (46) LSNK-RYCCYFDYS--CCLYLR-----
           T-1 conotoxin LeDr243 Q3YEH0  (46) LSRK-PCCSIHDNS--CCGLG------
              T-1 conotoxin Mr5A P69764  (46) LQDR-NACCIVRQC--C----------
```

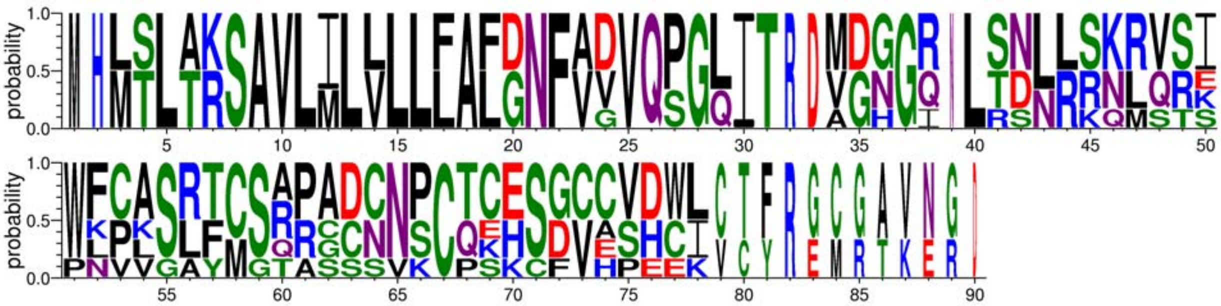
```
T-1 conotoxin Tx-D021 Q9BH21  (46) LQDR-SGCCVIDSN--CCG--------
 T-1 conotoxin Pn-014 Q9BH86  (46) FESR-YDCCKTFEC--CHWG-------
T-1 conotoxin TeAr193 Q3YEH2  (46) LESK-RNCCRRQIC--CGRTK------
  T-1 conotoxin Ca5.2 P0C667  (45) LVNA-RKCCPESPP--CCHYFGRRK--
  T-1 conotoxin Pu5.5 P0C640  (45) LLDA-RECCPQSPP--CCHYYYYGSWK
  T-1 conotoxin Vn-05 Q9BPF2  (46) LLNK-RSCCPNNPA--CCH--------
T-1 conotoxin TeAr154 Q3YEE8  (46) LLKR-VNCCPIDES--CCS--------
                   Consensus  (51) L    R CC       CC
```



The leader peptides of this large family contain a highly conserved Asp at position 29 and relatively well conserved Pro, Leu, and Ser at positions 32, 33, and 35, respectively.

j. P-Superfamily

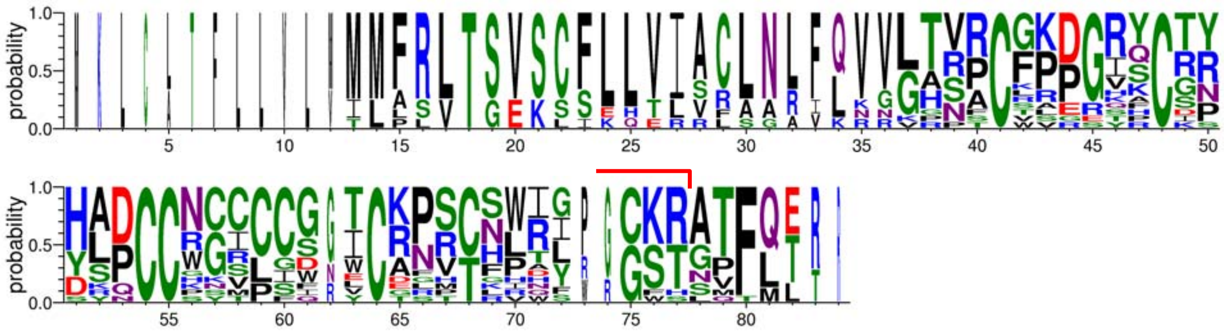
```
                                               |----------signal---------|---leader peptide--|
                                               1                                                50
     P-superfamily conotoxin Lt9a ABC74994   (1) --MTLTKSAVLILVLLLAFDNFADVQPGLIT--MGGGR-LSNLLSKRVRI
P-superfamily conotoxin Lt9a var ABC74995   (1) --MTLTKSAVLILVLLLAFDNFADVQPGLIT--MGGGR-LSNLLSKRVSI
P-superfamily conotoxin Lt9a var ABC74996   (1) --MTLTKSAVLILVLLLAFDNFADVQPGLIT--MGGGR-LSNLLSKRVSI
      P-superfamily conotoxin Gm9.1 Q9GU57   (1) MHLSLARSAVLMLLLLFALGNFVVVQSGLITRDVDNGQ-LTDNRRNLQTE
      P-superfamily conotoxin Rg9.1 Q8I6V7   (1) MHLSLARSAVLILLLLFALGNFVGVQPGQITRDADHGINLRSLRKQMSRS
             Spasmodic peptide Tx9a Q9GU58   (1) MHLSLARSAVLMLLLLFALGNFVVVQSGQITRDVDNGQ-LTDNRRNLQSK
                                 Consensus   (1) MHLSLTKSAVLILLLLLALGNFVDVQPGLITRDMGGGR LSNLRSKRVSI

                                                51                                     90
     P-superfamily conotoxin Lt9a ABC74994  (46) WFCASRTCSAPADCNPCTCESGVCVDWL------------
P-superfamily conotoxin Lt9a var ABC74995  (46) WFCASRTCSAPADCNPCTCESGVCVDWL------------
P-superfamily conotoxin Lt9a var ABC74996  (46) WFCASRTCSTPADCNPCTCESGVCVDWL------------
      P-superfamily conotoxin Gm9.1 Q9GU57  (50) WNPLSLFMSRRSCNNSCQSHSDCASHCICTFRGCGAVNG-
      P-superfamily conotoxin Rg9.1 Q8I6V7  (51) PLVKGAFCGQACSSVKCPKKCFCHPEEKVCYREMRTKERD
             Spasmodic peptide Tx9a Q9GU58  (50) WKPVSLYMSRRGCNNSCQEHSDCESHCICTFRGCGAVNG-
                                 Consensus  (51) WFCASRTCS PADCNPCTCESGVCVDWL   FR
```



This presently small family (in terms of full sequence information) has highly conserved leader peptides including a conserved IleThrArgAsp motif and a conserved Leu at position 40.

k. I-Superfamily

```
                                              |------------leader peptide------------|
                                              1                                                50
    I2-superfamily conotoxin 1 P69500    (1) ------------MMFRLTSVSCFLLVIACLNLFQVVLTSRCFPPGIYCTP
    I2-superfamily conotoxin 2 P69501    (1) ------------MMFRLTSVSCFLLVIACLNLFQVVLTSRCFPPGIYCTP
 I2-superfamily conotoxin ViTx Q7YZS9    (1) ------------MMFRLTSVSCFLLVIACLNLFQVVLTSRCFPPGIYCTP
      I2-superfamily conotoxin P69494    (1) ------------MMFRLTSVSCFLLVIACLNLFQVVLTRRCFPPGVYCTR
      I2-superfamily conotoxin P69498    (1) ------------MMLRLTSVSCFLLVIACLNLFQVVLTRRCFPPGTFCSR
      I2-superfamily conotoxin P69499    (1) ------------MMFRLTSVSCFLLVIVCLNLFQVVLTRRCVPPSRYCTR
    I2-superfamily conotoxin 1 P69495    (1) ------------MMFRLTSVSCILLVIAFLN--LVGLTNACTSEGYSCSS
    I2-superfamily conotoxin 2 P69496    (1) ------------MMFRLTSVSCFLLVIVCLN--LVVLTNACRLEGSSCRR
    I2-superfamily conotoxin 3 P69497    (1) ------------MMFRLTSVSCFLLVIACLN--LVVLTNACLRDGQSCGY
  I1-superfamily conotoxin R11.2 Q7Z099  (1) ------------------------------------GAVPCGKDGRQCRN
  I1-superfamily conotoxin R11.5 Q7Z098  (1) ------------------------------------GAVPCGKDGRQCRN
 I1-superfamily conotoxin R11.11 Q7Z0A0  (1) ------------------------------------GHVSCGKDGRACDY
 I1-superfamily conotoxin R11.12 Q7Z0A4  (1) ------------------------------------GHVPCGKDRRKCGY
 I1-superfamily conotoxin R11.13 Q7Z0A3  (1) ------------------------------------GHVPCGKDGRKCGY
 I1-superfamily conotoxin R11.18 Q7Z097  (1) ------------------------------------GAVPCGKDGRQCRN
   I1-superfamily conotoxin R11c Q7Z096  (1) MKLCLTFLLVLMILASVTGEKSSKHTLSRAARVKNRGPSFCKADEKPCKY
  I1-superfamily conotoxin R11.3 Q7Z090  (1) MKLCLTFLLVLMILASVTGEKLSEQTLRRAAR--KNKGPRCWVGRVHCTY
  I1-superfamily conotoxin Ar11a P0C607  (1) MKLCATFLLVLVTLPLVTGEKSS--ERSLSGAILRGVRRTCSRRGHRCIR
                              Consensus  (1)             MM RLTSVSC LLVIA LN   V LT  C  DG  C

                                                                 recognition sequence
                                                                            |---|
                                              51                               84
    I2-superfamily conotoxin 1 P69500   (39) YLPCCWGICCD-TCRNVCHLRI--GKRATFQE--
    I2-superfamily conotoxin 2 P69501   (39) YLPCCWGICCD-TCRNVCHLRF--GKRATFQE--
 I2-superfamily conotoxin ViTx Q7YZS9   (39) YLPCCWGICCG-TCRNVCHLRI--GKRATFQE--
      I2-superfamily conotoxin P69494   (39) HLPCCRGRCCSGWCRPRCFPRY--GKRATFQ---
      I2-superfamily conotoxin P69498   (39) YLPCCSGRCCSGWCTRRCSPRY--GKRATFQE--
      I2-superfamily conotoxin P69499   (39) HRPCCRGTCCSGLCRPMCNLWY--GKRATFQE--
    I2-superfamily conotoxin 1 P69495   (37) DSNCCKNVCCWNVCESHCGHH---GKRATFQ---
    I2-superfamily conotoxin 2 P69496   (37) SYQCCHKSCCIRECKFPCRWD---GKRATFQ---
    I2-superfamily conotoxin 3 P69497   (37) DSDCCRYSCCWGYCDLTCLIN---GKRATFQ---
  I1-superfamily conotoxin R11.2 Q7Z099 (15) HADCCNCCPIG-TCAPSTNWILPGCSTGPFMTR-
  I1-superfamily conotoxin R11.5 Q7Z098 (15) HADCCNCCPIG-TCAPSTNWILPGCSTGQFMTR-
 I1-superfamily conotoxin R11.11 Q7Z0A0 (15) HADCCNCCLGG-ICKPSTSWIG--CSTNVFLTR-
 I1-superfamily conotoxin R11.12 Q7Z0A4 (15) HADCCNCCLSG-ICKPSTSWTG--CSTSTFLLTR
 I1-superfamily conotoxin R11.13 Q7Z0A3 (15) HADCCNCCLSG-ICKPSTSWTG--CSTSTFLLTR
 I1-superfamily conotoxin R11.18 Q7Z097 (15) HADCCNCCPFG-TCAPSTNRILPGCSTGMFLTR-
   I1-superfamily conotoxin R11c Q7Z096 (51) HADCCNCCLGG-ICKPSTSWIG--CSTNVFLTR-
  I1-superfamily conotoxin R11.3 Q7Z090 (49) HKDCCPSVCCF---KGRCKPQSWGCWSGPT----
  I1-superfamily conotoxin Ar11a P0C607 (49) DSQCCGGMCCQ---GNRCFVAIRRCFHLPF----
                              Consensus (51) HADCC   CC   CKP C      GKRATFQ
```



The red line shows the XG-(K/R)$_n$ motif for amidation after residue X in some family members (see the alignment of the I2-superfamily members above).

References for Supplementary Information

1. Morris, R. P. et al. Ribosomally Synthesized Thiopeptide Antibiotics Targeting Elongation Factor Tu. *J. Am. Chem. Soc.*, 5946-5955 (2009).
2. Maqueda, M. et al. Genetic features of circular bacteriocins produced by Gram-positive bacteria. *FEMS Microbiol. Rev.* **32**, 2-22 (2008).
3. Duquesne, S., Destoumieux-Garzon, D., Peduzzi, J. & Rebuffat, S. Microcins, gene-encoded antibacterial peptides from enterobacteria. *Nat. Prod. Rep.* **24**, 708-34 (2007).
4. Roy, R. S., Allen, O. & Walsh, C. T. Expressed protein ligation to probe regiospecificity of heterocyclization in the peptide antibiotic microcin B17. *Chem. Biol.* **6**, 789-99 (1999).
5. Buczek, O., Bulaj, G. & Olivera, B. M. Conotoxins and the posttranslational modification of secreted gene products. *Cell Mol. Life Sci.* **62**, 3067-79 (2005).
6. de Vos, W. M., Jung, G. & Sahl, H.-G. in *Nisin and Novel Lantibiotics* (eds. Jung, G. & Sahl, H.-G.) 457-464 (ESCOM, Leiden, 1991).
7. Cotter, P. D., Hill, C. & Ross, R. P. Bacterial lantibiotics: strategies to improve therapeutic potential. *Curr. Protein Pept. Sci.* **6**, 61-75 (2005).
8. Chatterjee, C., Paul, M., Xie, L. & van der Donk, W. A. Biosynthesis and Mode of Action of Lantibiotics. *Chem. Rev.* **105**, 633-684 (2005).
9. Willey, J. M. & van der Donk, W. A. Lantibiotics: Peptides of Diverse Structure and Function. *Annu. Rev. Microbiol.* **61**, 477-501 (2007).
10. Li, Y. M., Milne, J. C., Madison, L. L., Kolter, R. & Walsh, C. T. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. *Science* **274**, 1188-93 (1996).
11. Mitchell, D. A. et al. Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S. *J. Biol. Chem.* **284**, 13004-12 (2009).
12. Liao, R. et al. Thiopeptide biosynthesis featuring ribosomally synthesized precursor peptides and conserved posttranslational modifications. *Chem. Biol.* **16**, 141-7 (2009).
13. Jennings, C., West, J., Waine, C., Craik, D. & Anderson, M. Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis*. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10614-9 (2001).
14. Ziemert, N., Ishida, K., Liaimer, A., Hertweck, C. & Dittmann, E. Ribosomal synthesis of tricyclic depsipeptides in bloom-forming cyanobacteria. *Angew. Chem., Int. Ed. Engl.* **47**, 7756-9 (2008).
15. Philmus, B., Christiansen, G., Yoshida, W. Y. & Hemscheidt, T. K. Post-translational modification in microviridin biosynthesis. *ChemBioChem* **9**, 3066-73 (2008).
16. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097-100 (1990).
17. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188-90 (2004).
18. Duquesne, S. et al. Two Enzymes Catalyze the Maturation of a Lasso Peptide in *Escherichia coli*. *Chem. Biol.* **14**, 793-803 (2007).