Drosophila melanogaster known antimicrobial peptides

BLASTP
TBLASTN

Musca domestica protein and genomic database

(http://www.ncbi.nlm.nih.gov)

Filter using signal peptide

Putative _Musca domestica_ antimicrobial peptides

Pattern recognition
(CAMP)

Putative _Musca domestica_ peptides

*Filter with signal peptide and
characteristics of known
antimicrobial peptides*

Musca domestica proteins in GenBank and Uniprot
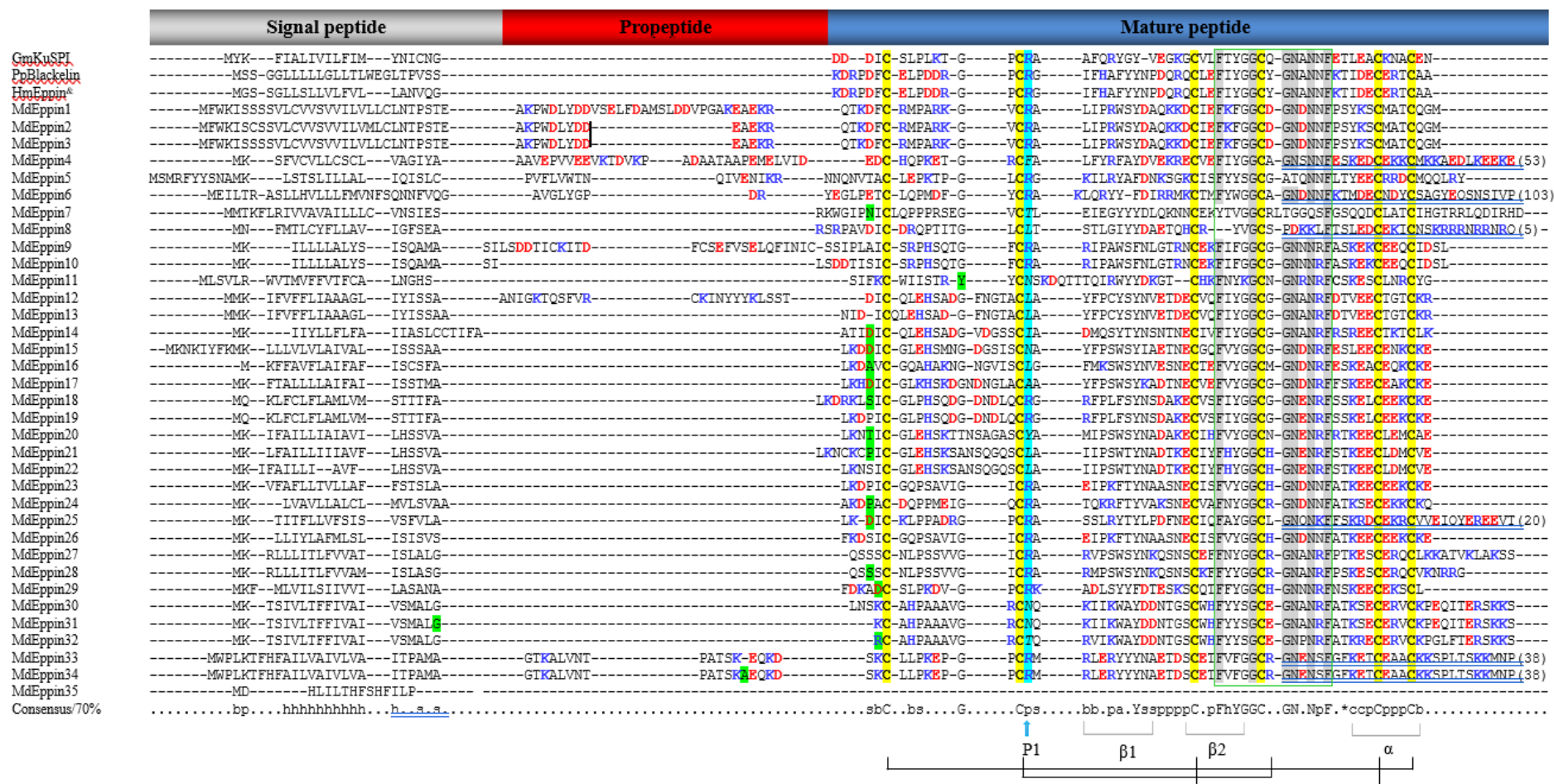
(http://www.ncbi.nlm.nih.gov)

Figure S1. The strategy for database searches for putative *M. domestica* antimicrobial peptides. For the BLASTP search, the expect threshold was set to 0.05 with the BLOSUM62 matrix. The Existence 11 and extension 2 were chosen in Gap costs. For the TBLASTN search, two databases, the whole genome shotgun and nucleotide collection databases, were used, with the following algorithm parameters: expect threshold of 0.05, the BLOSUM62 matrix and the Existence 11 and extension2 for gap costs. In PHI-BLAST, the word size was set to 3 and expect threshold 0.05, with the BLOSUM62 matrix. For short input sequences, an automatically adjust parameters are used in both BlASTP and TBLASTN.
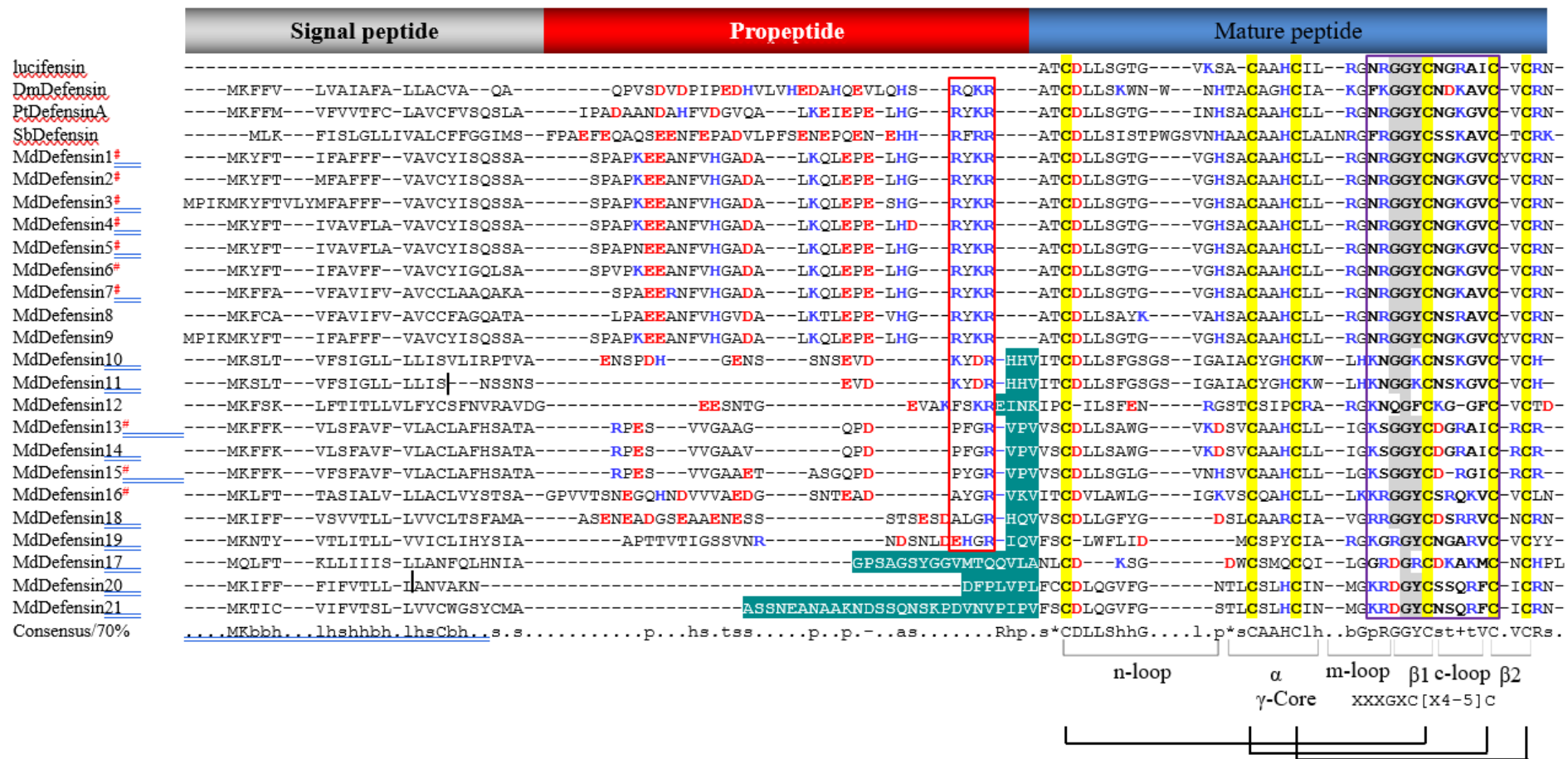
Figure S2. Multiple sequence alignment (MSA) of defensins. MdDefensins all are derived from the housefly. Cysteines involved in the formation of disulfide bridges are colored in yellow. Conservation replacements are shadowed in grey. Consensu/70% (Con.): - (negative), * (Ser/Thr), l (aliphatic), + (positive), t (tiny), a (aromatic), c (charged), s (small), p (polar), b (big), h (hydrophobic). # represents the defensins previously known. Phase 0 introns are marked by a black perpendicular line. The RXXR motif and its variants are highlighted in red box. The external N-terminus residues are shadowed in cyan. Dm: *Drosophila melanogaster* (GenBank: P36192.1), Lucifensin: *Lucilia sericata* (PDB:2LLD), Pt: *Protophormia terraenovae* (GenBank: P10891.2), Sb: *Simulium bannaense* (GenBank: AJP36711.1).

1

Figure S3. MSA of MdEppins. Cysteines involved in the formation of disulfide bridges are colored in yellow. Conservation replacements are shadowed in grey. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. The P1 amino acids are italicized and shadowed in cyan-blue. Conserved disulfides, α-helix and two-stranded β-sheets are also indicated at the bottom. The phase1 intron were boxed in green. ᵟ: The signal peptide and the kuntiz-domain in the human eppin are displayed. Conserved disulfide bridges are indicated at the bottom. Gm: *Galleria mellonella* (GenBank: AAK40037.1), Hm: *Homo sapiens*(GenBank:AAG00547.1), PP: *Pseudechis porphyriacus*(GenBank: sp_B5G6G6.1).

	Signal peptide	Mature peptide

```
AalKSPI     ---MRLVVAVLAVMLIIGLAQA----------------------------------------------------------------
AaeKSPI     ---MRHIGVFVGVLALALVLLVVEA-------------------------------------------------------------
AcKSPI      ------MWFFLILAIAANRNVIA-----------FPQNDLSDNNIEIQTLPTT----FEVDGFIFDGPANRPLRQTSTIITTTTTAATVDPQF-
CsKSPI      ---MKLTVLFWAVLLFSVCVLS--------------------------HQDETASVFPNNHKLRTS-GGPRSEA-------------
CpKSPI      -MNGRVLVLFALLGALVVLTEA----------------------------------------------------------------
DsM1        ---MMSQTLTLCCLGLVACVYGNTV-------------------------------------------------------------
DftPal9.2   ---MKCVAIFAIISLALMALVKA---------------------------------------------------------------
MdMuslin1   ---MRCFVIFLAFIVAIVASVSA---------------------------------------------------------------
MdMuslin2   ---MRFLIVIFAILMTLALANG---------------------------------------------------------------
MdMuslin3   ---MRFFCILIALVMLIALVSS---------------------------------------------------------------
MdMuslin4   ---MRFLAILIAVVMAIVLVSA---------------------------------------------------------------
MdMuslin5   ---MRFLAILIAIVMALVLVSA---------------------------------------------------------------
MdMuslin6   ---MRFLQVVFVLFVAVFATAYA--------------------------------------------------------------
MdMuslin7   ---MRFFCILIALVMLIALVSS---------------------------------------------------------------
MdMuslin8   ---MRFLIVIFAILMTLALANG---------------------------------------------------------------
MdMuslin9   ---MRFLFVALVLFMAIICVVA---------------------------------------------------------------
MdMuslin10  ---MKFLHLILALVLAICAFNTQPASA----------------------------------------------------------
MdMuslin11  ---MKLFAVFALVLMALLAFASA--------------------------------------DYRVAGLDSETEAD------------
MdMuslin12  MKYFSVLLFFILSLSALMSLTQA--------------------------------------------------------------
MdMuslin13  ---MKLLLGFVMLTVLQCSTLA------------LPQRPSDRFFPQRLPGAGNNGNPFLPSTNNNQGNLGNNNQGNSGNNNNIVSSTTAAPTTA
MdMuslin14  ---MKLFPLFLAMAIFVISFVNS--------------------------------------------------------------
MdMuslin15  -MFTKIVITLMVLAATLAPLNA---------------------------------------------------------------
MdMuslin16  -MFTKIVITLMVLAATLAPLNA---------------------------------------------------------------
MdMuslin17  --MVTLKWIFAIALIVIASNADA----------------------------------RSPPRPS-PGQVDES-------------
MdMuslin18  MKMASSKWIFTIALFVIVAAVLPNVAEA------------------------------QRGRPGRPGRPGQGNQG-----------
MdMuslin19  MKMASSKWIFTIALFVIVAAVLPNVAEA------------------------------QRGRPGRPGRPGQGNQG-----------
MdMuslin20  -MLAKFVIVFILLSVLSFTCLA---------------------------------------------------------------
MdMuslin21  --MFKLVVLISLLMALFSSPSHG-------------------------------YTVFSDEDT-----VDTVTET------------
MdMuslin22  ---MRFALGFFLVICLVAMAVA---------------------------------------VPAKNKQ-----------------
MdMuslin23  -MNSKQLLISLCLCLFVTLMMTAATALA---------------------------------------------------------
MdMuslin24  MFSLKLITSFLAVSFVAKEISS---------------------------------------------------------------
MdMuslin25  MYLIPLITTFLLALSLNPWVTEA--------------------------------------------------------------
```

Continued

	Signal peptide	Mature peptide
MdMuslin26	MFALRSIAIFVLLFSLLGVHG	---
MdMuslin27	MRSTTFIVILVLALSLYPYTAEA	---
MdMuslin28	MFSNSSFVILTCFIVCLVASIQA	---
MdMuslin29	MKFVASIALAILALLLAVAIGEA	---
MdMuslin30	MKFVASIALAILALLLAVAIGEA	---
MdMuslin31	MIASSPIVIFFLLFSLLMGVQG	---
MdMuslin32	---MRYTTSVAICMALVALCYIATPSEA	---
MdMuslin33	MQSSTLVAIIISTLILCQVAVEA	---

Mature peptide

```
AaeKSPI        --RSDAETGVC----A--CPR-------IYMPVC------G-----SNLKTYNND----C-LLR--CEINSDLGR---ANNLRKIADQA----CDNLTDNVNDFIPQEY---
AalKSPI        IPAKEKPTAKC--NAA--CTD-------DYTPVCGGVK--GS---KDKPISFGNE----CVMQKYNCENK--------KSLTVLSQGE---CPGGGGVRLQ---------
AcKSPI         --------DQC--VAT--CRT------TPEYNPVC------G-----TDQIDYKNPGQLSC---ASMCGKD-----------VSLKHYGR---CTTTKIRGR----------
CpKSPI         -RGRGGADGLC----A--CPR-------IYLPVC------G-----SDLETYSND----C-LLR--CEVESNRGR---ALGLRKLSDGA---CDNLADNLAELPVEY----
CsKSP1         LVDLR-E-SVCEKYKEGTCTK-------EYDPVC--GS-DG------KTYSTE----CVL----CQQNRNGK-----NTVAVAFKGA--PS-----------------
DaCow-219-270  ----DEELDNC--KP---CPV-------AKPTFLC------G-----ADNRTYSSL----CRLDYHNCIHS----------TSIRIACKGF---C-P
Dsm1-23-81     ---STND-TAC--PTF--CPS-------IYKPVC--GT-DG------QNFKEFAST----CNLLSHNCRRERNSV-----QAYAATDAAW---CS
Dsm1-96-156    ---KLEV-KEC--FKP--CSM-------IYQPVC--IT-NG------KYRAELANS---CLLENFNCALQVSGAQPA---ELFRLLREEK---C
MdMuslin1      ------VDDRS-)-FRP--CGR-------IRDFVCGKYP-DG-----TMETFENQ----CEMGIKACQKG---------VEIDKLYRGE---C-GHPSEHQN---------
MdMuslin2      --------QRNC--LKP--CGR-------NLDPRCGQHK-DG-----TMETFANP----CVLQVSECQKG---------ESLMCLEDLWKVFC-L-----------------
MdMuslin3      -------ENDQC--QRL--CGR-------NIAYVCGQFS-DD-----TKKTFANS----CVMEVEACQKG---------IV----NYGT---C-EEPDPYQPTPNGL----
MdMuslin4      --------DPC--MKP--CAR-------NFDFVCGQFP-DG-----TKKAFSNP----CVMEIEACQKG---------VEITQLHRGT---C-DKPVYE
MdMuslin5      -------EEDPC--LKP--CAR-------NFDFVCGQFP-DG-----TTKTYNNP----CGMEIEACQKG---------IEIVMLHRGT---C-DKPIKD
MdMuslin6      ---------SVC--VPD--CPV-------EYKQVCGQHV-DG-----S-RKTYVNS----CRMTFEACRTG---------KPITQVNRGP---C-KK
MdMuslin7      -------ENDQC--QRL--CGR-------NIAYVCGQFS-DD-----TKKTFANS----CVMEVEACQKG---------IGNVILST---C
MdMuslin8      --------QRNC--LKP--CGR-------NLDPRCGQHK-DG-----TMETFANP----CVLQVSECQKG---------EKIRELHRGN--CANKF-----------------
MdMuslin9      -------NEKC--TRR--CGY-------KYDPVCGKTP-DG----KFITFSSE----CIMESAACVTG---------QNITKAYTGE---CTEDN
MdMuslin10     --------ELC----P--CPR-------NYDPVC------G-----SNLVTYSNR----C--EFE-CSKREVERS---GRSLGLARSGP---C
MdMuslin11     -------GEVC----K--CAR-------ILDPVC------A-----TDSQTYPNS----C--EFL-CAQKKLARR---GRSIGLAHAGS---C
MdMuslin12     ---------IYC----P--CDL-------KNKKPVC------G-----SNGITYVNR----C--EFE-CTQREYRKL---GRVLNSVKDAP---C
MdMuslin13     --------FLSC--LQS--CPS-----TMEYNPIC------G-----SDNINYHNNGRLVC---AQRCGKN----------VSALRSGI---CN-----------P-----
```
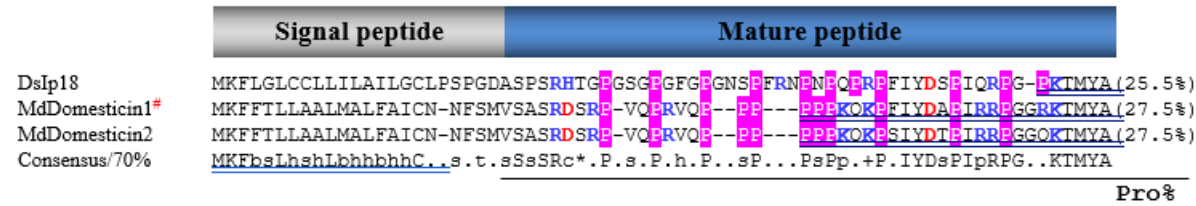
P1   β1   α   β2
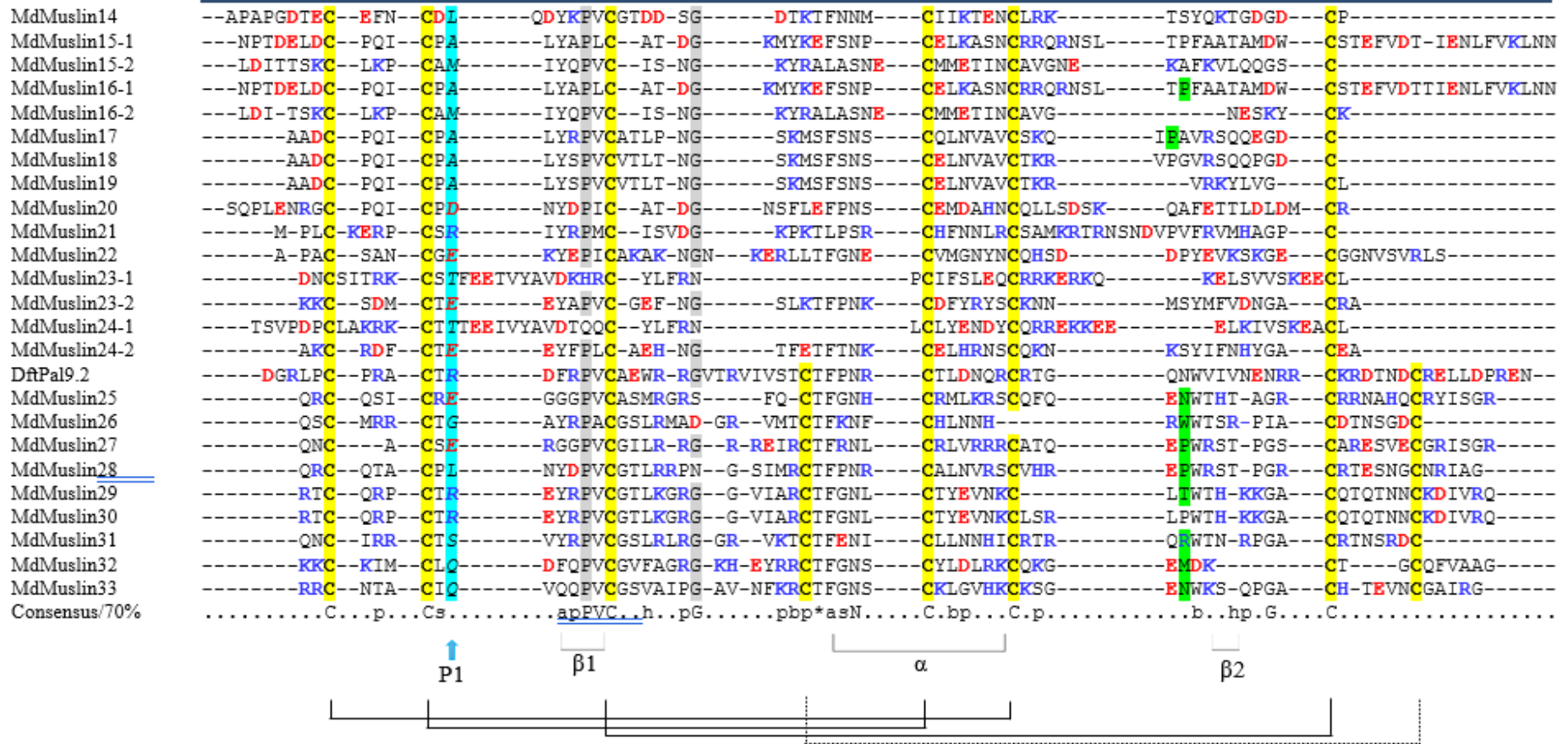
Figure S4. MSA of MdMuslins. Cysteines involved in the formation of disulfide bridges are colored in yellow. Conservation replacements are shadowed in grey. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. The P1 amino acids are italicized and shadowed in cyan-blue. Residues split by phase 1 introns are shadowed in green. α-helix and β-sheets are also indicated at the bottom together with the potential disulfides being showed with black lines, and the fourth bridge by dotted line. Aae: Aedes aegypti (GenBank: ABF18209.1). Aal: *Aedes albopictus* (GenBank: JAC06964.1), Ac: *Apis cerana* (GenBank: AGW24880.1), Cs: *Channa striata*(GenBank: CDG86164.1), Cp: *Culex pipiens pallens* (GenBank: AFN41343.1), DaCOW: *Drosophila ananassae*(GenBank: XM_001953924.3:219-270), Df: *Drosophila ficusphila* (GenBank: XP_017047190.1), Ds: *Drosophila simulans*(GenBank: XP_002105007.1).
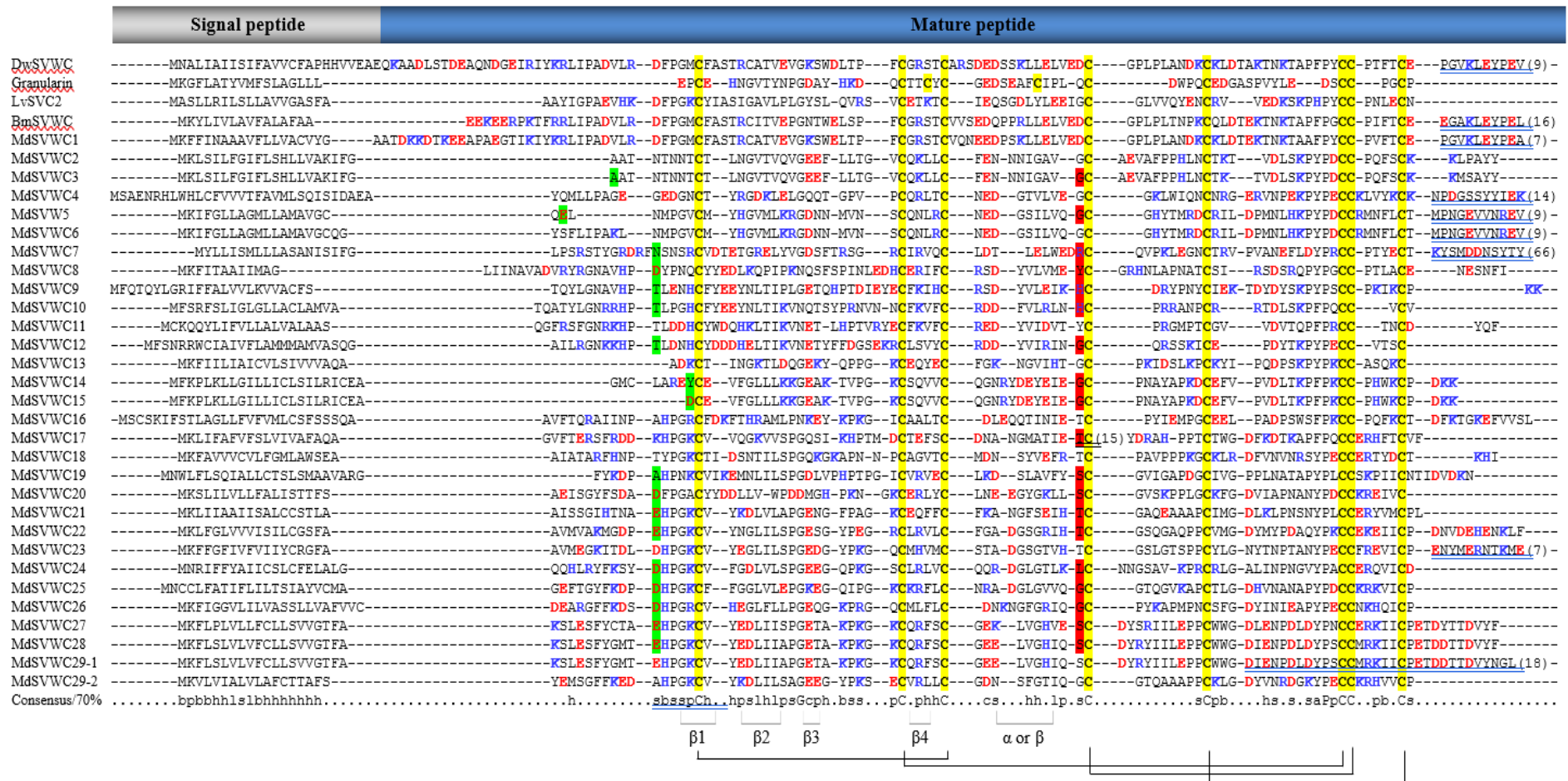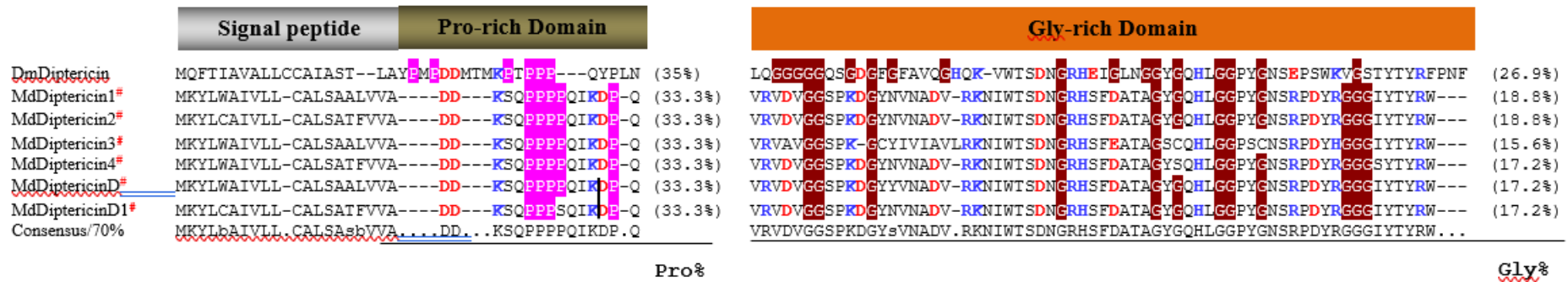
Figure S5. MSA of SVWC AMPs. Cysteines involved in the formation of disulfide bridges are colored in yellow. Conservation replacements are shadowed in grey. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. Residues split by phase 1 and 2 introns are boxed in green and red, respectively. α-helix and β-sheets are also indicated at the bottom together with the potential disulfides being showed with black lines. Bm: *Bombyx mandarina* (GenBank: XP_028031732.1), Dw: *Drosophila willistoni* (GenBank: XP_015032364.1), Granularin: *Lymnaea stagnalis* (GenBank: AAS20460.1), Lv: *Litopenaeus vannamei* (GenBank: HQ541159.1).

Figure S6. MSA of Crustins. Cysteines involved in the formation of disulfide bridges are colored by yellow. Conservation replacements are shadowed in grey. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. Phase 1 intron are shadowed in green. Disulfide bridges in the WAP domain are displayed by black lines, along with the ones in C-terminus of MdCrustin3 and 4, and ArwaprinThr1 in red. Ar: *Athalia rosae* (GenBank:XP_012251600.1), Bt: *Bombus terrestris* (GenBank:XP_020719905.1), Fb: *Farfantepenaeus brasiliensis*(GenBank: ABQ96197.1), Fp: *Farfantepenaeus paulensis* (GenBank: ABM63361.1), Mr:*Macrobrachium rosenbergii* (GenBank: ABQ41252.1), Pp: *Portunus pelagicus* (GenBank:AFN37210.1), Si: *Solenopsis invicta* (GenBank: XP_011162709.1),St: *Scylla tranquebarica* (GenBank: AFI56572.1). The Gly-rich, Cys-rich, Aromatic-rich, Pro-peptide, WAP are boxed in light green, blue, green, red and golden, respectively. The expand C-terminal domain in MdCrustin1and 2 is boxed in purple. While the C-terminal domain in MdCrustin3 and 4, and ArwaprinThr1 is boxed in orange.

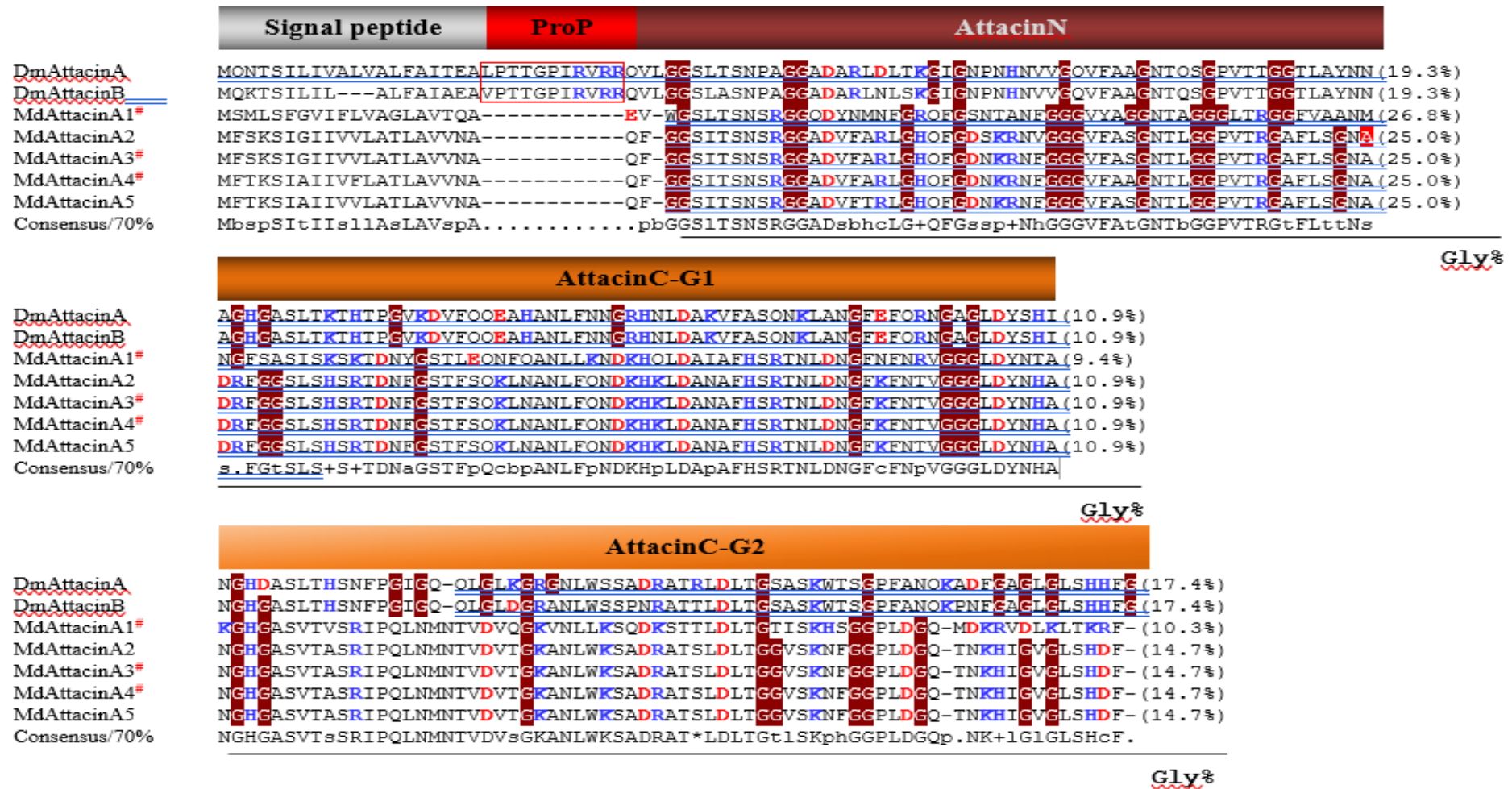|	Signal peptide	Mature peptide
DsIp18	MKFLGLCCLLILAILGCLPSPGDASPS	RHTGEGSGEGFGEGNSPFRNENEQERPFIYDSEIQRPG-EKTMYA (25.5%)
MdDomesticin1#	MKFFTLLAALMALFAICN-NFSMVSAS	RDSRE-VQERVQE--PP---PPEKQKPFIYDAEIRREGGRKTMYA (27.5%)
MdDomesticin2	MKFFTLLAALMALFAICN-NFSMVSAS	RDSRE-VQERVQE--PP---PPEKQKPSIYDTEIRREGGOKTMYA (27.5%)
Consensus/70%	MKFbsLhshLbhhbhhC...s.t.sSsSRc*.P.s.P.h.P..sP...PsPp.+P.IYDsPIpRPG..KTMYA	

Pro%

Figure 7. Domesticins. The percentage of proline in mature peptide are calculated. Proline residues in mature peptide are shadowed in pink. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. # represents the peptides previously known. DsIp18: *Drosophila serrata* (GenBank: XP_020804442.1).

	Signal peptide	Pro-rich Domain		Gly-rich Domain	
DmDiptericin	MQFTIAVALLCCAIAST--LAYEMEDDMTMKETPPE---QYPLN	(35%)		LQGGGGGQSGDGFGFAVQGHQK-VWTSDNGRHEIGLNGGYGQHLGGPYGNSEPSWKVGSTYTYRFPNF	(26.9%)
MdDiptericin1#	MKYLWAIVLL-CALSAALVVA----DD---KSQPPPEQIKDE-Q	(33.3%)		VRVDVGGSPKDGYNVNADV-RKNIWTSDNGRHSFDATAGYGQHLGGPYGNSRPDYRGGGIYTYRW---	(18.8%)
MdDiptericin2#	MKYLCAIVLL-CALSATFVVA----DD---KSQPPPEQIKDE-Q	(33.3%)		VRVDVGGSPKDGYNVNADV-RKNIWTSDNGRHSFDATAGYGQHLGGPYGNSRPDYRGGGIYTYRW---	(18.8%)
MdDiptericin3#	MKYLWAIVLL-CALSAALVVA----DD---KSQPPPEQIKDE-Q	(33.3%)		VRVAVGGSPK-GCYIVIAVLRKNIWTSDNGRHSFEATAGSCQHLGGPSCNSRPDYHGGGIYTYRW---	(15.6%)
MdDiptericin4#	MKYLWAIVLL-CALSATFVVA----DD---KSQPPPEQIKDE-Q	(33.3%)		VRVDVGGSPKDGYNVNADV-RKNIWTSDNGRHSFDATAGYSQHLGGPYGNSRPDYRGGGSYTYRW---	(17.2%)
MdDiptericinD#	MKYLWAIVLL-CALSAALVVA----DD---KSQPPPEQIKDE-Q	(33.3%)		VRVDVGGSPKDGYYVNADV-RKNIWTSDNGRHSFDATAGYGQHLGGPYGNSRPDYRGGGIYTYRW---	(17.2%)
MdDiptericinD1#	MKYLCAIVLL-CALSATFVVA----DD---KSQPPESQIKDE-Q	(33.3%)		VRVDVGGSPKDGYNVNADV-RKNIWTSDNGRHSFDATAGYGQHLGGPYGNSRPDYRGGGIYTYRW---	(17.2%)
Consensus/70%	MKYLbAIVLL.CALSAsbVVA....DD...KSQPPPPQIKDP.Q			VRVDVGGSPKDGYsVNADV.RKNIWTSDNGRHSFDATAGYGQHLGGPYGNSRPDYRGGGIYTYRW...	

Pro%   Gly%

Figure S8. MSA of Diptericins. The percentage of prolines and glycines in the proline-rich and the glycine-rich domains are calculated, respectively. Proline residues in the mature peptide are shadowed in pink, and the glycine residues in the mature peptides are shadowed in dark red. Basic (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. The position of a phase 0 intron is denoted by black line. # represents the peptides previously known. Dm: *Drosophila melanogaster* (GenBank: AAB82521.1).
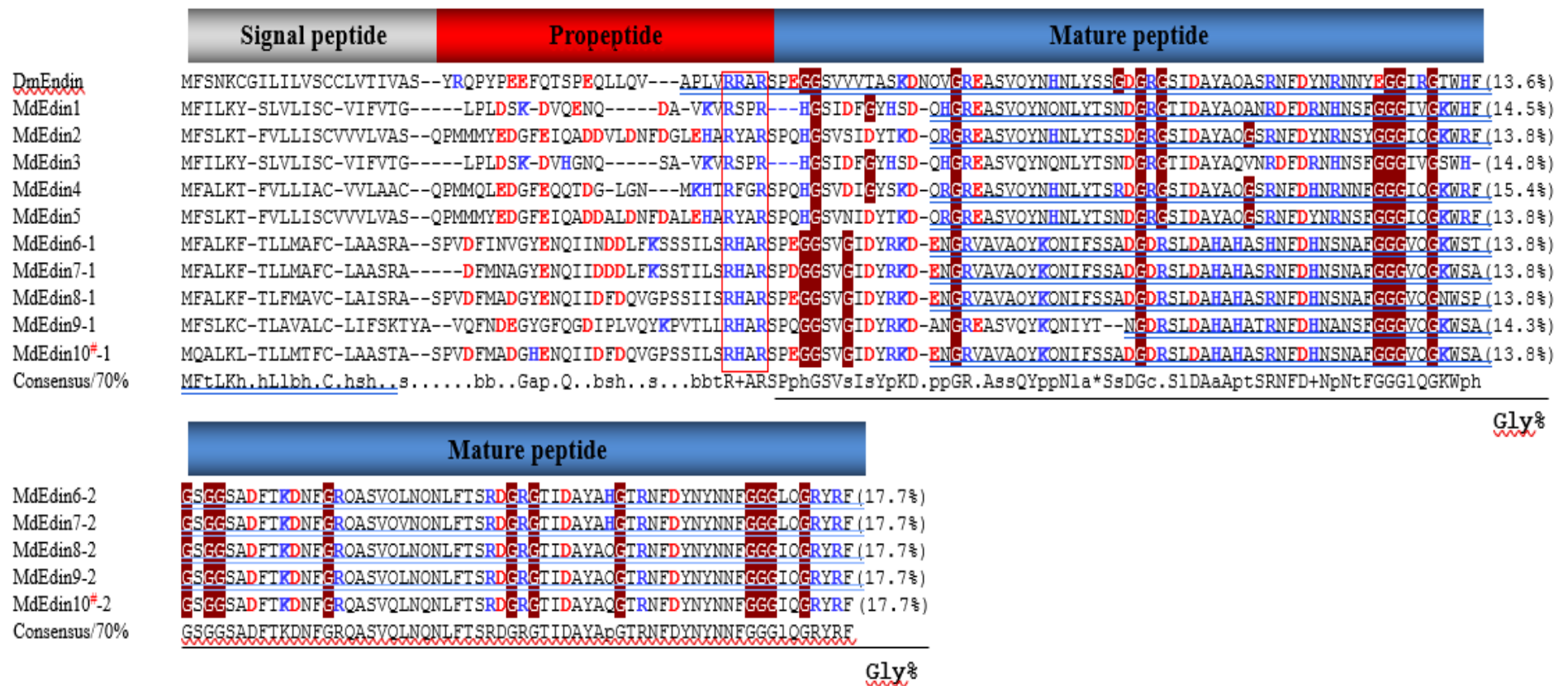
Figure S9. MSA of Edins. Glycine residues in the mature peptides are shadowed in dark red. The RXXR motif in the propeptides are in red box. Basic (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. The percentages of glycines in the mature peptides are calculated. # represents the peptides previously known. Dm: Drosophila melanogaster (GenBank: NP_730278.1).

Figure S10. MSA of AttacinA. Glycine residues in the mature peptides are shadowed in dark red. Basic (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. The percentage of glycines in the N-terminal attacin domain, G1 and G2 domain are separately calculated. The RXXR motif is boxed with red color, which does not exist in attacinA in *M. domestica*. The position of phase 2 intron is shadowed in red. # represents the peptides previously known. DmAttacinA: *Drosophila melanogaster* (GenBank: ABS52579.1), DmAttacinB: *Drosophila melanogaster* (GenBank: Q9V751.2).

Signal peptide | **ProP** | **AttacinN**

```
DmAttacinC    MSKIVLLIVVIIGVLGSLAV--ALPQRPYTQFLIYYPPPPTPPRIYRARR (33.3%) QVLGGSLTSNPSGGADARLDLSKAVGTPDHHVIGQVFAAGNTQTKPVSTPVTSGATLGYNN(14.8%)
MdAttacinC1#  -MKCILLFAACVGILAIASVTEAQPQRVYVQKLSYYPPSTRAPQIMRARR (14.8%) SPDSPPPSSGIKIKTNS--VSSNIFTSPSNKLDGTVTQSKTYFPNGQFSRSSSGSLDWKNS(6.8%)
MdAttacinC2#  -MKCILLLAACVGILAFASVTEAQPQRVYVQKLFYYPPPTRAPQLMRARR (22.2%) SPDSPPPSSGIQIKTNS--VSSNIFTSPSNKLDGTVTQSKTYFPNGQFSRSSSGSLDWKNS(6.8%)
MdAttacinC3-1 -MKCILLFAACVGILAVASVTEALPQRVYVQKLFYYPPPTQAPRVMRARR (22.2%) SPDNPAAHTGVQIKTNS--VSSNILTSPSNKLDGTVTQSKTYFPNGQFSRSSSGSLDWKNS(6.8%)
MdAttacinC3-2 -MKCIPLFAACVAILAVASVTEALPQRVYVQKLFYYPPPTQAPRVMRARR (22.2%) SPDNPAANTGIQIKTNS--VSSNILKTPSNKVDGTVTQSKTYFPNGQFSRSSSGSLDWKNS(6.8%)
MdAttacinC4#  -MKCILLFAACVGILATASVTEAHPQRVYVQKLFYYPPPTQAPQLMRARR (22.2%) SPDNPAANTGIQIKTNS--VSSNILTSPSNKLDGTVTQSKTYFPNGQFSRSSSGSLDWKNS(6.8%)
MdAttacinC5#  -MKCIVLLAACVGILAISTITDAQPQRVRVQKLFYYPPPTQAPRTMRARR (22.2%) SPDCHSRSNGISVKTNS--ASGTLLNGASNRIGGTVTQSKTYFPNGQFSRSSSGSIDWKNS(11.9%)
MdAttacinC6   -MKCILLFAACVGILAVASVTEALPQRVYVQKLFYYPPPTQAPRVMRARR (22.2%) SPDNPAAHTGVQIKTNS--VSSNILTSPSNKLDGTVTQSKTYFPNGQFSRSSSGSLDWKNS(6.8%)
MdAttacinC7   -MKCILLLAACVGILAIASVTEAQPQRVYVQKLFYYPPPTRAPQLMRARR (22.2%) SPDSPPPSSGIQIKTNS--VSSNIFTSPSNKLDGTVTQSKTYFPNGQFSRSSSGSLDWKNS(6.8%)
Consensus/70% -MKCIllbAACVGILAhASVtEApPQRVYVQKLPYYPPPTpAPplMRARR          SPDsPsss*GIpIKTNS..VSSNIbs*PSNKLDGTVTQSKTYFPNGQFSESSSGSLDWKNS
```

Pro%   Gly%

AttacinC-G1

```
DmAttacinC    HGHGLELTKTHTPGVRDSFQQTATANLFNNGVHNLDAKAFASQNQLANGFKFDRNGAALDYSHV (9.4%)
MdAttacinC1#  LGVGGSLSRDITTGHSDSFTKSLGVNIFNNDRNSLDAMYSQSHTKLNNGFGFNREGGSLKWTNE(14.1%)
MdAttacinC2#  LGVGGSLSRDITTGHSDSFTKSLGVNVFNNDRNSLDAMYSQSHTKLNNGFGFNREGGSLKWTNE(14.1%)
MdAttacinC3-1 LGVGGSLSRDITKGHSDTFTKSLGVNIFNNDRNSVDAFYSQSNTKLNNGFGINREGGSLKWTNE(14.1%)
MdAttacinC3-2 LGVGGSLGRDITKGHSDTFTKSLGVNIFNNDRNSLDAMYSOSNTKLNNGFGFNREGGSLKWTNE(15.6%)
MdAttacinC4#  LGVGGSLSRDITKGHSDTFTKSLGVNIFNNDRNSVDAFYSQSNTKLNNGFGFNREGGSLKWTNE(14.1%)
MdAttacinC5#  MGVGGSLSRDINKGHSDTFTKSLGVNIFNNDRNSVDAAYSQSRTNLNNGFGFNKESGSVNWSNI(14.1%)
MdAttacinC6   LGVGGSLSRDITKGHSDTFTKSLGVNIFNNDRNSVDAFYSQSNTKLNNGFGINREGGSLKWTNE(14.1%)
MdAttacinC7   LGVGGSLSRDITTGHSDSFTKSLGVNIFNNDRNSVDAFYSQSNTKLNNGFGFNREGGSLKWTNE(14.1%)
Consensus/70% LGVGGSL*RDITpGHSD*FTKSLGVNIFNNDRNSlDAbYSQSpTKLNNGFGFNREGGSLKWTNE
```

Gly%

AttacinC-G2

```
DmAttacinC    KGHGATLTHANIPGLGKQLELGGRANLWQSQDRNTRLDLGSTASKWTSGPFKGQTDLGANLGLSHYF-G----------(16.2%)
MdAttacinC1#  GGHGLNAGVTRFEGIGKQASIVGNTNLFTSNDGNTRLDAYGSGTKWLNGPFENQREFNFGFGGSHAFKG----------(21.7%)
MdAttacinC2#  GGHGLSAGVTRFEGIGKQASIAGNTNLFTSNDGNTRLDAFGSGTKWLNGPFENQREFNFGFGGSHAFKG----------(21.7%)
MdAttacinC3-1 GGHGLNAGVTRFEGFGKQASIGGNTNLYTSNDGNTRLDAFGSGSKWLNGPLENHREFNFGFGGSHAFKGGTSIKNYSLK(20.5%)
MdAttacinC3-2 GGHGLNAGVTRFEGFGKQASVGGHTNIFTSNDGNTRLDAFGSGSKWLNGPLENHREFNFGIGGSHYFRG----------(23.2%)
MdAttacinC4#  GGHGLNAGVTRFEGFGKQASVGGHTNIFTSNDGNTRLDAYGSGSKWLNGPLENHREFNFGFGGSHAFKG----------(23.2%)
MdAttacinC5#  GGHGATVAVNRFEGFGKQASVGGHTNLFTSSDGNTRIDAFGSGSKWLSGPMQGDKEMNVGIGGSHSFKG----------(23.2%)
MdAttacinC6   GGHGLNAGVTRFEGFGKQASIGGNTNLYTSNDGNTRLDAFGSGSKWLNGPLENHREFNFGFGGSHAFKG----------(23.2%)
MdAttacinC7   GGHGLNAGVTRFEGIGKQASIVGNTNLFTSNDGNTRLDAYGSGSKWLNGPFENQREFNFGFGGSHAFKG----------(21.7%)
Consensus/70% GGHGLsAGVTRFEGbGKQAS1tGpTNlFTSNDGNTRLDAaGSGSKWLNGPbENpREFNFGbGGSHhF+G.........
```

Gly%

Figure S11. MSA of AttacinC. Proline residues in the propeptide and glycine residues in the mature peptides are shadowed in pink and dark red, respectively. The percentages of glycines in the N-terminal attacin domain, G1 and G2 domain, and the prolines in the propeptides are separately calculated. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. The RXXR motif is boxed in red. Phase 2 intron is shadowed in red. # represents the peptides previously known. Dm: *Drosophila melanogaster* (GenBank: NP_523729.3).
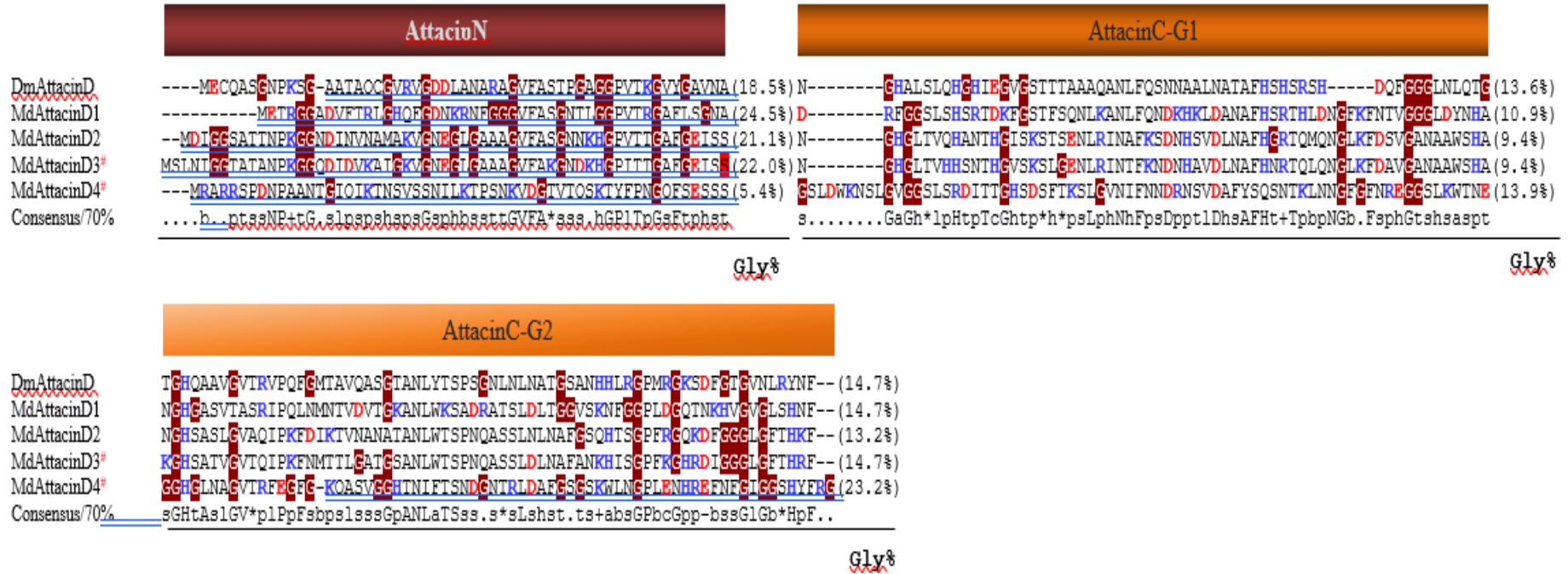
Figure S12. MSA of AttacinD. Glycine residues in the mature peptides are shadowed in dark red. Basic (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. The percentages of glycines in the N-terminal attacin domain, of G1 and G2 domains are separately calculated. The phase 2 intron is shadowed in red. Note: no signal peptide exists in AttacinD. Dm: *Drosophila melanogaster* (GenBank: NP_524391.2). # represents the peptides previously known.