

## Key to Gene Annotations

- **General organization:** The medial acid-base dyads (ABDs) are “parsed” into “strings that initiate with a dyad. N-term domain is denoted the head, C-term the tail.
- **Highlights:** Yellow, tyrosine; Green, cysteine; Pink, VPV; Gray, ABD in predicted head and tail domains.
- **Colored font:** Red, glycine residues in ABD domains; Bold-faced red, last G residue in head and first G residue in tail; Blue, alanine residues in ABD domains; Green, repeated string domains.
- **Underscores:** Acid-base triads or tetrads.
- **Commentary:** Notes on distinctive gene features (e.g. orthologues, localization patterns) in green font at top of some pages. Predicted homology domains (e.g. PDZ, coiled-coil) are given in Supplement Table 2.
- **Secondary structure:** PSIPRED predictions for a subset of proteins. Yellow, amino acids predicted in  $\beta$ -strand; pink, amino acids predicted in  $\alpha$ -helix; no highlight, amino acids predicted in disordered (random coil) domain.

## Articulin. Orthologue of PTET.51.1.G0130289.

MNNQSSKSVRFNEGHKILQPVNTNTTSTGSSKLTKVTSAIQELRTSFNNQAQ  
 KLTSSRVSDQLNESVATVTRSPGHS **C**APSY **C**QP **C**QPM **C**QPM **C**QSV **C**QPQF  
 GFQQQQMCSPQAYVVQGGSDQRHLEKEIKLKKKNKKLKEESKEEVIVSTIAE  
 TPKRQRPQQHVQDPEIDTIKREIQEMERIIKQMEQQPKQQSNSEIELYLEDNK  
 K **C**LKKM **C**KKIKSLEKELYEVGRQRDEALIIKQQLERENQEMFDRIGELESLLKV  
 ADKKVFDLTVQLERQNGYVKQLEDEVERLRKKKKKKKQIEIQ

ERVV  
 EKIV  
 EKPVEVIKTVHVNQPHQVQEVKPVVEII  
 KEVI  
 KEVPSEPKIV  
 EKIIIEIPKIE **Y****Y**QQVPQ**Y**IEVPKLQTIE **VPV**VQRIEVP **Y**EV**P****Y****Y**  
 RDVP **Y**EVI  
 KEVP **Y**EVV  
 REVI  
 KEVP **Y**EVI  
 KEVI  
 KEVP **Y**EVI  
 KEIPV **Y**IE **VPV**  
 DRIV  
 ERRVE **VPV**  
 ERIVE **VPV**  
 DRVVE **VPV****P****Y**EV**P****Y****Y**  
 ERVVEVP **Y**  
 ERIVEVPR  
 DR **Y**M  
 DR **Y**I  
 DRPV  
 DR **Y**VE **VPV**  
 ERRVEVP **Y**  
 ERIVEVP **Y**  
 EKIVEVP **VPV**  
 EKIVEVP **VPV**  
 EKIVEVP **VPV**  
 DRFV  
 ERYV

RDDAELEMLNIENRELQRII **G**IWEDRANKLENEIIKERRISDKLRFDIEELEYMVEDGR  
 SFNNQQQEQRFRYLKELKSKYESKIIIEARKGVVIKHHVIPTEVQQTVIQQPISTGISQNL  
 PQFGGQMLNQ

## Articulin. Orthologue of PTET.51.1.G0190258

Secondary structure next slide

MNNQSSKSVRFNEG HKVLQPVNTNTTSSGSSKSSKVTS AIQELRTSFNNQAQ  
 KLTSSRVSDQLNESVATVTRSPGHS **C**APSI**C**QPC**C**QPM**C**QPMYSSV**C**QPQF**C**  
 QQQPV**C**PPQAYVVQGGNDQRHLE**E**KEIKKLLKQNKLLKQSKEEIIIVSTIAETPK  
 KHRPQQHVQDPEIDTIKREIQEMERIIIRQMEQQPKQSSNSEIELYIEDNKK**C**LK  
 KM**C**KKIKSLE**E**KELYEV**G**RQ

RDEALIIKQQL  
 ERENQEMFDRI**G**ELESLLKVA  
 DKKVFDLTVQL  
 ERQNGYVKQLEDEV  
 ERLRKKKKKKQIEIQ  
 ERVV  
 EKIV  
 EKPVEVIKTVHVNQPQQVHDVKPVEII  
 KEVI  
 KEVPSEPKIV  
 ERIVEIPKIEYVYQQVPQYIEVPKVQTV**E**VPV**V**QRIEVPYEVPPY  
 RDVPYEVI  
 KEVPYEVI  
 KEVI  
 KEVPYEVIKQVI  
 KEVPYEVI  
 KE**VPV**YIE**VPV**  
 DRIV  
 EKRVE**VPV**  
 ERIVE**VPV**  
 DRVVE**VPV**PYEIPYPY  
 ERVVEVPY  
 ERIVEVPY  
 EKIVEVPR  
 DRYV  
 DRYV  
 DRYM  
 DRYI  
 DRPV  
 DRYVE**VPV**  
 EKRVEVPY  
 EKIVE**VPV**  
 EKIVH**VPV**  
 EKIVE**VPV**  
 EKIVE**VPV**  
 DRYV  
 DRYV  
 RDESELEMLNIEN  
 RELQRII

**G**IWEDRANKLENEVI**K**ERRISDKLRFDIEELEYMVEDGRAFNNQQQEQRHYL**K**ELKSK  
 YESKIIIEARKGVVVIKHHVIPSEVQQTVPTFIQQPISTGISQNLPPQFGGQKINQ



# Paramecium PTET.51.1.G0130289

1 M N N Q S S K S V R F N E G H K V L Q P V N T N T T S S G S S K S S K V T S A I Q E L R T S F N N Q  
51 A Q K L T S S R V S D Q L N E S V A T V T R S P G H S C A P S I C Q P C Q P M C Q P M Y S S V C Q P  
101 Q F I C Q Q Q P V C P P Q A Y V V Q G G N D Q R H L E K E I K K L K K Q N K K L K Q S K E E I I V S  
151 T I A E T P K K H R P Q Q H V Q D P E I D T I K R E I Q E M E R I I R Q M E Q Q P K Q Q S N S E I E  
201 L Y I E D N K K C L K K M C K K I K S L E K E L Y E V G R Q R D E A L I I K Q Q L E R E N Q E M F D  
251 R I G E L E S L L K V A D K K V F D L T V Q L E R Q N G Y V K Q L E D E V E R L R K K K K K Q I E  
301 I Q E R V V E K I V E K P V E V I K T V H V N Q P Q Q V H D V K P V E I I K E V I K E V P S E P K I  
351 V E R I V E I P K I E Y V Y Q Q V P Q Y I E V P K V Q T V E V P V V Q R I E V P Y E V P Y Y R D V P  
401 Y E V I K E V P Y E V I K E V I K E V P Y E V I K Q V I K E V P Y E V I K E V P V Y I E V P V D R I  
451 V E K R V E V P V E R I V E V P V D R V V E V P V P Y E I P Y P Y E R V V E V P Y E R I V E V P Y E  
501 K I V E V P R D R Y V D R Y V D R Y M D R Y I D R P V D R Y V E V P V E K R V E V P Y E K I V E V P  
551 V E K I V H V P V E K I V E V P V E K I V E V P V D R Y V D R Y V R D E S E L E M L N I E N R E L Q  
601 R I I G I W E D R A N K L E N E V I K E R R I S D K L R F D I E E L E Y M V E D G R A F N N Q Q Q E  
651 Q F R H Y L K E L K S K

## Articulin

MSQIHNDLGIKSIDELRQAFKQVQLPNKQPQQQQSQQQSPQKEVERVIETT  
 APTSSHQQIHCHPICFPSHICHPINYQPFIIVQEQPVNERKSNKKKKKYVKEVQ  
 EEIQVQIPVKKSQRKIEQPNFNVQREVDHLEQLIRQVKIQEQPAQVLEYHISVP  
 EQEKPKDNTEMIAMLSALKQEPIQLPKDDYDKIENARLQEEIKNTKISYHELVS  
 KYEELQKRKRK

EKIV  
 KEIEY  
 KEIPKIVEVP  
 RDVV  
 REVVKPFEVI  
 KEVI  
 KEV  
 EKPSQMNT  
 KEIPIYVPQY  
 KEVTVN  
 KEVPVY  
 KEVPV  
 EKEVKVY  
 KDIPVYKVQPVYYDVPVY  
 RDVPIYQKVPVY  
 KEVPVYSEPINVY  
 REIPII  
 KEV  
 DRVPE

TIEHPNLSRTRRVISPSQNGRSYIRSPYSRIEY

6 *Paramecium* 4223 epiplasmin family1 CAI39404.1

MSQRPPVPQNQAPQPQQFQPAPTYAPQYAPQYAPTYAPAYAPTYAPTYAPQY  
APAPIAPLTYSVARPVAPVAPVAPVVAQPVIAQPVVAQPVLQQSVIAQPVVQQS  
VHATIKGESRIEYIPYQKAVMEYEEQEYVQYVPR

ERKVTDY<sup>Y</sup>AVE<sup>Y</sup>QTE<sup>Y</sup>VPQVFQ  
EK<sup>Y</sup>TE<sup>Y</sup>VPV  
DR<sup>Y</sup>Q  
ERVE<sup>Y</sup>Y<sup>Y</sup>PV  
ERQVVH

QQVVQQPVVQQVVQQPVVQQVVQQPVVQQVVQQVVPQPVVQQVVQQPLSVVQPVQT  
VPLTYAPQYAAPIVSSRVIPSYPQYPSYPQYQQAPQQHQAPQQPPRSNLNVH

MSQRPPVPQNQAPQQPQAPQYHPQAPQYAPQYAPQYAPQYSPAPLATYPAQY  
 APATYAPAYAPVAPLTYSVARPVAPVVAQPVVQAPVLQQSVIAQPVVQQPVHAT  
 IKGESRIEYVPYQKAVMEYEEQEYVQYVP

RERKVTDYYAVEYQTEYVPQVFQ  
 EKYTEYVPV  
 DRYQ  
 ERVEYYPV  
 ERQVVH

QQVVQQPVVQQVVQQPVVQQVVQQPVVQTVAPQPVVQQFVQQPVSVIQPVQT  
 YPVQYAAPVSSRVIPSYPTYPQYPQYHPAPQQVQQQPPRSNLNNNI

1 M S Q R P P V P Q N Q A P Q Q P Q A P Q Y H P Q A P Q Y A P Q Y A P Q Y A P Q Y S P A P L A T Y P A  
 51 Q Y A P A T Y A P A Y A P V A P L T Y S V A R P V A P V V A Q P V V Q A P V L Q Q S V I A Q P V V Q  
 101 Q P V H A T I K G E S R I E Y V P Y Q K A V M E Y E E Q E V V Q Y V P R E R K V T D Y Y A V E Y Q T  
 151 E Y V P Q V F Q E K Y T E Y V P V D R Y Q E R V E Y Y P V E R Q V V H Q Q V V Q Q P V V Q Q V V Q Q  
 201 P V V Q Q V V Q Q P V V Q T V A P Q P V V Q Q F V Q Q P V S I V Q P V Q T Y P V Q Y A A P I V S S R  
 251 V I P S Y P T Y P Q Y P Q Y H P A P Q Q V Q Q Q P P R S N L N N N I

8

*Paramecium* EPI2 epiplasmin family 2 CAI28404.1

MSNIPQSQHPQQAQQPPQAPQPYYGQPSYAQPYGAPLSPLRYSYAPPVVQQV  
VPQTYVPQQVVPQTYVPQ QVVAQPVVAQPVVAQSVVAQPTIKGESRIEYVPY

EKTVLEYEEVRQRIQVP  
REKFVTDYYAVEYQTEYVPQVFQ  
EKFTERYVPV  
DRYQ  
ERVEYYPV  
ERQVVH

QQQVQQVVAQPVQQVVTQSVVQPVQYAPQPVQYVQQPVQYAPQPVQYASQPV  
QYAPAPLQQTTYVPAPVASLPLAQTQVPTRTVQARPQQPLDRTQAQNPRPQPA  
QQPQQKQKSFLDRLFDRD



MSNHPASTKPPQKTAPAPQQPIPYSPRRTYAPPVAFASPSYLPQQSFVGGAP  
 VQYVPQPVAVQPVAVQPVAVQPAQQVIKGESRIEYIPY

EKSVIEYEEVRQKIQVPR  
 EKYTEYQAVEYQTDYIPQVFY  
 DKVTEYVPV  
 DRFQ  
 DRVEYYPV  
 ERQVVH

QPVQQVVAQPVVQSVVQQVPQYVAPVQSVVQPVYQQPQISYAPYVQPNFAPS  
 RIAPVSYAPPLSYGAPVSHPRRYVNHIVINQSPPAKPPAQKPPQKEKKTFLD  
 NIFS

1 M S N H P A S T K P P Q K T A P A P Q Q P I P Y S P P R T Y A P P V A F A S P S Y L P I Q Q S F V G  
 51 A P V Q Y V P Q P V A V Q P V A V Q P V A V Q P A Q Q V I K G E S R I E Y I P Y E K S V I E Y E E V  
 101 R Q K I Q V P R E K Y I T E Y Q A V E Y Q T D Y I P Q V F Y D K V T E Y V P V D R F Q D R V E Y Y P  
 151 V E R Q V V H Q P V Q Q V V A Q P V V Q S V V Q Q V P Q Y V A P V Q S V V Q P V Y Q Q P Q I S Y A P  
 201 Y V Q P N F A P S R I A P V S Y A P P L S Y G A P V S H P R R Y V N H I V I N Q S P P A K P Q P A Q  
 251 K P Q P Q K E K K T F L D N I F S

*Paramecium* EPI38 epiplasmin family 2 CAI39389.1

MSNRPPTQQPGQPATTAPQYQPQWNPTFSPPRAYQAPVQLQSPAYIPQYYTQ  
 PVAQSYVAPPVVQQPIVYQQPVVTQSVVAQPVVTQPVVQQQPIIKGESRVEYIPY

EKSVIEYEEVRQKIQVP  
 REKYVTEYQAIEYQTEYVPQVFY  
 DKVTEYVPV  
 DRYQ  
 ERVEYYPV  
 ERQVVH

QPQIQTVAQPVVQQSYVAQPVVVTQSYVAQPVVQQSYVTQTVPVVSQPIQYAQYP  
 QYAQPATQTIYGPVPIGNGPIPASQLPAAVKTGQFQQTPQK

1 M S N R P P T Q Q P G Q P A T T A P Q Y Q P Q W N P T F S P P R A Y Q A P V Q L Q S P A Y I P Q Y Y  
 51 T Q P V A Q S Y V A P P V V Q Q P I V Y Q Q P V V T Q S V V A Q P V V T Q P V V Q Q Q P I I K G E S  
 101 R V E Y I P Y E K S V I E Y E E V R Q K I Q V P R E K Y V T E Y Q A I E Y Q T E Y V P Q V F Y D K V  
 151 T E Y V P V D R Y Q E R V E Y Y P V E R Q V V H Q P Q I Q T V A Q P V V Q Q S Y V A Q P V V T Q S Y  
 201 V A Q P V V Q Q S Y V T Q T V P V V S Q P I Q Y A Q Y P Q Y A Q P A T Q T I Y G P P V P I G N G P I  
 251 P A S Q L P A A V K T G Q F Q Q T P Q K

11 *Paramecium* EPI20 epiplasmin family 2 CAI39407.1

MSNVPNSHHPPQAPKQGPPYQPTQFQPGFAPQYAPAPVTYGPPLTSSPLRYS  
QPLYQPSVVAQPVYAPPVVQQPVYAPSVVQQPVVTQSVVAQPVVAAQPIK**G**E  
SRIEYVPY

EKTVLE**Y**EEVRQKIQVP  
RER**Y**VTD**Y****A**VE**Y**QTE**Y**VPQVFQ  
EK**Y**TE**Y**VPV  
DR**Y**Q  
ERVE**Y**YPV  
ERQVVH

QPAQQVQQVVAQPVVQQPVQVVQQPVQYVQQPVVQQPLVQSIPVQTVRPPVYAP**G**  
SLPLGQTVSPRLPPQAQTKPQAQLPQTKQPQQQKTKSFLDRLFDRD

*Paramecium* EPI18 family2 CAI39409.1

MSNIPPSQHPPQAPQPGPYQQPTFQPGFAPQYAPAPVAYGPPLTQSPLRYSQ  
PLYQAPVVQQPVYAQPVVQQPVYTQPVVQQPVYAQPVVQQPVYAQPVVQQP  
IVTQSIVAQPQPVVAAQPIQGESRIEYVPY

EKTVIEYEEVRQKIQVP  
KEKYVTDYYAVEYQTEYVPQVFQ  
EKFTYVPVDRYQ  
ERVEYYPV  
ERQVVH

QQVQPVQQVVAQPVVQQPVQYVQQPVQVVQPQPVQVVQQQPVYQQQPVYQQ  
PLVQSIPVQTVRPPVYG PATTLLPLGQTVSPRPLGTTVPAKPLDKTQGPKQPAQQQ  
NKQKSFLDRLFDRD

MAEYQPQGIVQQPQLYGTPQTITGQPQFVRPIVLGNSTYQQQLVTNQNKGQS  
GFYQGGGVLVNQPVLGASAISSNIVSTGQAVK**G**ESRIEYIPY

EKTIME**Y**EEVRRQ**Y**QVPITRQITE**Y**QAIQ**Y**ETE**Y**IPQVIQ  
EKVIE**Y**MPV  
EKFA  
ERVE**Y**Q

TVTRQNVLQNTVQQVQQQTQIQPIVTTQTYQTTTPIVQTVQQPIVTQVAQPINVPQ  
TYYSTYQ

MANLPPQQYSPPSNLPTNNREPQTFQQPTIGQPQVARSNLPIQSLPDTALQYLPQQIAQQ  
 PYSSQPLQQPLYNQLQPTKLGPQQPIFGQSQIQQIAQPPQEGNVVKGQSRIEYIPYERTI  
 TEYEEVRRQVQVPITKQVTDYYAVQYDIEYIPQVIQ

EKQIEYVPV  
 ERVAERTEYYTV  
 EKQNVIIQQPIGYQSIIQTNYIPEQFQYNQII  
 ERQSALAYQPYQQQTQIQT  
 REVTQYPV

AQQQQYIQQLPVQQSVPLPAPQAQYLPTQYQQYASPIGIQQPQYQIQQTVPLNYGQQFQA  
 SIIQPAPKQPIATATVPQGYQIPTQTVPQQYINQQVEQDPTIHSKRPPQQQLAQTAYNPQL  
 QQTVGPNQFAQSVPQQQSLQQDLGRTRPYQQQQQPQPQQQQLPQQQQGVPQKPNKEK  
 SFLEKLF D

```

1 M A N L P P Q Q Y S P P S N L P T N N R E P Q T F Q Q P T I G Q P Q V A R S N L P I Q S L P D T A L :
51 Q Y L P Q Q I A Q Q P Y S S Q P L Q Q P L Y N Q L Q P T K L G P Q Q P I F G Q S Q I Q Q I A Q P P Q :
101 E G N V V K G Q S R I E Y I P Y E R T I T E Y E E V R R Q V Q V P I T K Q V T D Y Y A V Q Y D I E Y :
151 I P Q V I Q E K Q I E Y V P V E R V A E R T E Y Y T V E K Q N V I Q Q P I G Y Q S Q I Q T N Y I P E :
201 Q F Q Y N Q I I E R Q S A L A Y Q P Y Q Q Q T Q I Q T R E V T Q Y P V A Q Q Q Q Y I Q Q L P V Q Q S :
251 V P L P A P Q A Q Y L P T Q Y Q Q Y A S P I G I Q Q P Q Y Q I Q Q T V P L N Y G Q Q F Q A S Q I Q P :
301 A P K Q P I A T A T V P Q G Y Q I P T Q T V P Q Q Y I N Q Q V E Q D P T I H S K R P Q Q Q L A Q T A :
351 Q Y N P Q L Q Q T V G P N Q F A Q S V P Q Q Q S L Q Q D L G R T R P Y Q Q Q Q Q P Q P Q Q Q Q L P Q :
401 Q Q Q G V P Q K P N K E K S F L E K L F D

```

MNNPRPSQLPAQKATPSQLPPTSQQQVPETYQAQPNWVPASSQIQPGVLSQP  
IPLQQTGLIQNPQYLQQPVYAQQQLIQSSAPVVTQGNAIKGESRIEYIPYT

KEVTEYVTQEVVEYVP  
RERKITDYYAVEYVTEHIPQVIQ  
EKYIEYVPVETI  
KERTEYQAVTKQSVVQAPIDY

QQIQKTQQYQVAAPIQYAQTYTQPIQYAQTTVQPIQYAQTTTQFVPTTQSQYVSAP  
VTTSQFVSQPYPTTTYGAPITNYGQLTQNTAYLPAATYAQPGIGQVTTTTQQYSTG  
WQQVYPASTGNFQPQLQQQQVGGQLLIQGGQSPKYA

*Paramecium* EPI51 epiplasmin family 5  
GSPATP00033927001

MSYYKPFYQSAYDRTAPVYPPSQYGPAYYGPYDRPYSYQSRAPTRBEQWSEY  
IPVEQRYTDYVPETKVEYRPV

EKSYTDYIEVKHETDYVPVPRL  
EKRVEYIPV  
DRYDEHVDYVPVQ

NSHVVKGPQSRAGYGYQSQYLPPPPAPTSSNYRYSRVSRYRPGATFYGY  
RYL



*Paramecium* EPI48 epiplasmin family 5  
GSPATG00018713001

MSYLUSPURSYARPYQYSPTRSYVSSPYYTSYPPR**G**NSWV

ERIPVEQRYTEYVPEQRIEY

KPV

ERRYTDYVEIEHY

RDYVPVPR**L**

ERRVEYVPI

ERYDEAVDYVPV

ERSSVVRQPL

SNSLAYSRYSRYP**S**RYYY

## Articulín. First half of sequence

MKKNTKHKQNYRKNSGSSMEVPVEIPYQVEI

EKIV  
 EKIVQKPVVQ  
 EKIV  
 EKIVEVPVEIPYQVEI  
 EKIV  
 EKPVIQ  
 EKIV  
 EKVVEVPVEVPYVEV  
 EKIV  
 EKIV  
 EKPVIQ  
 EKII  
 EKIV  
 EKVVEVPVEVPYEVQV  
 EKIK  
 EKPVIQ  
 EKIV  
 EKVVEVPVEV  
 EKIV  
 EKPVIQ  
 EKII  
 EKIV  
 EKVVEVPVEVPYEVER  
 EKII  
 EKIV  
 EKPVVQ  
 EKIVKQIV  
 EKVVEVPVEIQI  
 EKIV  
 EKPVIQ  
 EKIV  
 EKIVEIPVEVPYQVEI  
 EKIV  
 EKPVVQ  
 ERIV  
 EKIV  
 EKIV  
 EKVVEVPVEVPYEVER  
 EKII  
 EKIV  
 EKPVVQ  
 EKIVKQIV  
 EKVVEVPVEIQI  
 EKIV  
 EKPVIK  
 EKIVEVPVEVPYEI  
 EKIV  
 EKPVIK  
 EKIVEIPVEVPYEI  
 EKIV  
 EKPVIK  
 EKII  
 EKIVEVPVEVPYQVLV  
 EKIV

## Second half of sequence

EKPVIK  
EKIVEVPVEVPYQVEV  
EKIK  
ERPQQ  
EKIV  
EKIV  
EKVVEVPVEVPYKVEV  
EKIIKKIV  
EKVVEVPVEIII  
EKPA  
EKIV  
EKVVEVPVEVIV  
EKMVEVPQKKQQEFKKNQPNQQLS  
EKAPQII  
EKIVEVPVEVPYKVEV  
EKIV  
EKVVEVPVE  
  
LKLKKRQSKSPQNIQYI

*Pseudocohnilembus persalinus* KRW98423.1  
First half of sequence

Articulin. VYV motifs.

MV

EKL  
 EKEIIV  
 DRPVEVIKNVYV  
 EKEVEVPVYI  
 EKIV  
 DRPYEVIKTV  
 EKAIIEVPIEVPVIEVPVYEVIV  
 EKRVEVPYETKVYI  
 DKLVEIPIEVPVIEVPVYEVIRVEVPV  
 EKRIEIPV  
 EKTVYV  
 DKLVEIPIEVPVIEIPYEVVV  
 EKRVEVPV  
 EKTVYV  
 DKYV  
 DRLVEIPIEVPVIEIPYEVIRVEVPV  
 EKRIEIPV  
 EKTVYV  
 DKYI  
 DRLVEIPIEVPVIEIPYEVVV  
 EKRVEVPV  
 EKRIEVPV  
 EKTVYV  
 DKYI  
 DRLVEIPIEVPVIEIPYEVIV  
 EKEVFRTV  
 EKEIEVPVYI  
 EKIV  
 EKPV  
 EKIV  
 EKIVEV  
 EKFIIV  
 EKPVRQQPPQ  
 EKIVEVFFV  
 DREIEVPVYI  
 EKIV  
 DRIVEV  
 EKRVEVPV  
 EKIV  
 ERIVEVPIEVPVIEVPVYQIETAKQVPTIQYQ  
 DRIV  
 EKIV  
 ERIVEIPIEVPVIEVPVYEIETARRAQPQEIQ  
 ERIV  
 EKIV  
 EKRVEVPV  
 ERIV  
 ERIVEVPIEVPVIEVFFV  
 DRIV  
 EKIV  
 EKRVEVPV  
 DRIV  
 EKIV  
 ERIVEVPIEVPVIEVPVYEIETARRAQPQEIQ

*Pseudocohnilembus persalinus* KRW98423.1  
Second half of sequence

ERIV  
EKIVEVPV  
EKIV  
ERFIEIPIEVPVEVPYEIEAARRAQPQEIQ  
ERIV  
EKIV  
EKPVYI  
EKEVEVPVYV  
DRIV  
EKIVEI  
EKIV  
EKIVEVPVEI

FKEVESKQAANFQTQYVEVEVEVPVYVDRIVEVPVEKIVEVPVEKIVEVFRE  
YPNNIQQAPQVIKEYVEVEVEVPVYVDRIVEQTIENQIFSDNSERITQLQQQ  
NGQLQAYSDELKKNLSQLSQKISSLYEQIETSKTSINQTQSKCENCQDMTRL  
VLAGEIESLRIFDQTEVVLIEPRARSSQYVMTGMTGEIDGRASQYLQPGKD  
YKSPIPRASFTNNRYSQYNNNQRVVVSPNPRDSSRERLLMGSQVQRTSHTI  
TRSQQANMKNSQQANVFSHTNLNSNIQQQQRQQFQRNL

```

1 M V E K L V E K E I I V D R P V E V I K N V Y V E K E V E V P V Y I E K I V D R P Y E V I K T V E K
51 A I E V P I E V P I E V P V E V P Y E V I V E K R V E V P Y E T K V Y I D K L V E I P I E V P V E V
101 P Y E V R V E V P V E K R I E I P V E K T V Y V D K L V E I P I E V P V E I P Y E V V V E K R V E V
151 P V E K T V Y V D K Y V D R L V E I P I E V P V E I P Y E V R V E V P V E K R I E I P V E K T V Y V
201 D K Y I D R L V E I P I E V P V E I P Y E V V V E K R V E V P V E K R I E V P V E K T V Y V D K Y I
251 D R L V E I P I E V P V E I P Y E V I V E K E V F R T V E K E I E V P V Y I I E K I V E K P V E K I V
301 E K I V E V E K F I E V E K P V R Q Q P Q E K I V E V F V D R E I E V P V Y I E K I V D R I V E V
351 E K R V E V P V E K I V E R I V E V P I E V P V E V P Y Q I E T A K Q V P T I Q Y Q D R I V E K I V
401 E R I V E I P I E V P V E V P Y E I E T A R R A Q P Q E I Q E R I V E K I V E K R V E V P V E R I V
451 E R I V E V P I E V P V E V F V D R I V E K I V E K R V E V P V D R I V E K I V E R I V E V P I E V
501 P V E V P Y E I E T A R R A Q P Q E I Q E R I V E K I V E V P V E K I V E R F I E I P I E V P V E V
551 P Y E I E A A R R A Q P Q E I Q E R I V E K I V E K P V Y I E K E V E V P V Y V D R I V E K I V E I
601 E K I V E K I V E V P V E I F K E V E S K Q A A N F Q T Q Y V E V E V E V P V Y V D R I V E V P V E
651 K I V E V P V E K I V E V F R E Y P N N Q I Q Q A P Q V I K E Y V E V E V E V P V Y V D R I V E Q T
701 I E N Q I F S D N S E R I T Q L Q Q Q N G Q L Q A Y S D E L K K L N S Q L S Q K I S S L Y E Q I E T
751 S K T S I N Q T Q S K C E N C Q D L M T R L V L A G A E I E S L R I F D Q T E V V L I E P R A R S S
801 Q Y V M T G M T G E I D G R A S Q Y L Q P G K D Y K S P I P R A S F T N N R Y S Q Y N N N Q R V V V
851 S P N P R D S S R E R L L M G S Q V Q R T S H T I T R S Q Q A N M K N S Q Q A N V F V S H T N L T N
901 S N I Q Q Q Q R Q Q F Q R N L

```

*Pseudocohnilembus persalinus* KRX09694.1  
First half of sequence

Articulin. VYV motifs.

MRSSIINGQQQLVSSVNRGSFFQRQTLVSQSPKRLVSATNYTVVPSQQIL  
TGSVKSVTQNTSSQQLVQIISPVKKVVQQVQQLQQSSPNTITHSHT

EKVI  
EKPV<sup>YY</sup>  
EKEV  
EKVI  
EKPV<sup>YY</sup>  
EKEV  
EKVI  
EKPV<sup>YY</sup>  
EKIV  
EKPVEV  
EKIV  
EKIVE<sup>VPV</sup>  
EKI<sup>VYV</sup>  
EKPVEI  
EKI<sup>VYV</sup>  
DKIV  
EKIIEVPIE<sup>VPV</sup>EVP<sup>Y</sup>E<sup>VPV</sup>EVQKRIE<sup>VPV</sup>  
EKI<sup>VYV</sup>  
DKIVE<sup>VPV</sup>  
EKT<sup>VYV</sup>  
DK<sup>Y</sup>I  
EKLVEIPIE<sup>VPV</sup>EVPYEVIV  
EREKRVEIPV  
EKT<sup>VYV</sup>  
DKVVEVPQYI  
EKLIEIPIE<sup>VPV</sup>EVP<sup>Y</sup>EVIV  
EREKRVEIPV  
EKT<sup>VYV</sup>  
DKVVEVPQYI  
EKLIEIPIE<sup>VPV</sup>EVP<sup>Y</sup>EVIV  
EREKRVEIPV  
EKT<sup>VYV</sup>  
DK<sup>Y</sup>I  
DKLVEIPIE<sup>VPV</sup>EVP<sup>Y</sup>EVMV  
EREKRVEIPV  
EKT<sup>VYV</sup>  
DKIVEIEIPK<sup>Y</sup>I  
EKLVEIPIE<sup>VPV</sup>EVP<sup>Y</sup>EVIV  
EREKRVEIPV  
EKT<sup>VYV</sup>  
DK<sup>Y</sup>I  
DRPV  
EKT<sup>VYV</sup>  
DK<sup>Y</sup>V  
DRLVEIPIE<sup>VPV</sup>EVP<sup>Y</sup>EVIV  
EREKIIIPV  
EKT<sup>VYV</sup>  
DK<sup>Y</sup>V  
DRLVEIPIEVPEVP<sup>Y</sup>EVIV  
ERERLV

*Pseudocohnilembus persalinus* KRX09694.1  
Second half of sequence

EKRVEIPV  
EKT~~YYV~~  
DKYI  
DRLVEIPIE~~VP~~EIP~~Y~~EVVV  
EKERRVEIPV  
EKT~~YI~~  
DKFI  
EKLVEIPIE~~VP~~EVP~~Y~~EVIV  
EREKLI  
EKRVEIPV  
EKT~~YYV~~  
DK~~YV~~  
DRLVEIPIE~~VP~~EIP~~Y~~EVVV  
EKERRVEIPV  
EKT~~YV~~  
DK~~YV~~  
DRLVEIPIE~~VP~~EIP~~Y~~EVLV  
EKRVE~~VP~~  
EKT~~YQ~~  
ERIVEIPIEIPVEVP~~Y~~EV

LIERNIVDENRVKQVQAELEEWRKFD~~M~~KLVEENRYRE~~Q~~IRLLELEV~~K~~DW~~M~~RK  
YEEVNTETHYEVVSVKSRKSIYNSSTS~~Q~~LNLNRSQVRAPEQVIDAVKRASLY~~G~~  
DMKNSHTYNSNSNLHQQQY~~T~~NSQVQNGVT~~L~~SYQNSSQSQT~~T~~YVNGGSNIK  
QMSPNYRTTQSRQFIVGNNQ

## Articulin

MIPNTALQAGQIGSGVGPFGAGQFQGVPGFASPQYNFALQGQLQAALQAAGTPVGAIPGLL  
 GSAAAGLGFANLIPQQVANPALSVSAPVPNQSHQFFAENQRLQGQLQRGNALISELKALVTQ  
 LASQPPQVVNIPQVV

DRVVPVQQVIDQIVP  
 REV  
 EKPFPPVDVPV  
 EKIV  
 ERQVPFPV  
 ERKVDVPYVVT  
 RDVEVPRYV  
 DKEITVMKYVDVPVDVPIYVPVYQ  
 DKVVEVPVYV  
 DKIMEVPV  
 DKPVYVQ  
 RDIIV  
 ERPIII  
 EKKVEIPV  
 ERQVVV  
 EKPIIVEV  
 ERLI  
 EKPVY

STKTVDIPIEHERDVLTSVVDQPVNYQVVVPKVVDTPVEVPVNVVPVDVPIEVPVPVDRDVPV  
 PFQLNIDVPVDVPVARPVPVERIIQQPIPLEQPRLVEQHVPVPHVVPVAQEVIVQQPFAVPQPY  
 TVQQEVPIPHVVPVQPYAVPQPVPVPTPVAVPQPVTVQQPVVTIQQAVPVAQPVGVQSTVQI  
 PLQQSGYVQRSIGLAPQGFSPVGVQGTFLPQNAVYAGNVGFGQGGFVGGPAAFNVGQAG  
 LIGTPGRMTAQNVAGFGQFGGAQLQNGSF



## Articulín

MAQVVGTPGRVIGGNAFGANLIGNRVGVAPANVLGAQQLGLNRFGTPLGAGV  
VNANAFQLNGLIGARGLANSVVQASSASAGNIAAAQLQRSLVGAGSRVGIAPLA  
ANQALLTNEFANWRARASTVNGLIATPPAQLGLVDLQARALPYGAFQAQVPQQ  
VVNLPQVV

DRVVPVPHVI  
DRIVPRAVDTPYQVDVPV  
ERIV  
DRPVQFPV  
DRPYDVPYVVT  
RDVEVPRVV  
DKPIAVPRYVDVPFDVPIVPRYNDIIVEVPVYV  
EKRVEVPV  
DKPIYVQ  
KDVVV  
EKPVII  
ERKVEVPV  
ERQIVVPKPVYVEVERIVEVPVY

SQRVVDVPIEHERSVLLTSIVDQPVNRPVTVPKVVDTPFEVPVNVVVDVPVQIPIQV  
DVERPVPVPFNVDVPVDVPIQRPIVERVFHNPVPIEQPRIIDQPIPFQHPVPVPQP  
VTVQQGVVPVPQPVRVPQVGVIPQAVPVPHPVAVPQPVAVPQPYAVEQPYAVQQQ  
VRVQEPVAVPNPVAVPQPYAVPQPYAVQQPVVRVQEPVPVGVQTVVQPQFAVGV  
QTVGLQQGVIGVSPRIGVQNINAGRFPVNVNGGLGLAGVAGVQNLAGIQNFAGIQNL  
AGIQNFAGLQGVAGVQNLASLRAVNGLAGQVNLAASSRVAAAPGLINNSLVLAGGVL  
GNSRVL

## Articulin

MAFVNNNQLLNSRLLNGQVVNGQVLNAAARPGFVGAQVLPGQNVFGSNVLGA  
 RAQFGVGVGVSVDSPRIGQQVLPQGFVGSRVGLQVPLHPQFFNAPVPQTLVQQ  
 TVVENIVEQPQIVENPVEQIF

ERPVPVAVP  
 REVTIQRDVPVPV  
 DRPV  
 ERPVPEEQIIQ  
 KELLI  
 ERPAFFEEVI  
 ERPVPRPVYVEQII  
 DRQI  
 EKPVVIEQIVEVPV  
 ERPVYIEVENPI  
 EKEVVV  
 EREIPVPV  
 ERQVRIEQPVYHFI  
 EKPVEVPVLFENIVQ  
 RDIHI  
 DRPIY

SQQAVPRPVDFPVHAEQVVQRPVEVPRQYPVQVPRPVPAPVQVPRDVAVPVPVER  
 QIPIERPVEVPFAVDRYVDVPVPVDVPVPIGRPVPQPVQVPQPYQVIQPVAVPQPYHV  
 PEPVPAQPYQVPQPVPVPQAVPVPHPVVPVPQPTQYIEQVPVVERVPVPHNVPVPQ  
 PVAVPHPVVPVVEQVPVVEKVPVPVYNNFYAQPQVVGANFRVNQVAGAFPVQGYGIQ  
 GVQGIQGIQGIQGFVGVAGLNRAVGLNVANAGLLGSSIIGRAL

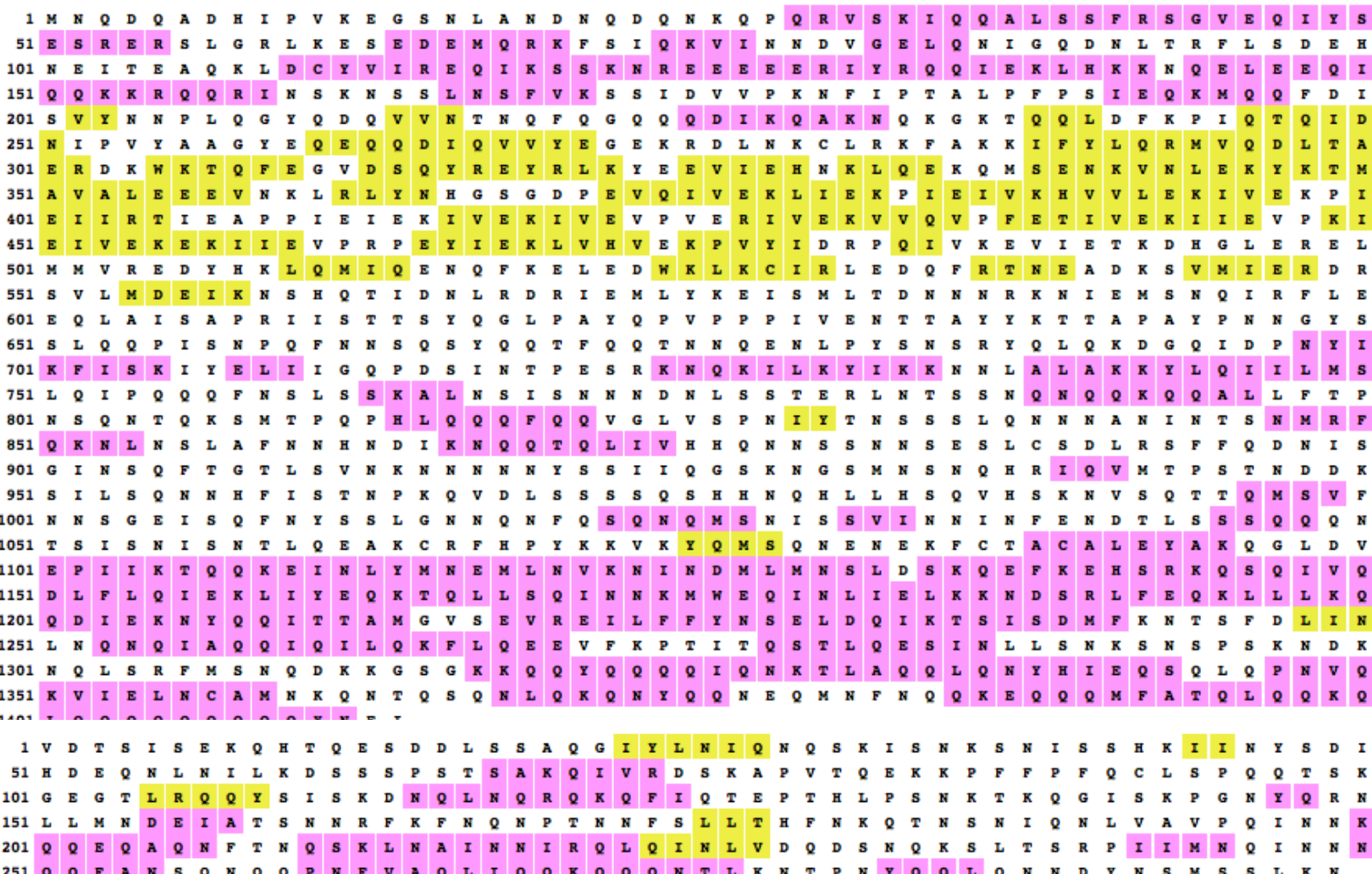
## Predicted secondary structure next slide

MNQDQADHIPVKEGSNLANDNQDQNKQPQRVSKIQQALSSFRSGVEQIYSES  
 RERSLGRLESEDEMQRKFSIQKVINNDVGELQNIGQDNLTRFLSDEHNEIT  
 EAQKLDGYVIREQIKSSKNREEEEEERIYRQQIEKLHKKNQELEEQIQKKRQ  
 QRINSKNSSLNSFVKSSIDVVPKNFIPTALPFPSIEQKMQQFDISVYNNPLQ  
 GYQDQVVNTNQFQGGQQDIKQAKNQKGTQQLDFKPIQTQIDNIPVYAAGYE  
 QEQQDIQVVYEGEKRD<sup>L</sup>NK<sup>L</sup>LRKFAKKIFYLQRMVQDLTAERDKWKTQFE<sup>G</sup>  
 DSQY

REYRLKYEEVIEHNKLQ  
 EKQMSENKVNL  
 EKYKTM<sup>A</sup>V<sup>A</sup>LEEEVN  
 KLRLY<sup>N</sup>H<sup>G</sup>S<sup>G</sup>DPEVQIV  
 ECLI  
 EKPIEIVKHVVL  
 EKIV  
 EKPIEIRTIE<sup>A</sup>PPIEI  
 EKIV  
 EKIVE<sup>VPV</sup>  
 ERIV  
 EKVVQVPFETIV  
 EKII EVPKIEIV  
 EKEKII EVPRPE<sup>Y</sup>I  
 EKLHVH  
 EKPV<sup>Y</sup>I  
 DRPQIV  
 KEVIE

TKDH<sup>G</sup>L<sup>E</sup>R<sup>E</sup>LMMVREDYHKLQMIQENQFKELEDWKLK<sup>L</sup>IRLEDQFRTNEADK  
 SVMIERDRSVLMDEIKNSHQITIDNLRDRIEMLYKEISMLTDNNNRKNIEMSNQIR  
 FLEEQLAISAPRIISTTSYQGLPAYQPVPPIVENTTAYYKTTAPAYPNNGYSSL  
 QQPISNPQFNNSQSYQQTFQQTNNQENLPYSNSRYQLQKDGQIDPNYIKFISK  
 IYELIIGQPDSINTPESRKNQKILKYIKNNLALAKKYLQIILMSLQIPQQQFNSLSS  
 KALNSISNNNDNLSSTERLNTSSNQNQKQKQALLFTPNSQNTQKSMTPQPHLQ  
 QQFQQVGLVSPNIYTNSSSLQNNANINTSNMRFQKNLNSLAFNNHNDIKNQ  
 TQLIVHHQNNSSN<sup>S</sup>ESL<sup>S</sup>SDLRSFFQDNISGINSQFTGTLSVNKNNNNNYSSII  
 QGSKNGSMNSNQHRIVMTPSTND<sup>D</sup>KSILSQNNHFISTNPKQVDLSSSSQSHH  
 NQHLLHSQVH<sup>S</sup>KNVSQTTQMSVFNNSGEISQFNYS<sup>S</sup>LGNNQNFQSQNQMSNIS  
 SVINNINFENDTLSSSQQNTSISNISNTLQEA<sup>K</sup>RFHPYKKVKYQMSQNEKEF  
<sup>C</sup>TAC<sup>A</sup>LEYAKQGLDVEPIKTQQKEINLYMNEMLNVKNINDMLMNSLDSKQEFIKE  
 HSRKQSQIVQDLFLQIEKLIYEQKTQLLSQINNKMWEQINLIELKKNDSRLFEQKLL  
 LKQQDIEKNYQQITTAMGVSEVREILFFYNSEL<sup>D</sup>QIKTSISDMFKNTSFDLINLNQN  
 QIAQQIQLQKFLQEEVFKPTITQSTLQESINLLSNKSN<sup>S</sup>PSKNDKNQLSRFMSNQ  
 DKKGSGKKQQYQQQQIQNKTLAQQQLQNYHIEQSQLQPNVQKVIELN<sup>C</sup>AMNKQNT  
 QSQNLQKQNYQQNEQMNFNQQKEQQQM<sup>F</sup>ATQLQKQLQQQQQQQQYNEIV  
 DTSISEKQHTQESDDLSSAQGIYLNINQNSKISNKSNISSHKIINYSDIHDEQNLNILK  
 DSSSPSTSAKQIVRDSKATPVTQEKKPFFPFQ<sup>L</sup>SPQQTSKGEGLTRQQYSISKDN  
 QLNQRQKQFIQTEPTHLPSNKTQKQGISKPGNYQRNLLMNDEIATSNNRFKFNQNP  
 NNFSL<sup>L</sup>THFNKQTN<sup>S</sup>NIQNLVAVPQINNKQQEQAQNF<sup>T</sup>NQSKLNAINNIRQLQINLV  
 DQDSNQKSLTSRPIIMNQINNNQQFANSQ<sup>N</sup>QQPNFVAQLIQKQQQNTLKNTPNY  
 QQLQNDYNSMSSLKN

Predicted secondary structure



## Predicted secondary structure next slide

MSIQSSNFGSLNKNTLDLQFREYGNIDDVLKYQKSRIIQSSDIEQQQGSGSN  
 NALILS **C**P ISHE **C**DSLNNRIIPNAQTILQETIQTDTANLLGSLEAGNIPQKSGL  
 QKNQSEFLEQKLNSMKQKHKQLQQQLSEVSKIDSVSNKKKNVVHSQSTQG  
 FFKGGVSTSNLDMFNF **G**SLKNQIKQMDVISNILSSIQLKPNNNQQ

KDIEVII  
 EKNDLYRILKILIDQLVSYI  
 EKLNQVVS  
 ERDIIDIQRQNA  
 EKLNNEL  
 KETIRILEEELKIYKNQSTETKIEL  
 EKQKALYDIL  
 EKKYLDLLKRKQEVKIIN  
 KETV  
 KEIVI  
 EKPIEIL  
 KEINNVRIVKQK  
 KEVRKIIRPQAEVKHTI  
 EKKLHIL  
 EREKIVPQIQVI  
 EKV  
 KEIPKIEFV  
 DKEKVI  
 EKPPFFV  
 EKLV  
 DRFVEVPRFIP  
 KEVLV  
 DRPVYV  
 DRPV  
 EKIVKVPQIV  
 EKEKLV  
 ERPVFI  
 EKPV  
 DRFI  
 EKETFI  
 DRPVYIQEPPVHI  
 EKVIEVPKII  
 EKQKII  
 DRPIYINQPRI  
 ERVI  
 EKPVFINKYEPIN

SQLVNSYSTPTISTQFNSPAANYQRFIPEQISRQNNFLSPFNSIQQTFSQDL  
 TSPSYIKQYV

Predicted secondary structure

```
1 M S I Q S S N F G S L N K N T L D L Q F R E Y G N I D D V L K Y Q K S R I I Q S S D I E Q Q Q G S G
51 S N N A L I L S C P I S H E C D S L N N R I I I P N A Q T I L Q E T I Q T D T A N L L G S L E A G N
101 I P Q K S G L Q K N Q S E F L E Q K L N S M K Q K H K Q L Q Q Q L S E V S K I D S V S N K K K N V V
151 H S Q S T Q G F F K G G V S T S N L D M F N F G S L K N Q I K Q M D V I S N I L S S I Q L K P N N N
201 Q Q K D I E V I I E K N D L Y R I L K I L I D Q L V S Y I E K L N Q V V S E R D I I D I Q R Q N A E
251 K L N N E L K E T I R I L E E E L K I Y K N Q S T E T K I E L E K Q K A L Y D I L E K K Y L D L L K
301 R K Q E V K I I N K E T V K E I V I E K P I E I L K E I N N V R I V K Q K K E V R K I I R P Q A E V
351 K H T I E K K L H I L E R E K I V P Q Q V I E K V K E I P K I E F V D K E K V I E K P F F V E K L V
401 D R F V E V P R F I P K E V L V D R P V Y V D R P V E K I V K V P Q I V E K E K L V E P V F I E K P
451 V D R F I E K E T F I D R P V Y I Q E P P V H I E K V I E V P K I I E K Q K I I D R P I Y I N Q P R
501 I E R V I E K P V F I N K Y E P I N S Q L V N S Y S T P T I S T Q F N S P A A N Y Q R F I P E Q I S
551 R Q N N F L S P F N S I Q Q T F S Q D L T S P S Y I K Q Y V
```

This may be the same gene as ALV1 or it may be an orthologue.

Predicted secondary structure next slide

MNQDQADHIPVKEGSNLANDNQDQNKQPQRVSKIQQALSSFRSGVEQIYSESR  
 ERSLGRLKESEDEMQRKFSIQKVINNDVGELQNIGQDNLTRFLSDEHNEILTEAQ  
 KLD<sup>C</sup>YVIREQIKSSKNREEEERIRQQIEKLHKKNQELEEIQQKKRQQRINSKN  
 SSLNSFVKSSIDVVPKNFIPTALPFPSIEQKMQQFDISVYNNPLQGYQDQVNTNQ  
 FQGQQQDIKQAKNQKGTQQLDFKPIQTQIDNIPVYAAGYEQQQDIQVVYE<sup>G</sup>  
 EKRD<sup>L</sup>LNK<sup>C</sup>LRKFAKKIFYLQRMVQDLTA

ERDKWKTQFE<sup>G</sup>VDSQ<sup>Y</sup>  
 RE<sup>Y</sup>RLK<sup>Y</sup>EEVIEHNKLQ  
 EKQMSENKVNL  
 EK<sup>Y</sup>KTMAVALEEEVNKLRL<sup>Y</sup>NH<sup>G</sup>SGDPEVQIV  
 EKLI  
 EKPIEIVKHVVL  
 EKIV  
 EKPIEIIRTIEAPPIEI  
 EKIV  
 EKIVE<sup>VPV</sup>  
 ERIV  
 EKVVQVPFETIV  
 EKIEVVKIEIV  
 EKEKIEVPRPE<sup>Y</sup>I  
 EKLHVH  
 EKP<sup>Y</sup>VI  
 DRPQIV  
 KEVIE

TKDH<sup>G</sup>LERELMMVREDYHKLQMIQENQFKELEDWKLK<sup>C</sup>IRLEDQFRTNEADKSVMII  
 ERDRSVLMDEIKNSHQTIDNLRDRIEMLYKEISMLTDNNNRKNIEMSNQIRFLEEQLA  
 ISAPRIISTTSYQGLPAYQPVPPIVENTTAYYKTTAPAYPNNGYSSLQQPISNPQFNN  
 SQSYQQTFQQTNQENLPYSNSRYQL

## Predicted secondary structure

1 M N Q D Q A D H I P V K E G S N L A N D N Q D Q N K Q P Q R V S K I Q Q A L S S F R S G V E Q I Y S  
 51 E S R E R S L G R L K E S E D E M R K F S I Q K V I N N D V G E L Q N I G Q D N L T R F L S D E H N  
 101 E I L T E A Q K L D C Y V I R E Q I K S S K N R E E E E R I Y R Q Q I E K L H K K N Q E L E E Q I  
 151 Q Q K K R Q Q R I N S K N S S L N S F V K S S I D V V P K N F I P T A L P F P S I E Q K M Q Q F D I  
 201 S V Y N N P L Q G Y Q D Q V V N T N Q F Q G Q Q Q D I K Q A K N Q K G K T Q Q L D F K P I Q T Q I D  
 251 N I P V Y A A G Y E Q E Q Q D I Q V V Y E G E K R D L N K C L R K F A K K I F Y L Q R M V Q D L T A  
 301 E R D K W K T Q F E G V D S Q Y R E Y R L K Y E E V I E H N K L Q E K Q M S E N K V N L E K Y K T M  
 351 A V A L E E E V N K L R L Y N H G S G D P E V Q I V E K L I E K P I E I V K H V V L E K I V E K P I  
 401 E I I R T I E A P P I E I E K I V E K I V E V P V E R I V E K V V Q V P F E T I V E K I I E V P K I  
 451 E I V E K E K I I E V P R P E Y I E K L V H V E K P V Y I D R P Q I V K E V I E T K D H G L E R E L  
 501 M M V R E D Y H K L Q M I Q E N Q F K E L E D W K L K C I R L E D Q F R T N E A D K S V M I E R D R  
 551 S V L M D E I K N S H Q T I D N L R D R I E M L Y K E I S M L T D N N N R K N I E M S N Q I R F L E  
 601 E Q L A I S A P R I I S T T S Y Q G L P A Y Q P V P P P I V E N T T A Y Y K T T A P A Y P N N G Y S  
 651 S L Q Q P I S N P Q F N N S Q S Y Q Q T F Q Q T N N Q E N L P Y S N S R Y Q L



*Tetrahymena* XP\_001012752.2

This may be the same gene as ALV2 or it may be an orthologue.

Predicted secondary structure next slide

MSIQSSNFGSLNKNTLDLQFREYGNIDDVLKYQKSRIIQSSDIEQQQGSGSNNAL  
 ILS **C**PISHE **C**DSLNNRIIPNAQTILQETIQTDANLLGSLEAGNIPQKSGLQKNQS  
 EFLEQKLNSMKQKHKQLQQQLSEVSKIDSVSNKKKNVVHSQSTQGFFKGGVST  
 SNLDMFNF **G**SLKNQIKQMDVISNILSSIQLKPNNNQQ

KDIEVII  
 EKNDL **Y**RILKILIDQLVS **Y**I  
 EKLNQVVS  
ERDIIDIQRQNA  
 EKLNNEL  
 KETIRILEEELKI **Y**KNQSTETKIEL  
 EKQK **ALY**DIL  
 EKK **Y**LDLLKRKQEVKIIN  
 KETV  
 KEIVI  
 EKPIEIL  
 KEINNVRIVKQK  
 KEVRKIIRPQ **A**EVKHTI  
 EKKLHIL  
EREKIVPQIQVI  
 EKV  
 KEIPKIEFV  
DKEKVI  
 EKPPFFV  
 EKLV  
 DRFVEVPRFIP  
 KEVLV  
 DRPV **Y**V  
 DRPVNQFI

NNNRIQQKIKQKRLKKQLK **C**LKQQKKKNQSRDQYLLKNQLTDLQKRKLLLIDQSIFK  
 NLQYIQKKLQKFLRLQKSKKQLIDLFTSINHVQKELLKNLYSQINMSQQILSQLIHILLQ  
 LFQHNLTHLLLIIRDLYLSKLVDKITF **C**HLLILYSKHFLKIQLHLHTLNSMFEKYL FVIVK  
 QYSLIKISLFLKFQISLYLIISQSFTFQKINLF **C**KIKNYQK **C**IFVFQQTILQPLINLLLSDKF  
 IQQKQFMILFIVSKQKIDILIFYNQYVNFLF

*Tetrahymena* XP\_001012752.2

## Predicted secondary structure

1 M S I Q S S N F G S L N K N T L D L Q F R E Y G N I D D V L K Y Q K S R I I Q S S D I E Q Q Q G S G

51 S N N A L I L S C P I S H E C D S L N N R I I P N A Q T I L Q E T I Q T D T A N L L G S L E A G N I

101 P Q K S G L Q K N Q S E F L E Q K L N S M K Q K H K Q L Q Q Q L S E V S K I D S V S K K K N V V H S

151 Q S T Q G F F K G G V S T S N L D M F N F G S L K N Q I K Q M D V I S N I L S S I Q L K P N N N Q Q

201 K D I E V I I E K N D L Y R I K I L I D Q L V S Y I E K L N Q V V S E R D I I D I Q R Q N A E K L N

251 N E L K E T I R I L E E E L K I Y K N Q S T E T K I E L E K Q K A L Y D I L E K K Y L D L L K R K Q

301 E V K I I N K E T V K E I V I E K P I E I L K E I N N V R I V K Q K K E V R K I I R P Q A E V K H T

351 I E K K L H I L E R E K I V P Q I Q V I E K V K E I P K I E F V D K E K V I E K P F F V E K L V D R

401 F V E V P R F I P K E V L V D R P V Y V D R P V N Q F I N N N R I Q Q K I K Q K R L K K Q L K C L K

451 Q Q K K K N Q S R D Q Y L L K N Q L T D L Q K R K L L L I D Q S I F K N L Q Y I Q K K L Q K F L R L

501 Q K S K K Q L I D L F T S I N V Q K E L L K N L Y S Q I N M S Q Q I L S Q L I H I L L Q L F Q H N L

551 T H L L L I I R D L Y L S K L V D K I T F C H L L I L Y S K H F L K I Q L H L H T L N S M F E K Y L

601 F V I V K Q Y S L I K I S L F L K F Q I S L Y L I I S Q S F T F Q K I N L F C K I K N Y Q K C I F V

651 F Q Q T I L Q P L I N L L L S D K F I Q Q K Q F M I L F I V S K Q K I D I L I F Y N Q Y V N F L F

*Tetrahymena* >XP\_001017441.1

MSSNENQSEQEKAQVLQGTPVYQYPAYTSQYVQQQPLIQSQLVTQPVQYV  
QQPVQYVQQTVQGVHQSALSQSQAVASGLVATQIQ

REVVKGESRIEYIPY  
EKSVIEYEAVQRVDYVPK  
EKKITDYYAVEYQTEYLPQVYQ  
DRYIEYIPT  
ERIQ  
ERVEYQAIQKQIVHQPV

QQQQQVLQTVVQPATTYLQSSQVTPVVPVQQQFVPTASVVHTTIPATQVVQGA  
THVVQGASTVVQNVPTAQFVQTVPTAFSTQTLGVPVVRAGLTTSGVRTLGPV  
PYSVVEEAIKKSQKKDE

*Tetrahymena* EPI21 XP\_012653403.1, THERM\_000481318

MSEYSFFAQLYAQMNEQNLQNSWIDYGLAEENKIYFCQLKGFKKEEISEYFI  
 QQILINKQKIFRKEKENFKEETNNLQLLANITKMSDQPEQAQQQANLQQQQQ  
 PPVQTQVLQYPFQYVQQPLVQSQVVAQPIYTNGQVPTTQFLGSTLIQGATQ  
 QVSAGTVLHQSGVQVS~~GV~~VPVQTTLQ

REVKGESRIEYIPY  
 EKTVEYEA~~V~~QRIEYVPK  
 EKKVTDY~~Y~~AVEYQTEYIPQVYQ  
 DRYIEYVPT  
 ERVQ  
 ERVEYQAI  
 EKQVVHQP

AQEVSQVSQ~~VP~~VSSVSQ~~VP~~VAAQLTSSQVHVLPTTQTVISGGQQVAYQNQS  
 FAYAPVQQSVVYTTPAQLGQSVYRGPTVLTNSIAQKPVQQNAKKEEKTFLER  
 IFE

1 M S E T S F F A Q L T A Q N E E Q N L Q N S W I D Y G L A E E N K I Y F C Q L K G F K K E E I S E Y F I  
 51 F I Q Q I L I E K Q K I F R K E K E N F K E E T N N L Q L L A N I T K M S D Q P E Q A Q Q Q A N L Q Q Q Q Q  
 101 Q Q Q Q P P V Q T Q V L Q T F F Q T V Q Q P L V Q S Q V V A Q P I Y T N G Q V P T T Q F L G S T L I Q G A T Q  
 151 Q G A T Q Q V S A G T V L E Q S G V Q V S G V P V Q T T L Q R E V V K G E S R I E Y I P Y E K T V I  
 201 E T E A V Q R I E T V P K E K E V T D T A V E T Q T E T I P Q V T Q D R T I E T V P T E R V Q E R I  
 251 V E T Q A I E K Q V V K Q P A Q E V S Q T V S Q V P V S S V S Q V P V A Q Q L T S S Q V H V L P P T  
 301 T Q T V I S G Q Q Q V A T Q R Q S F A T A P V Q Q S V V Y T T P A Q L G Q S V Y R G P T V L T N S I A Q K P V Q Q N A K K E E K T F L E R I F E  
 351 A Q K P V Q Q N A K K E E K T F L E R I F E