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 β-strand; pink, amino acids predicted in α-helix; no highlight, amino acids predicted in disordered (random coil) domain.

2

Euglena 80K articulin Gene.55749_G

Articulin
Predicted secondary structure next slide

MSYLAGAPVTAAASYPVVPGAAAYPVGASYPAVAQASVLRANAQFAGQRVIR

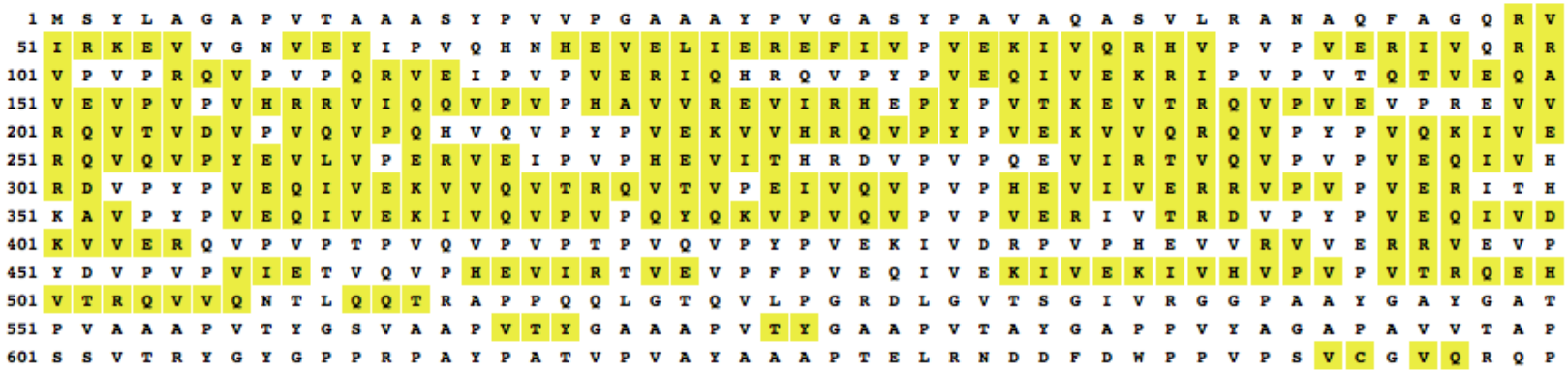
KEVVGNVEYIPVQHNHEVELI
EREFIVPV
EKIVQRHVPVPV
ERIVQRRVPVPRQVPVPQRVEIPVPV
ERIQHRQVPYPVEQIV
EKRIPVPVTQTVEQAVEVPVPVHRRVIQQVPVPHAVV
REVIRHEPYPVT
KEVTRQVPVEVP
REVVRQVTVDVPVQVPQHVQVPYPV
EKVVHRQVPYPV
EKVVQRQVPYPVQKIV
ERQVQVPYEVLVP
ERVEIPVPHEVITH
RDVPVPQEVIRTVQVPVPVEQIVH
RDVPYPVEQIV
EKVVQVTRQVTVPEIVQVPVPHEVIV
ERRVPVPV
ERITHKAVPYPVEQIV
EKIVQVPVPQYQKVPVQVPVPV
ERIVT
RDVPYPVEQIV
DKVV
ERQVPVPTPVQVPVPTPVQVPYPV
EKIV
DRPVPHEVVRVV
ERRVEVPYDVPVPVIETVQVPHEVIRTVEVPFPVEQIV
EKIV
EKIVHVPVPV

TRQEHVTRQVVQNTLQQTRAPPQQLGTQVLPGRDLGVTSGIVRGGPAAYGAYGA
TPVAAAPVTYGSVAAPVTYGAAAPVTYGAAPVTAYGAPPVYAGAPAVVTAPSSVT
RYGYGPPRPAYPATVPVAYAAAPTELRNDDFDWPPVPSVCGVQRQPQ

3

Euglena 80K articulin AAB23240.1

Predicted secondary structure



4

Euglena 86K articulin Gene.26294_G

Articulin
Predicted secondary structure next slide

MSWVAAQTQQGAFALDAADGRIDGKYFGSNVGVSVPGAPVTYGAAAPVTSYA
APAAVTSFAATAAPVTSFGAYGAYAPACPPCATGPRVVNDPLETRFVDVVKQV
ETIRSVDVPVPHEVVRTVDVPEHYDVPVPHAVHVQVPYPV

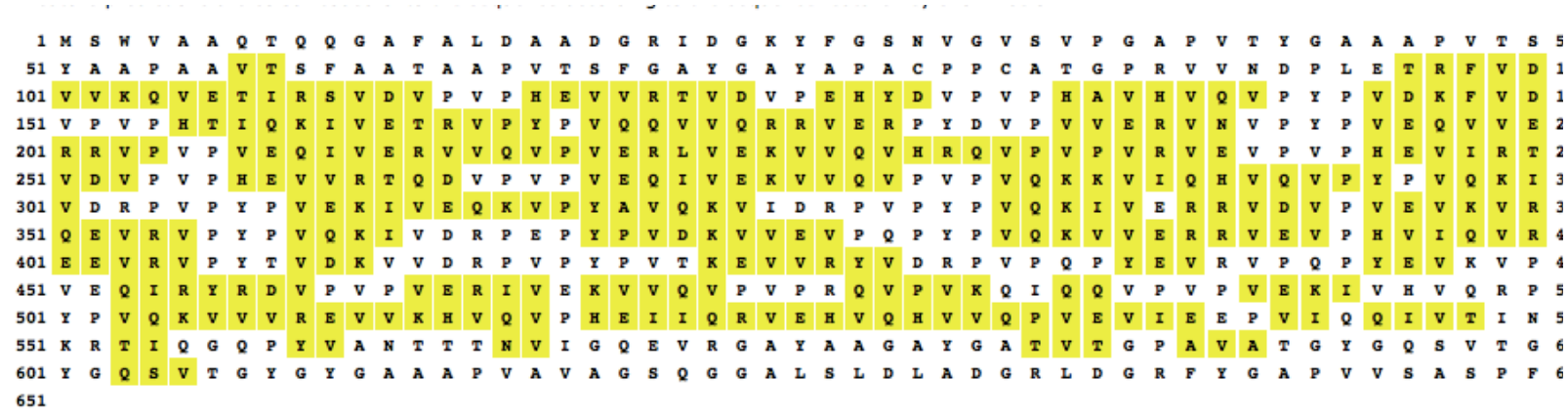
DKFVDVPVPHTIQKIVETRVPYPVQQVVQRRV
ERPYDVPVV
ERVNVPYPVEQVV
ERRVPVPVEQIV
ERVVQVPV
ERLV
EKVVQVHRQVPVPVRVEVPVPHEVIRTVDVPVPHEVVRTQDVPVPVEQIV
EKVVQVPVPVQKKVIQHVQVPYPVQKIV
DRPVPYPV
EKIVEQKVPYAVQKVI
DRPVPYPVQKIV
ERRVDVPVEVKVRQEVRVPYPVQKIV
DRPEPYPV
DKVVEVPQPYPVQKVV
ERRVEVPHVIQVREEVRVPYTV
DKVV
DRPVPYPVT
KEVVRYV
DRPVPQPYEVRVPQPYEVKVPVEQIRY
RDVPVPV
ERIV
EKVVQVPVPRQVPVKQIQQVPVPV
EKIVHVQRPYPVQKVVV
REVVKHVQVPHEIIQRVEHVQHVVQPVEVIEEPVIQQIV

TINKRTIQGQPYVANTTTNVIGQEVRGAYAAGAYGATVTGPAVATGYGQSVTGYGQ
SVTGYGYGAAAPVAVAGSQGGALSLDLADGRLDGRFYGAPVVSASPF

5

Euglena 86K articulin AAB23241.1.

Predicted secondary structure



6

Euglena Gene.3890_GEFR01000334.1
(First third of sequence)

Articulin. Predicted secondary structure at end of sequence

…FFFELFISTMSKPVTTTSGRVRYDGGSQMRVLGGATSGSASSTPIFVSETSP
ITTTHYGGSVTRTYGASTVGVGGISSHYVGGTTVAGSSLAGSTRTHAVGGTAIA
GSSALAGSSLTGSAAFGTGGSVLVGGSALVGGASSVVGGASVVGGSRVRHIG
GEGPIYVDSGYTTSGYSNVCCAPAGVRSTFVGSRREVLEAATVTQHI

EKVPVEIDVDIDTY
REIVEVPI
EKIVEVPYPV
EKIVQVPV
DRVVEVPV
ERLV
ERIVRVPEVRTVEVPVEHCI
EKLVTVPEIL
EREVAI
EKIV
ERVVRVPEIHEIEVPV
EKLI
ERILRVPEIHEIEVPV
ERIV
EKVIRVPEIRQVEVPV
ERII
EKIIRVPEIHEIEVPV
EKRIETVVRVPEVHQIEVPV
EKII
ERVVRVPEVHQIEVPV
EKIV
ERVVQTHT
RDVIEVPV
EKVV
ERIVRVPQETLIEVPV
EKIV
ERVVRVPQHTTVEVPV
EKVI
ERVVQTHT
RDIIEVPV
EKVV
ERVVRVPQETIIEVPV
EKVV
ERIVTVPQHTTVEVPV
EKIV
ERVVQTHT
RDVIEVPV
EKVV
ERIVKVPQETIIEVPV
EKLI
ERVCRVPQETIVEVPV
EKII
ERVSQTHV
RDIVEVPV
EKLV
ERIVTVP
KETVIEVPV
EKII
ERIVKVPQETIVEVPV
EKLV
ERVCQTHT

7

Euglena Gene.3890_GEFR01000334.1
(Second third of sequence)

RDIIEVPV
EKIV
ERVVKVPQETIIEVPV
EKVI
ERVVTVPQETI
REVPV
EKVV
ERVVQTHT
RDIIEVPV
EKII
ERVVKVPQETIVEVPV
ERVI
EKVCSTHT
RDIIEVPI
EKII
ERIVKVPQETIIEVPV
EKCI
ERVVTVHSTDVVEVPV
EKVI
ERIVRVPQETIIEVPV
EKII
ERVCEVQV
EKIVEVPVQQIVRVPVEVPVEHIV
ERVVKVPVETI
KEVPV
EKIVEIIRPVEVIHHIEVPV
ERIV
ERVVEVL
KEVPV
DRVV
ERLVEVPVTHTVEVPVVSVV
EKLVEVP
KEII
REVPV
EKIV
ERVVTVNVDNYIEVPV
EKII
ERVVTVPV
EKRIEVPVEVVV
ERI
KEVQV
EKIVEVPV
EKII
EKIVKVPV
EKIVEVPVECVV
ERIVQVPVTQTVEVPV
ERIV
EKIVEVQVQQIVEVPV
EKVV
ERVVKVPVDHVIEVPV
ERVV
ERI
REVPV
EKIIQVPV
EKIKIVTV
DRIVEVPVEV

8

Euglena Gene.3890_GEFR01000334.1
(Final third of sequence)

EKIVTVQV
EKIVQVPV
EKIV
EKLVEV
EKIV
ERTVTV
ERVVEVPKVV
EKIV
EKIVEVP
REYI
REVTVTSI
KEVPV
EKIV
ERIVEVPITNVVEVPV
ERVV
ERVVEIPL
ERVV
ERIVEVPV
EKII
ERIIEVPV
ERVVEHIIEVPV
ERLV
ERIVEIPV
EKII
EKIIEVPI
EKII
ERVIEVEVPIEI
EKVVEI
EKIVEVPVVV
EKVVEV
EKIVEVSGGVEIV
EKIVEV
EKIVEVPVVV
EKVIET
EKII
EKGVV

TEIVPEVSTELYSTSYSTGSKGTSSSISASGGASYSVGGGYGVSGGASTGGGYAVS
GGTGGAGYGSIGTSGGTGYSSITTGGSGTGYSSIGGTGAGVGYSSATTGGAGAGA
GTGYSSISGSGAPGGAGGYSSTSYSSGTYGAGGGTYGAGAPGAGGYGSYSSTSY
SGTGGSGSAPGRTGYTATSGSGTRSGSATRSGSGSRA

9

Euglena Gene. 3890_GEFR01000334.1

Predicted secondary structure

N.B Sequence submitted to PSPRED in 5 sectors. Longer sectors yield predictions of no secondary structure using both PSIPRED and YASPIN.

10

Euglena 15434_G
(First half of sequence)

Articulin

MTVEESRTFSQTPITTYQSGARAVSNTRILSTGATTLSSATSIVDGSRRYAR
SSAAVCCQPFDPCNSVVTRQLQTRYLDPVETTHVEKVPMEIEVDVEGYRE
VIEVPV

ERIVEVPV
ERIRTVPQMQVVEVPVNRHV
ERLVVVPEL
KERQVFIPKTI
ERVVKVPEVYEYEVPVQRTV
EKVVVVP
REVLVEVPV
DKVTTV
DRVV
EKHVPVEHIV
ERLVEQ
ERVVETQVKVQQIVEVP
REVIKNYEV
ERIV
ERVVEVPQL
KEVQVPV
ERVY
ERIVPVPK
RDTVQIPV
ERIH
ERIVPVPQTNIVEVPV
DKLI
ERVCTVPKQCTVEVQV
EKVV
DRIVEVPRTITQEVLIDMVV
EKVVTVPS
EKIIEVPV
EKVVESIVQIPKVVTHDVPV
EKVI
ERVVTVPHEQRIEIPIQRVV
ERVVQVPV
DRVV
ERVVEVPHVVT
REVPVESVV
ERVVTVPVNRVVEVPV
DRYV
DRVVQVPV
EKVVNRVVEIPKVIT
REVELQQIV
EKIVEIPTTKVVEVPVIRYI
EKIVEVPVT
REVEVPVQKYV
EKIVEIPY
ERVVEVQVEVPV
ERIV
EKVVEVPFTV
DRVVEISI
EKVI
EKIIEV
ERRVEVPVEVIV
EKVV

11

Euglena 15434_G
(Second half of sequence)

EKVVEIEVPVII
EKVVEI
EKIVEV
ERRVEVPFDTIVEV
ERVVEV
EKIVEIPIEVVKNV
ERVV
EKTIE

TSYDVTPVVEQKVSSFSASGGGSYSTSASYSTGGGYASGGGGYVSGGGGGFT
SGGGGYVSGGGYSSGGGTGGSAFIGGGGSGGSAFISGGGSGGSAFISGGGSG
GSAFISGGGGSGSGSRVTGGFSSSGALPGAAGVTGTTQLYSSQTSSLSGSTGA
WGLPKQY

12

Euglena 22739_G

Articulin

…AASPVVDDLPGRPGGGSPELVAMAPALGSGVAGGQAVGDQGEVVDLMQIGSE
QEIVEALEVESFDSRAVHQVALLEEDSPLWRSPVKPNGVV

ERIVPV
ERLIEVPKVIHCVE
ERIVEVPIQRIV
ERLVPMEVTHI
KEVPV
ERIV
EKLVEVPVVTI
KEVPVETVV
ERVEVVTV
EKVV
EKLVTV
EKIVEVPTFVEVIHVKVIEVPV
ERVI
EKIV
ERPVPVESVV
ERLVEVPVQHVVQVPV
EKIV
ERIVEV
EKPVTVTH
EKIVEVPV
DRIV
EKVVTQEVPRIVEVPV
ERLV
ERLVEVPVVTI
KEVPI
ERVV
ERVEVITV
EKVV
EKVVPV
EKIVEV
EKVVRVVE
DRIVEVPVARVIQQVV
ERPVVVETRVETI
REVPV
REIVQIPVYEVI
ERI
KEVQVPYT
REVPVEHVV
EKIVEFVEVPIYV
DRIVEVPP

SVRSTSRAQQKISAHQRQGCGASRRNGSGRW

13

Euglena 54789_G

Articulin

…YVRHYTDLQPTHREYVEMVPVEVSVQEEIIQDIIEVPTFERRRRMVWEEFGG
AAPTVVPFRPVTPERPVPVPLQRPADTFAPVPVPHEGAQTVAPSERHGPVG
PAEHTLQVVRVPQVQRVEVPV

ERVV
EKVVTVPQVQRVEVPV
ERLV
EKVVKVPEVRIVEVPV
DRVI
ERRVEVPV
DRVV
ERVVKVPEVTTVEVPVETI
KEKIVTVPVQRTIEVPI
EKFV
ERIVKIPEVRTVEIPV
ERII
EKIVQVPI
EKIVEVPV
DRII
ERVIEVPVTKVV
EKVIEVPV
DRVVEVPV
EKIV
EKFVEVEVLV
ERIV
ERLVEVEVP
KEKII
ERIIEVPV
ERVI
EKVVEVYVPV
ERVVEV
ERIVEVPV
ERIV
ERIVEV

PYSEDLAGAQVPSEQMAEKGSEEATRSAQAQSGSHPVSHVSSEVSSSALPPVGSI
SSWMPGSSYASTQPSQSVASSAQSAGAVDVSTTWGPSGYRPAAARSGPSLAGY
VPSAGARSEPPGTAAGVVDFRQYSAQLAPEAEQQRF

14

Euglena 61926_G

Articulin

…FFSPFPLSFLSFTLQSADSQRYMSYSGAYRADYPVRSTSPRLSRPYPFSRPA
ERVVSPTRVVDGRYRSPSRLVSESCFPAQSTVVETTSACIRHTQEIVEVPVAQ
PVEVVRTVPMEVPTLEVVERARVQVQEHVIEVP

REHIV
EKIV
EKVVLKTIEVPVEHIV
EKIIEVPV
ERLV
ERIV
ERIVEVPQ
DRLV
ERVI
EKIVEVPVEYPV
ERVV
EKIVEVLVEIPV
ERIV
ERIVEVPVEVPV
ERIV
EKIVHVTTEVPRVQRQ
ERIIEVPVDHIV
ERIVEVPV
DRPISAIQ
ERLVPHTV
ERIVKYEVPV
DRVIEVPV
EREVI
REVHV
DRPVIHQV
ERPVVHTVTRPVIHEV
ERPVIHEV
ERPVIHEV
ERPVVHRV
ERPV

AVPVSVVRQEPAVCYETAPYRQRYVNGTTTYSTRAPR

15

Eutreptiella CAMPEP_0113738536

Articulin

MPIPKTARDVAAALDNADGVADGKFHGLPIVVGDSRPSSRNVYSSPRSTRYV
SDEFIRTGLPATRVGYPATRVGYPATRSRYMSGPVIHHHEPPTIVRHSTAPSA
SRIIRSEGPLVRRERIDQATSRTHIEKRVVPVTRVIEESVEPVRRFVQPVERRV
EVQPVQTVERIVEVPVERRVVDVQQVEHVVEVEVPVPVEVTKEVKVEVPVEK
PVYVEHVVEKIVQVDVNHETQVPIEKIVETLRVEEVVHIQEVPVQQTVEKLVTV
EVPHIIEVPV

EKVV
EKIVEV
EKII
ERIVPV
EKIV
EKEVILKIVE
EKIIEVPV
ERVKHV
EKIVEVPKIIETIV
EKIVEVPVKHVVEVPV
EKIV
EKLVQVPVMYT
KEVPIETVV
EKEIPGPV
REV
EKIVQVPV
EKLVEV
EKIIEVPVQVPV
EKIVEVPVEVPV
ERLV
EKLVEVPV
EKIVEVMV
EKVIEVPVEVPV
ERII
EKVVEVPVHV
EKIVEIPLEI
EKIV
EKIIEVEIPVEIPVEI
ERIV
EKLVEVPVVV
EKVVEVEIPIEI
EKIIEI
EKIVEIPIEI
EKIVEI
EKIIEVEVPIEI
EKVIEV
EKIVEVPIEI
EKVI
EKTIEV

GTEPTTVTKYSSFSTSSAPIETKRFSSQGMSRSSTQSSSTTPGGEIQYSSSAGSAESTGT
GTGWGTPGTNRSAWLGNMNN

16

Eutreptiella CAMPEP_0113740288

Articulin

MTVTITPTMTIQQTPGRFTX

EREVGYRKLH
RDVLQEIHEVLV
EKPVEVIQTEVMEVPVKHIE
ERVVAETH
ERIVEVPV
EKIV
ERVV
EKVVVKTVEVPV
ERIV
EKIVEVPV
ERIVEVYV
EKIVEVP
RDKIV
ERIV
EKFVEVPQIV
ERIV
EKIIEVPVDYPV
ERVV
EKLVQVPVEVPV
ERIV
ERIVEVPVEVPV
DRIV
EKIVEVEVEVPVIQ
DRERVVEVPTDNI
REKVVEVPV
ERLVEVPS
EREVVKQVQVPVEIPVEVEPEVDVVIV
DKAV
EKT
ERPKRITLADQNYVPPRYSL
DKTY
KERSS

ASYQYTAPEKSVEYVTHAPSEKRYTRSSSGQVVHEYAAPAKSVEYVTS
GPERGQVLREYTREYYEDPAHEGDYRISSGRHGTAYEAARALDAADGRLDG
KYQGKDIYVNGEKVRSGGXACCLRRPQPRLHKSQPAGEQQVAVAITRVQAXV
TMCERLSILPRSPSLPPQQGHLVLAGDLHLPLYTQKKQVRDRVNKATGTGWV
MAIGVNQRQVKGQPSSAIRVLRMSSKKHSMGSAEPTGSCRVEGENKHSEVR
GRALESDNSPKHHHLPALTSNQAIFFRNEWLQSSWWANVVSGESLC

17

Articulin

Eutreptiella CAMPEP_0113741250 (First half of sequence)

…GAPHSAPISSNCTIDNMSSRVTTGSASGGRVRYDGGSQMRVRDPKISVTE
TGTPGNEVIVVSAERDVVPTGTVTSSTYQGGSTIVRERGYPVTSVGHSYFRS
GSPVAVGGSSVARSAVRVGQVQAGGAYYRQGGRALYSQSVRRELLDEATE
TTMI

EKVPIEIDVEVDNY
REVVEIPV
EKIVEVPYPV
ERVVEVPVEHTV
EKIITVPEL
REIEIPV
ERIV
EKVVRVPQETLVEVPV
EKVI
ERVVRVPEVTEIEVPV
ERIV
EKVVRVPEVHEIEVPV
EKII
ERVVNTHS
RDIIEVPV
ERVI
ERVVEVEV
EKIVHVPV
EKVVTVPHEVVVNSVV
EKIVTVEVETI
KEVPVESVVEVIRPVEVTRHVEVPVEHIV
ERIVEVV
KEVPV
ERV
REKIVQVPVTNTVEVPVVSVV
EKLVEVP
REIV
REVPV
EKIV
ERIVQVPVDNYIEVPV
EKSV
EKLVQIPV
EKTVEVPI
ERVV
EKLVEVPVETIVEVPV
EKL
RESIKVVTV
EKLIEVPV
ERVV
EKIIQVPVEHSVQVPI
ERIV
EKIVEVQVQQIIEVPV
EKEV
ERVVKVPV
EKVVEVPI
EKIV
ERIVQVPV
EKIIEVPI
EKVVEVIKVVQN

18

Eutreptiella CAMPEP_0113741250
(Second half of sequence)

ERIVEVPVEV
ERLVSVPV
EKIIEVPL
ERIV
EKAVPV
DRII
EKVVEV
ERIVEVPMIV
EKFV
DRTVTV
EKIV
EKIVEVPVTNVVEVPI
ERTV
EKVVEVPV
EKIV
ERIVEVPV
EKIVEV
EKIVENIIEIPI
ERVI
EKIIEVEIPIEI
EKIVEI
EKVVEVEIPIEI
EKIVEI
EKIVEVEIPIEI
EKIVEI
EKIVEIEVPVIV
EKVVEI
EKIVEVPIEI
EKLV
ERTIVE

TAPTVETTVTTETVEVPRGAGYSSYSSSSSFSSGSRGAGGGGGGAAGGGSGGSTSYSTWSSQSGX

19

Eutreptiella CAMPEP_0113737346 (First half of sequence)

Articulin. Head domain a perfect repeat, followed by a bit of a third repeat, followed by a novel sequence (blue font)

…NDGLFAEPDHTLSQLALPGPPGGEAMAVATQRLAGKLKATVSPKHLVFS
KLIPALKRACKRSDLLEEMEADGLTTIPPPSDWVLAKMAREWRCWLPSGP
GYFDYASQPVPWELVGGRHWPNSRGDFDERDTGHEGFSTQDFLNRPEAK
AAGLTAAEVIALRLYTGPGYIPINRSLRVNSGRFAVTQWALDCAIGKLALAERE
GLLLRGLRLLPKEEWQQQYEDCRCADDAMDLWISDPAYSSTTTDMAVATGT
DFGGPCTFVFHAQCDLSPADGLIGNAASVQWVSQYPDEVERLLPSNSCFLSL
PQGMRSELPEGMGDRKIFQFFSRFLWDYERCCPPVVTEVEEYVVRINEVMW
HVYRGLGSEKPPPEVTADAELTAFFLAPDPVGVDGMDGGLPEVLEEEM

NDGLFAEPDHTLSQLALPGPPGGEAMAVATQRLAGKLKATVSPKHLVFSKLI
PALKRACKRSDLLEEMEADGLTTIPPPSDWVLAKMAREWRCWLPSGPGYF
DYASQPVPWELVGGRHWPNSRGDFDERDTGHEGFSTQDFLNRPEAKAAG
LTAAEVIALRLYTGPGYIPINRSLRVNSGRFAVTQWALDCAIGKLALAEREGLL
LRGLRLLPKEEWQQQYEDCRCADDAMDLWISDPAYSSTTTDMAVATGTDF
GGPCTFVFHAQCDLSPADGLIGNAASVQWVSQYPDEVERLLPSNSCFLSLP
QGMRSELPEGMGDRKIFQFFSRFLWDYERCCPPVVTEVEEYVVRINEVMW
HVYRGLGSEKPPPEVTADAELTAFFLAPDPVGVDGMDGGLPEVLEEEM

NDGLFAEPDHTLSQLALPGPPGGEAMA

ASSLNGLPTFCAVSQEDPVAVETRIEVPTVEKVTAVQAEDIQEVPMVEVPV

ERVIEVPV
EKII
ERI
ERIPVETIREVPVEV
EKIVHVPVEHVV
EKVVEVPV
ERVV
EKVLQVPVEV
EKVV
EKVVEVPV
EKIV
EKVFQVPVEV
EKIIHVPVEHVV
EKVVEVPV
ERVV
EKVLQVPVEV
EKIIHVPVENVV
EKVVEVPV
ERVV
EKVLQVPVEV

20

Eutreptiella CAMPEP_0113737346
(Second half of sequence)

EKVV
EKVVEVPV
EKVV
EKVEKVPVEV
EKIIHVPVEHVV
EKVVEVPV
ERVV
EKVLQVPVEV
EKIIHVPVXXVV
EKVVEVPV
ERVV
EKVLQVPVEV
EKVV
EKVVEVPV
EKVV
EKV
EKVPVEV
EKIIHVPVEHVV
EKVVEVPV
ERV

21

Eutreptiella CAMPEP_0113744366

Articulin. Predicted secondary structure next slide

MAEARTGPSAMDALEDADRRTATRQTRSAAEALDAVDGVMDGKFFGRPIVAT
GPSRVVGGRSVVDSIDAFDGRRYRSSGGVHGSAAEALDAADGVMDGRFYGR
PIVETRSPQRLRGVTAQALDAADGVIDGRFYGRPIVETRGATTRYDDHVEVRR
VHGGTRHRSKVVEVPVYHHENRIPVEVQVPVQIPVEVPVRI

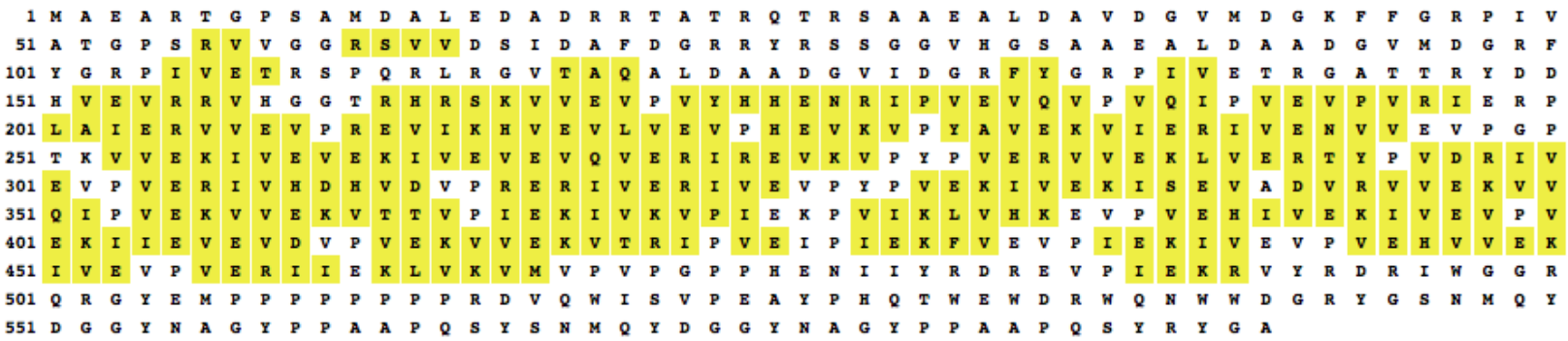
ERPLAI
ERVVEVP
REVIKHVEVLVEVPHEVKVPYAV
EKVI
ERIVENVVEVPGPTKVV
EKIVEV
EKIVEVEVQV
ERI
REVKVPYPV
ERVV
EKLV
ERTYPVDRIVEVPV
ERIVHDHVDVP
RERIV
ERIVEVPYPV
EKIV
EKISEVADVRVV
EKVVQIPV
EKVV
EKVTTVPI
EKIVKVPI
EKPVIKLVH
KEVPVEHIV
EKIVEVPV
EKIIEVEVDVPV
EKVV
EKVTRIPVEIPI
EKFVEVPI
EKIVEVPVEHVV
EKIVEVPV
ERII
EKLVKVMVPVPGPPHENIIY
RDREVPI
EKRVY
RDRIW

GGRQRGYEMPPPPPPPPRDVQWISVPEAYPHQTWEWDRWQNWWDGRYGSNMQY
DGGYNAGYPPAAPQSYSNMQYDGGYNAGYPPAAPQSYRYGA

22

Eutreptiella CAMPEP_0113744366

Predicted secondary structure



23

Eutreptiella CAMPEP_0113747498
(first half of sequence)

Articulin. Predicted secondary structure next slide

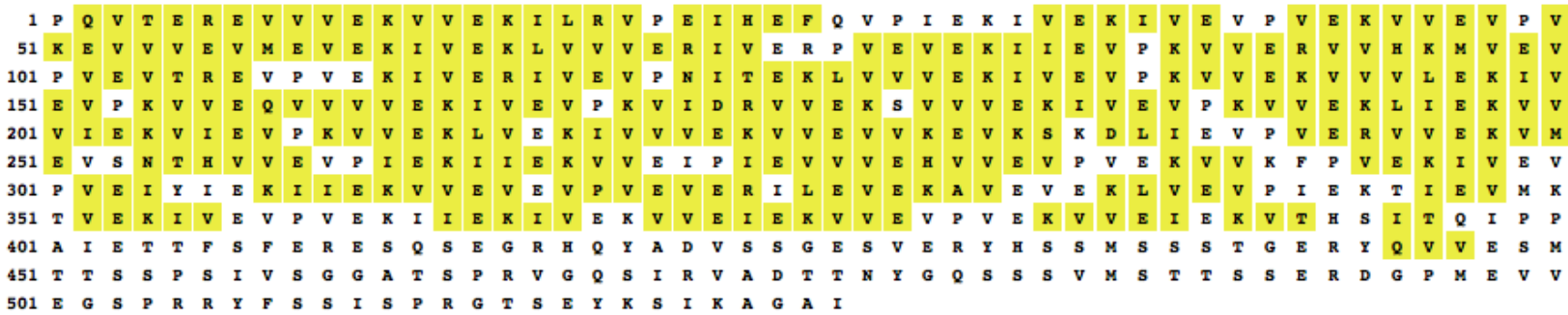
…PQVTE

REVVV
EKVV
EKILRVPEIHEFQVPI
EKIV
EKIVEVPV
EKVVEVPV
KEVVVEVMEV
EKIV
EKLVVV
ERIV
ERPVEV
EKIIEVPKVV
ERVVHKMVEVPVEVT
REVPV
EKIV
ERIVEVPNIT
EKLVVV
EKIVEVPKVV
EKVVVL
EKIVEVPKVVEQVVVV
EKIVEVPKVI
DRVV
EKSVVV
EKIVEVPKVV
EKLI
EKVVVI
EKVIEVPKVV
EKLV
EKIVVV
EKVVEVV
KEVKS
KDLIEVPV
ERVV
EKVMEVSNTHVVEVPI
EKII
EKVVEIPIEVVVEHVVEVPV
EKVVKFPV
EKIVEVPVEIYI
EKII
EKVVEVEVPVEV
ERILEV
EKAVEV
EKLVEVPI
EKTIEVMKTV
EKIVEVPV
EKII
EKIV
EKVVEI
EKVVEVPV
EKVVEI
EKVTH

24

Eutreptiella CAMPEP_0113747498 (second half of sequence)

SITQIPPAIETTFSFERESQSE**G**RHQYADVSSGESVERYHSSMSSSTHGERY
QVVESMTTSSPSIVSGGATSPRVGQSIRVADTTNYGQSSSVMSTTSSERDGP
MEVVEGSPRRYFSSISPRGTSEYKSIKAGAI



25

Eutreptiella CAMPEP_0113748194

MSIYVGEQRYSSPQQYAAERQRYVSEPRPLDHIRVADAGYKYGTGDSKVVQG
KRVRSQVVNVPVVHHDVRVPVEVEYPVEVPVEVPYEVSIPREVIREVTVPVERI
VESVN

ERKVI
KEIPVPQEFETIV
EKVV
EKRVSVPGPVKYV
EKIVEV
EKIVEV
ERPVWRI
KEIPVPYPV
EKVV
EKIVNVPKYI
DRKVEVPQDRLVQVP
REREYI
REVV
KEVRVPYPV
ERIV
EKVVKVPQ
ERFV
EKKVPFIV
EKVVEHV
EKVEFT
KEVHVTV
EKEVIEEVEVRVPV
EKVI
EKRVPKII
EKVVEI
EKPRIVQKIV
EKTVHMPY
EKIT
EKTIEVPI
ERHVHVPV
ERVI
EKQIEVPI
ERVV
EKVVQVPQEVD

GGDVDKVVWRDIPVFIDKVVYNDYTESGHLFAYHNMPTPPPPLPNFKWVPYSYNALTLGT
GRSAKGGPPLARSLVPPRGPPVPARLPGGRVGTIGRLPLPLGDARXXGNLGQLNIIDRFIF
VDPEGNVVDERADLWTERGLDPNIMHQKVSAAQAAGFHRSDKFWEHHFDELAHAAAQR
PIVDTQHFDHPVELNYHGHVSPTGRAYYPXLLAGGTATGTRVPTSLPAVLTVSLRGAASAL
RVWLPTDEDSGVAGALAGTGLLAAPLGPRAFDSAAPNMDEEAVCGGGAPPHPVGQPEP
REEGSX

26

Eutreptiella CAMPEP_0113750188 (First half of sequence)

Articulin

MHAAAPRTPPRQGGSVTRPGSGVXXXXXXPHPHQLSGGVPCREAGSKNASHDGMV
MQVVQESSQMGSRLAPSYNRGYLTSQEVAVARSRGSTMPSSQSSSLTRSDVIVVDE
GRTYYDSGMLSHPGSRVRHGGYS

ERSVV
EKREIEIEVEVDNI
REIVEIPV
EKIIEVPV
ERVV
EKIVQVPQVT
ERDVLV
EKVV
ERIVQVPVETRVEVPMVQTI
EKVMAVQVENV
REVPV
EKYR
EKLVEVPV
ERVIEVGV
EKLV
ERVVQVHVDNI
REVPV
EKVI
EKLVQIPVETI
REVPV
ERIV
EKVVEVRVPVPV
ERVI
EKIVEYPVQNI
REVVV
EKVV
ERI
REVPV
ERLVEVPV
ERIV
ERIVEVPIHNI
KEVPVVRLV
EKVVQVPVEIPV
ERVI
EKIVKVPVENI
REVPV
EKIV
ERVVEVPV
ERVV
ERVV

27

Eutreptiella CAMPEP_0113750188 (Second half of sequence)

EKQVPVEV
EKIVPVENIV
EKLVTVQVEHI
KEVPV
EKIV
ERIVEVPV
ERVV
EKIV
EKVV
ERVVEIPIEI
EKIV
ERVEVVEVPIEI
EKII
ERTVEVEVEVPVIV
EKVVEI
EKIVEVPIEI
EKIV
EKTIVE

SVGSAPSASVTVTTETSSSASGYSSGGSGYVAGSWSSASKEQAAPGGANGYS
AFSNSAHRHANMLAALDAADGKLDGQVFSADVKTILPGSAPTAYVPPPTTRDKL
AXEGQPQCQPIHRQPPVWLQQFRDPNPRIRVDS

28

Eutreptiella CAMPEP_0113750414 (First third of sequence)

Articulin

XVVMSGATQHPSYATYPSLQAPTYVGLASSAHHAAAEYRGPYASLQNTVEAS
GIGALMPATISGPPPLLPVTAQYQPPSGQEVGQGWMEEQSTVTPGNMQVESL
TEISTHASKAKKSNKVSSGSGRSGLVRYVEVPV

EKVV
EKVVEV
EKIV
EKAVPV
ERIV
EKVVEV
ERIV
EKAIPV
EKIVTIPQVTT
REEKIV
DRVVKVPQKHEVVVPVEHIV
ERVVQVPTTV
EKVVVV
EKVV
ERPVEIPVATEHIVVV
EKVV
ERVVRVPEVLEVPVEHIVKQI
EKVPEVRTVEVPV
ERII
ERLVEVPQVT
EKEVPV
ERIV
ERIVKVPEVHTVEVPV
ERLL
EKLVEVPLVV
EREVPV
ERLV
ERIQLVP
EVRTVEVPVEVV
RERIVKVPEVHTVEVPV
EKYV
ERLVEVPQIIE
KVVVNTV
ERIQTVPEVCTIEVPV
EKRIETINKVPMETI
KEVPV
EKII
ERIVEIPQLTTIEVPV
EKVV
ERVQQVTVQNI
KEVPV
EKXV
ERIRKVPMVTI
KEVPV
ERIH
EKIVEVPQVLTIEVPV
EKVV
ERVQQVVVQNI
KEVPV
EKLV
ERIQKVPVETI

29

Eutreptiella CAMPEP_0113750414 (Second third of sequence)

KEVPVENII
EKVVQVITQNV
REVPV
EKTV
ERIMKVPVETV
REVPI
EKTI
ERLQKVPMETI
REVPV
EKII
EKIVEVPLVI
ERAIPV
EKVV
EKIVEVPQLTTI
ERPVETII
ERVVQVNTENI
REVPV
EKTI
ERVHKVPMETI
KEVLV
EKVV
EKIVEVPQVI
ERAVPVETIV
ERILEVQVQNI
KEVPV
EKVV
ERVVQVPVEII
KEVPV
EKVMTVEVAKIVHQPVEVI
KEVPI
EKVVKVPEVHV
KEVPV
EKVI
ERIVQVPVETV
KEIPV
EKIVNVEVAKIVNVPVEVL
KEVPV
ERIVKVPEIHV
REVAV
EKLI
EKLVEVPV
EKII
EKHIYIDRHVEVI
KEVPV
EKIV
ERIVEVP
KETVKMVVV
EKVV
ERVVEVP
KETVRSVVV
EKVV
EKVVKVPQLHEIEVAVEQIY
EKIVEVPCETT
KEVPI
EKII
ERIVEVPV

30

Eutreptiella CAMPEP_0113750414 (Final third of sequence)

ERII
EKAVTV
EKIVEVPVRI
ERTVEVPVTEIVEV
EKLVYVNVP

HHAEGPVQMDDRVSPFDWSCAFVDCSNHVVPGYRFCATHQVDHGLSRSANSL
NIPIPHVGVPEPHKRTEKDYQALADMIREDGERRRSANSSSISQAHSSASSFSSIT
RMSGATKTKSYEGIKSEPIIARLTTDDGTRTINIPMHRVVVLGSSSELSLEYRFATF
EPQHCSLSSTSAGIVLRDLSDSGVFVNEKAVGKQVSVMLQDKDRVAIGKARVSF
VFHPT

31

Eutreptiella CAMPEP_0113750944

Articulin

…PQYSVQPLQMKQAEYQPAIYSMPQMQQAPAPVQENRVEYKEVIKEVEVPV
PVEVPVYQDEPEYYEVPEYVRVEVPVPHEVPRYVEVPQPYTV

EKVVEVPEPYEVV
KEVTVPVNVTYEVEVVETVNVPYPVEQVV
EKVVVVP
KEQIV
EKVVEVPV
EKVV
EKVVEVPQSVSVPYVVEIPVPYET
EKIIEVPKPYEVIKYIEVPTPVTKEV
EKIV
EKRIPNYVKV
REEIKVPSRRPKLWIKNVPV
ERV
ERVEVPYKVHKYVDVRKPYEVIKKI
EREV
EKVVEVKVL
KEKAVPHKI
DKIVEIPLPYYV
EKIVDVPQPYTVKKII
EKEVEVPYTVLV
KEEVRKPYEVTKVVNRPVPQKV
DKVVLKYV
EKPFAQHYEVRVPKYYDVPVPYEDIKY
RDVPFPV
EKLV
DKVVEVPVP
REVPFKVFNEIPVPV
EKIIEVPKPEPYDTVV
EKEVIVYQ

HEPHEVIQEIELVEHIVHPIEIIEEQVIKRKIKKLRKKVYGEPVLCDAPFTGQPVDQAPPDGK
SGKSGKSGKSKKGTGATNPQIESGPPLAFPAAAYGMPGMPGMPGMPGMPLFGMPMPY
GVPMAF

32

Eutreptiella CAMPEP_0113750946

Articulin

XFFRLGVSMTTYEELKARREREKRLFEVPSDVLSDRTTERRVVQPAYVQDD
MGALAIRARPWTSQSPSTGELSCLPITLCPXSVLS

REVPIQVPVEVP
REVPI
DRVFKHT
EKIPVPVR
RERKIRVPVTKIVKKPVTKYVDVPVEVIT
EKVLIKRKKKIVEVPV
ERIV
EKPV
EKIIEVKV
EKIVEVPV
ERVV
EKKVNVPYEQVV
EKVVNVPVEILVNKVV
EKVV
EKIVEVPVEQVV
ERTVEVTI
ERIIQKNIEIPV
ERVQENIITIPIIKII
EKVVHVPV
EKII
EREVPVQV
EKIV
EKYDT

SDPIYVDKTVNVVVEPTVHTTTTYETEYNYDVDGDWGGYGDNTEVTTEVITTTEY
DGDWGGDAGVTTTTTEYTYGDDGGYGGDDGYGGY

33

Eutreptiella CAMPEP_0113753042

MPEPEGSPADLQAADDIEIDVSVETLVHVDQRQVVDKPFEVSIEHLEEVPVE
QVLESVN

ERLVL
REVRVPI
ERVV
ERVVGVAIEVPY
ERVV
ERVVEVPV
EKVVP
KEIVKVI
EKIVEVPVTRVLKTKKITPV
ERIV
EKRVEVPTETTVIRRVEVPYTRIV
ERFEEVII
DKIV
EKRVEVFV
DKIV
EKIVHVPHDVII
EKVVEIEEEEEIVEVPV

THHETTYLDVPQRITLQDQGYKPPKAVILNHRQPVPLPTQHQPHMHHHQPTQHPIT
PIMLQQPYPPTPVTALHTDWLRDRVQAAEEKNHWLQGRRYEPLSHDVEAPVRRLS
APDPGPGSGHRMHGSTAVHEPPYAAPSSVSSRTRSRSADQYAA

34

Eutreptiella CAMPEP_0113754010
(First half of sequence)

Articulin

…PEVTADAELTAFFLAPDPVGVDGMDGGLPEVLEEEMNDGLFAEPDHTLSQL
ALPGPPGGEAMAVATQRLAGKLKATVSPKHLVFSKLIPALKRACKRSDLLEEM
EADGLTTIPPPSDWVLAKMAREWRCWLPSGPGYFDYASQPVPWELVGGRHW
PNSRGDFDERDTGHEGFSTQDFLNRPEAKAAGLTAAEVIALRLYTGPGYIPINR
SLRVNSGRFAVTQWALDCAIGKLALAEREGLLLRGLRLLPKEEWQQQYEDCR
CADDAMDLWISDPAYSSTTTDMAVATGTDFGGPCTFVFHAQCDLSPADGLIG
NAASVQWVSQYPDEVERLLPSNSCFLSLPQGMRSELPEGMGDRKIFQFFSRF
LWDYERCCPPVVTEVEEYVVRINEVMWHVYRGLGSEKPPPEVTADAELTAFFL
APDPVGVDGMDGGLPEVLEEEMNDGLFAEPDHTLSQLALPGPPGGEAMAASS
LNGLPTFCAVSQEDPVAVETRIEVPTVEKVTAVQAEDIQEVPMVEVPV

ERVIEVPV
EKII
ERIE
RIPVETI
REVPVEV
EKIVHVPVEHVV
EKVVEVPV
ERVV
EKVLQVPVEV
EKVV
EKVVEVPV
EKIV
EKVFQVPVEV
EKIIHVPVEHVV
EKVVEVPV
ERVV
EKVLQVPVEV
DKVA
EKVVEDVV
EKVVEVPV
ERVV
EKVLQVP

35

Eutreptiella CAMPEP_0113754468
(Second half of sequence)

XKVLQVPVEV

EKIIHVPVENVV
EKVVEVPV
ERVV
EKVLQVPVEV
EKIIHVPVEHVV
EKVVEVPV
ERVV
EKVLQVPVEV
DKVA
EKVVEVPVEQAVI
RDVPV

EKTVEMIVEGPNMLDAAIGLAEEGKPFVFEAANGYGLTARHVEAGQQLKNKR
ADNDRWRLHRMSPGEFTVESLSAPGLRLGVGNQSGHGFKAVLVPEGDQRA
LLRLVPARDGSQQRVSFESVSQPGSLLNHCNGLMWFFDRPANQHFSNDSSW
VLLSSEKENSKTSRRGGTAQLRLEGKYLNLDESSGFTIRPRAGQTIVQPVTFT
SAQDVYPGHNPHTWTLKGTNDXGGMGITSGWGLSLEGSGFREYIRFENWKA
YRMYRVTPPPGPELHPRKVDADRAVRRPRPGRRQPPAKKKRERKREEGLSQ
GWX

36

Eutreptiella CAMPEP_0113754992

Articulin

MAEARTGPSAMDALEDADRRTATRQTRSAAEALDAVDGVMDGKFFGRPIVAT
GPSRVVGGRSVVDSIDAFDGRRYRSSGGVHGSAAEALDAADGVMDGRFYGR
PIVETRSPQRLRGVTAQALDAADGVIDGRFYGRPIVETRGATTRYDDHVEVRR
VHGGTRHRSKVVEVPVYHHENRIPVEVQVPVQIPVEVPVRI

ERPLAI
ERVVEVP
REVIKHVEVLVEVPHEVKVPYAV
EKVI
ERIVENVVEVPGPTKVV
EKIVEV
EKIVEVEVQV
ERI
REVKVPFLTXV
ERVV
EKLV
ERTYPV
DRIVEVPV
ERIVHDHVDVP
RERIV
ERIVEVPYPV
EKIV
EKISEVADVRVV
EKVVQIPV
EKVV
EKVTTVPI
EKIVKVPI
EKPVIKLVH
KEVPVEHIV
EKIVEVPV
EKIIEVEVDVPV
EKVV
EKVTRIPVEIPI
EKFVEVPI
EKIVEVPVEHVV
EKIVEVPV
ERII
EKLVKVMVPVPGPPHENIIY
RDREVPI
EKRVY
RDRIW

GGRQRGYEMPPPPPPPPRDVQWISVPEAYPHQTWEWDRWQNWWDGRYGSNMQY
DGGYNAGYPPAAPQSYSNMQYDGGYNAGYPPAAPQSYRYGA

37

Eutreptiella CAMPEP_0113757424

Articulin. Repeat in ABD domain

MSYEEYGGYGDAEGYGEATTESWTEGYGEGGGEAYAEGYGEASGAAYGEASG
AAEAYGEASYSAYGEASGAAEAYGEAYGDGGGEEYTHTDSWVEHSGGGDGGG
AEQWDQHEWTEEHEWGEPGGYTSEYKMEALPDVDLGGYSEFTPTTDTYDGQA
GVISSSYQVEGEGQKVYTDWSYQEPTPIKEYVPYVSPEIEPEGEKVVEMIHGEPV
LTETKILAENQLDEVIVSVDLLSEGKVSEVETGRRIVSKQRVIENRVEQVVQVPVET
FEEHLVVN

EKPKPVRVNV
DRLVA
KDV
DKIIY
KEVKVPIEVK
KEVVVKQEVEVPV
EKVI
ERIRKVPC
ERII
EKVV
EKVV
EKIVEVPV
ERIVIEEEEIFV
EKII
EKIIEVPVPKIV
EKVIKVPV
ERIV
EKVIKVPV
ERIV
EKIVEVPV
ERIV
EKIMEVPVENII
EKTVEVLVPSRKQHYVEQVVVE
ERPVEYT
RERAPVYVE

Repeat

ERAPVYVE
ERAPVYVE
ERAPVYVE
ERAPVYVE
ERRVE
ERRAPVYMK
ERAPVYVE
ERAPVYVE
ERAPVYVE
ERRVE
ERHPVYVE
ERAPVYVE
ERAPVYVE
ERGPVYAK
EREPIY

TSSTYGTRRYADQEPVYANETSASPIYVDERSAGAYY

38

Eutreptiella CAMPEP_0113758370

Articulin

XFFRLGVSMTTYEELKARREREKRLFEVPSDVLSDRTTERRVVQPAYVQDDM
GALAIRARPWTSPSRDFVGHDKTVSVY**GG**TVVPSHNTVSTSVLS

REVPIQVPVEVP
REVPI
DRVFKHT
EKIPVPVR
RERKIRVPVTKIVKKPVTKYVDVPVEVIT
EKVLIKRKKKIVEVPV
ERIV
EKPV
EKIIEVKV
EKIVEVPV
ERVV
EKKVNVPYEQVV
EKVVNVPVEILVNKVV
EKVV
EKIVEVPVEQVV
ERTVEVTI
ERIIQKNIEIPV
ERVQENIITIPIIKII
EKVVHVPV
EKII
EREVPVQV
EKIV
EKYD

TSDPIYVDKTVNVVVEPTVHTTTTYETEYNYDVD**G**DWGGYGDNTEVTTEVITTTEY
DGDWGGDAGVTTTTTEYTYGDDGGYGGDDGYGGY

39

Eutreptiella CAMPEP_0113758614

Articulin

MTTTTGAAARALDLADGVEDGKFYGRRIVEGNSRVLTGSRVYTRSPSRSVVR
TEGSVIRALDAADGVIDGKFYGSRIVDGARLGYSTQTEYGSRIVDGSSQVYSSI
HRPRYSSITRGERRVVDVSSEATYKQKRVEVPVVHHETRVPVEYEVPVHIQVE
VPVRVEVPYPV

ERIVEVPQDVVQ
ERQVIVDVPVEVRVPYPV
ERVRHVPV
ERIIEVQGPTRFVENVVHV
ERLVEVPIEVPVVRTV
ERPYPVEQIVEVMVEQPVPVEHIVQVPEEYIVEEIIEVEQ
ERLVEQVVEVPVPVEHIVEQRVPVPVPHIVEEIVEVPVENIV
EKLIQIPV
ERV
REVPVEQIVRQARYV
DRPYEVL
REKVVHVPV
ERVRHVHVDVPV
ERIV
EKIVEVPFEVPV
EKVVQVPV
ERVVQVPVEHVV
ERPV

TRYSAPTYAAPPLRPVHHNPAFALDAADGRIDGQYFGARIAQPYNPALALDAADG
RIDGTYYGSTIAPQPYGPGYRPY

40

Eutreptiella CAMPEP_0113766106

MSIYVGEQRYSSPQQYAAERQRYVSEPRPLDHIRVADAGYKYGTGDSKVVQG
KRVRSQVVNVPVVHHDVRVPVEVEYPVEVPVEVPYEVSIPREVIREVTVPVERI
VESVN

ERKVI
KEIPVPQEFETIV
EKVV
EKRVSVPGPVKYV
EKIVEV
EKIVEV
ERPVWRI
KEIPVPYPV
EKVV
EKIVNVPKYI
DRKVEVPQ
DRLVQVP
REREYI
REVV
KEVRVPYPV
ERIV
EKVVKVPQ
ERFV
EKKVPFIV
EKVVEHV
EKVEFT
KEVHVTV
EKEVIEEVEVRVPV
EKVI
EKRVPKII
EKVVEI
EKPRIVQKIV
EKTVHMPY
EKIT
EKTIEVPI
ERHVHVPV
ERVI
EKQIEVPI
ERVV
EKVVQVPQEVD

GGDVDKVVWRDIPVFIDKVVYNDYTESGHLFAYHNMPTPPPPLPNFKWVPYSYNALTL
GTGRSAKGPVPARLPGGRVGTIGRLPLPLGDARMLGNLGQLNIIDRFIFVDPEGNVVD
ERADLWTERGLDPNIMHQKVSAAQAAGFHRSDKFWEHHFDELAHAAAQRPIVDTQH
FDHPVELNYHGHVSPTGRAYYPXLLAGGTATGTRVPTSLPAVLTVSLRGAASALRVWL
PTDEDSGVAGALAGTGLLAAPLGPRAFDSAAPNMDEEAVCGGGAPPHPVGQPEPRE
EGSX

41

Eutriptella CAMPEP_0113768498

Articulin

MAEARTGPSAMDALEDADRRTATRQTRSAAEALDAVDGVMDGKFFGRPIVAT
GPSRVVGGRSVVDSIDAFDGRRYRSSGGVHGSAAEALDAADGVMDGRFYGIV
ETRSPQRLRGVTAQALDAADGVIDGRFYGRPIVETRGATTRYDDHVEVRRVH
GGTRHRSKVVEVPVYHHENRIPVEVQVPVQIPVEVPVRIERPLAIERVVEVPRE
VIKHVEVLVEVPHEVKVPYAV

EKVI
ERIVENVVEVPGPTKVV
EKIVEV
EKIVEVEVQV
ERI
REVKVPYPV
ERVV
EKLV
ERTYPV
DRIVEVPV
ERIVHDHVDVP
RERIV
ERIVEVPYPV
EKIV
EKISEVADVRVV
EKVVQIPV
EKVV
EKVTTVPI
EKIVKVPI
EKPVIKLVH
KEVPVEHIV
EKIVEVPV
EKIIEVEVDVPV
EKVV
EKVTRIPVEIPI
EKFVEVPI
EKIVEVPVEHVV
EKIVEVPV
ERII
EKLVKVMVPVP

GPPHENIIYRDREVPIEKRVYRDRIWGGRQRGYEMPPPPPPPPRDVQWISVPEAY
PHQTWEWDRWQNWWDGRYGSNMQYDGGYNAGYPPAAPQSYSNMQYDGGYN
AGYPPAAPQSYRYGA

42

Eutriptella CAMPEP_0113754468

Articulin

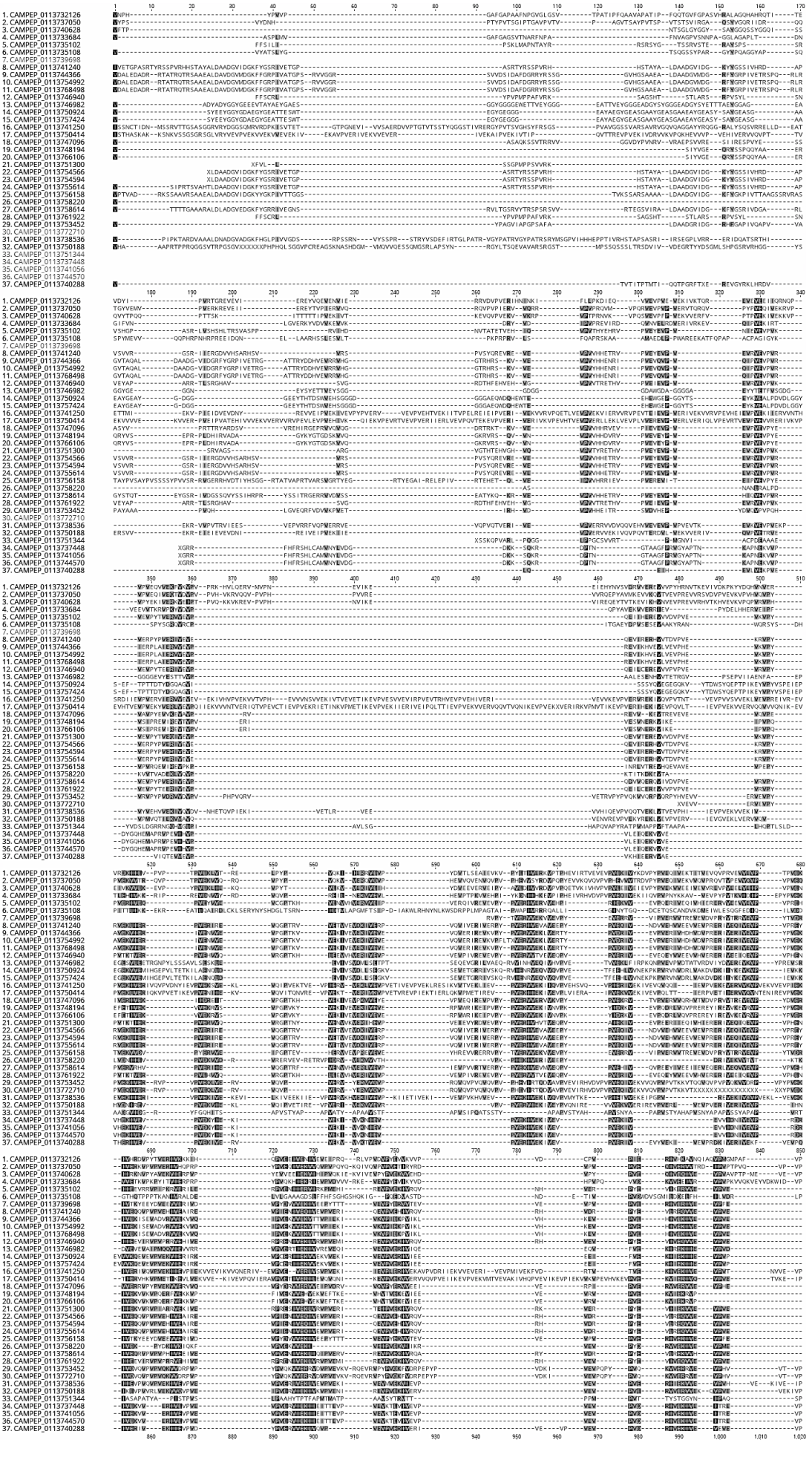
XKVLQVPVEV

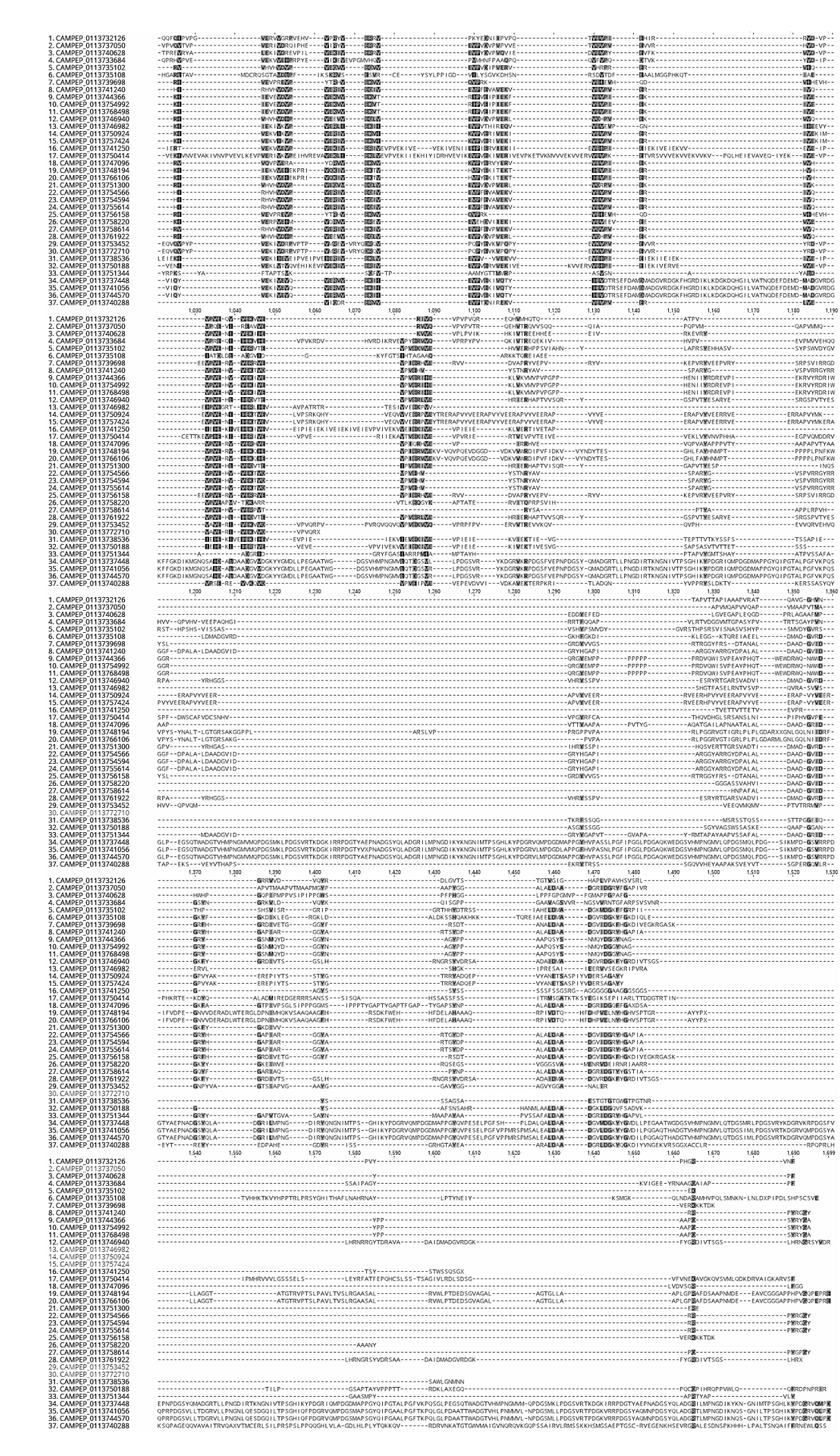
EKIIHVPVENVV
EKVVEVPV
ERVV
EKVLQVPVEV
EKIIHVPVEHVV
EKVVEVPV
ERVV
EKVLQVPVEV
DKVA
EKVVEVPVEQAVI
RDVPV

EKTVEMIVEGPNMLDAAIGLAEEGKPFVFEAANGYGLTARHVEAGQQLKNKR
ADNDRWRLHRMSPGEFTVESLSAPGLRLGVGNQSGHGFKAVLVPEGDQRA
LLRLVPARDGSQQRVSFESVSQPGSLLNHCNGLMWFFDRPANQHFSNDSSW
VLLSSEKENSKTSRRGGTAQLRLEGKYLNLDESSGFTIRPRAGQTIVQPVTFT
SAQDVYPGHNPHTWTLKGTNDXGGMGITSGWGLSLEGSGFREYIRFENWKA
YRMYRVTPPPGPELHPRKVDADRAVRRPRPGRRQPPAKKKRERKREEGLSQ
GWX

Alignment of *Eutreptiella* articulins into two groups

Group 1 N-terminal half





44

Alignment of *Eutreptiella* articulins into two groups

Group 1 C-terminal half

45

Alignment of *Eutreptiella* articulins into two groups

Group 2