

Supplementary Materials

SEC:Signal peptide

TAT signal peptide

>C6WIU3 ACTMD Actinosynnema mirum ATCC 29888 Amir_4331
MSRVKPRLAAGAGLSAAAVVGAFLVQPATADGFTGPVVGLGGKCLDVAGASSANGTAV
QLYTCNGTSAQSWTVDTGAKTIKALGKCLDVDDGRHDGALLQIWDCAVPEQRWTVDTQT
GRVVGQSGKCVDPAGNTADGTRLQVWTCDAWSPYQRWTPGGGTNPTTTTTSQVPGRS
DLATPAKKDIAMRLVSSAENSSLDWRAQFAYIEDIGDGRGYTAGIIGFCSGTGDMLELVE
HYTSVKPNNALAGYLPALRRVNGTPSHEGLDPGFQQAWRTAAQDPVFQQAQETERDRVYF
TFSVNQAKSDGLRALGQFAYYDAIVMHGGGTDQASFGTIRRKAMARAKTPAQGGDETAYL
HAFLDERVVMKLEEAHEDTSRVDTAQRVFLNNGNLDLNTPLAWKVYGDPPFRID

>tr|D8I4T1|D8I4T1_AMYMU Chitosanase OS=Amycolatopsis mediterranei (strain U-32)
GN=AMED_6991 PE=4 SV=1
MSKKLRPVLVGLTGAASVAALVITTPALSSAAPAPLAASSASVLATGDLAPAKKEIAMK
LVSSAENSSLDWKAQYKYIEDIGDGRGYTAGIIGFCSGTGDMLELVEAYTNSVPNNPLAK
YLPALRKVNGTASHSLGSAFESAWKQAAATTAFTAQNNERDRVYFNPSVNVQKADGLS
NLGQFAYYDAIVMHGPGDSSDSFGGIRKAAMKAKTPAQGGDEATYLLKAFFAARKVIMKQ
EEAHADTSRVDTEQAKFLSEGDFDLHTPLKWKVYGDSYTIIN

>AORI_1736 Amycolatopsis orientalis HCCB10007
MSDTLRPAMRRALGAVSVLALAVSGPVAISSLASAAPVPASASIFASGGDLASPEKKEIA
MQIVSSAENSSLDWKAQYSYIEDIGDGRGYTAGIIGFCSGTGDMLELVERYTNSVDPNPL
AEYLPALREVDGTESHEGLTGFENAWREAAGREDFKAAQNSERDRVYFDPAVNQAKSDG
LGALGQFIYYDAIVMHGPGTSSDSFGGIRSAALQQAQPPAQGGDEATYLLKAFLDARRVVM
KQEEAHADTSRIDTAQLVWLNDGNLHLHTPLKWAVYGDPEYID

>Q9LBG4_9PSEU Chitosanase OS=Amycolatopsis sp. CsO-2 GN=cho PE=4 SV=1
MHPSKHRTARTTRTALAVLVGVLPPLTLAAAGTGHAAPAAPAAAASVAVSVGLDDPAKKE
IAMELVSSAENSSLDWKAQYKYIEDIGDGRGYTAGIIGFCSGTGDMLELVQHYTDLKPGN
VLAKYLPALKKVNVTDSHSLGSAFVNDWRATAAKDTVFQRAQNDERDRVYFNPAVKQAKA
ERLRALGQFVYYDAIVMHGPGSSSDSFGGIRAAAMKAKTPAQGGDEATYLLNAFLDARKV
IMKQEEAHADTSRVDTEQRVFLNAKNFDLNPPLKWKVYGDPEYQING

>E1UPU6_BACAS Chitosanase OS=Bacillus amyloliquefaciens (strain ATCC 23350 /
BAMF_3066
MRSGLKKKAGFWKKTAVSSLIIFTMFFTLMMSGTVLAAGLNKDQKRRAEQLTSIFENGKTE
IQGYVEELDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRRLAKDES
DDTSLNKGFSAWRS LGNDKAFRAAQDEVNDRLYYQPAMKRSQAGLKTALAKAVMYDTV
IQHGDDPDSFYALIKRTNKKMGSPKDGTDKKNLNFVLDVRYDDLMPNPSDEDTQDEW
RESVARVDVFRDIVKAKNYNLDPPIHVSSEYGNFTIQ

>tr|Q0PVM7|Q0PVM7_BACAM Chitosanase OS=Bacillus amyloliquefaciens PE=4 SV=1
MKAKVDSWKKTAVSLLIIFTMFFTLMMNNTVFAAGLNKDQKRRAEQLTSIFENGMTIQYG
YVEHLDPGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRRLAKEESDDIS
NLKGFDSAWRS LGKDKDFRAAQDTVNDRLYYQPAMKQSDNIGLKTALAKAVMYDTIIQHG
GGDDPDSLNSLIKRTNKKAGGSPKNGVDEKKNLNFVLDVRYDDLMPNADPDTRDEWRESV
ARVDVLRSLAKANNYNLNGPINVYSEYGFVVIK

Figure S1. Cont.

>tr|A7Z8H9|A7Z8H9_BACA2 Csn OS=Bacillus amyloliquefaciens (strain FZB42)
 GN=RBAM_029740 PE=4 SV=1
 MKISLKKKAGFWKKTAVSSLIIFTMFFTLMMMSGTVFAAGLNKDQKRRAEQLTSIFENGKTE
 IQYGYVEALDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRRLAKDES
 DDISNLKGFASAWRSLGNDKAFRAAQDKVNDSLYYQPAMERSENAGLKTALAKAVMYDTV
 IQHGDGDDPDSFYALIKRTNKKMGSSPKDGTDEKKWLNKFLDVRYDDLMPDSEDTEQDEW
 RESVARVDVFRDIVKEKNYNLNGPIHVRSEYGNFTIQ

>tr|Q9ET84|Q9ET84_BACAM Chitosanase OS=Bacillus amyloliquefaciens PE=4 SV=1
 MKISLKKKAGFWKKTAVSSLIIFTMFFTLMMMSGTVFAAGLNKDQKRRAEQLTSIFENGKTE
 IQYGYVEALDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRRLAKDES
 DDISNLKGFASAWRSLGNDKAFRAAQDKVNDSLYYQPAMKRSENAGLKTALAKAVMYDTV
 IQHGDGDDPDSFYALIKRTNKKMGSSPKDGTDEKKWLNKFLDVRYDDLMPADEDTQDEW
 RESVARVDVFRDIVKEKNYNLNGPIHVRSEYGNFTIQ

>tr|E3E2V1|E3E2V1_BACA1 Chitosanase OS=Bacillus atrophaeus (strain 1942)
 GN=BATR1942_06760 PE=4 SV=1
 MKAKVDSWKKTAVSILIFTMFFTLMMNDTVFAAGLNKDQKRRAEQLTSIFENGMTEIQYG
 YVEHLDPGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRRLAKEESDDIS
 NLKGFDSAWRSLGNDKDFRAAQDAVNDRLYYQPAMKRSENAGLKTALAKAVMYDTVIQHG
 DGDDPDSFYSLIKRTNKKAGGSPKDGIDEKKWLNKFLDVRYDDLMPDTPDRDEWRESV
 ARVDVLRSLIAKANNYNLNGPINVYSEEGDFVIK

>sp|P33673|CHIS_BACCI Chitosanase OS=Bacillus circulans GN=csn PE=1 SV=2
 MHMSNARPSKSRTKFLLAFLCFTLMSALFGATALFGPSKAAAASPDDNFSPETLQFLRNN
 TGLDGEQWNNIMKLINKPEQDDLNWIKYGYCEDIEDERYTIGLFGATTGGSRDTHPDG
 PDLEKAYDAAKGASNPADGALKRLGNGKMGKGSILEIKDSEKVFCEGKIKKLQNDAAWRK
 AMWETFYNVYIRYSVEQARQRGFTSAVTIGSFVDTALNQGATGGSDTLQGLLARSGSSSN
 EKTfMKNEFHAKRTLVDVNTKYNKPPNGKNRVKQWDTLVDMGKMNLKNVDSEIAQVTDWEM
 K

>tr|A0EQW7|A0EQW7_9BACI Chitosanase OS=Bacillus sp. DAU101 GN=csn PE=4 SV=1
 MKISLKKKAGFWKKTAVSSLIIFTMFFTLMMMSGTVFAAGLNKDQKRRAEQLTSIFENGKTE
 IQYGYVEALDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRRLAKDES
 DDISNLKGFASAWRSLGNDKAFRAAQDKVNDSLYYQPAMKRSENAGLKTALAKAVMYDTV
 IQHGDGDDPDSFYALIKRTNKKMGSSPKDGTDEKKWLNKFLDVRYDDLMPDSEDTEQDEW
 RESVARVDVFRDIVKEKNYNLNGPIHVRSEYGNFTIQ

>tr|Q9RC18|Q9RC18_9BACI Thermostable chitosanase OS=Bacillus sp. KFB-CO4 GN=
 Q9RC18_BacKFB PE=4 SV=1
 MREADWWGKAAISLLVFTMFFTLMMSETVFAAGLNKDQKRRAEQLRRICEDGTTEMRYPY
 VARLDDARPSTCGPAGVTATRDGFEVVPVYKQAVANKKLPNYLAGLRRLEKEASDDTSK
 LKGPASAWKSLADDKAFRAAQDGVNDQVYYQPAMERSDNAGLTTALARAVMYHTVRQRGD
 GDDGDSRYALIKRTPKGAGGSPKEGIDEQKCLNKFSHVRYDDLMPGANHRRRDEWRESVG
 RVHVLRSIANQNNYNLNGGIHVRSEYGNFVIK

>tr|D0FHB6|D0FHB6_BACIU Chitosanase OS=Bacillus subtilis PE=4 SV=1
 MKISMQKAAISLLVFTMFFTLMMSETVFAAGLNKDQKRRAEQLTSIFENGTTEIQYGYVE
 RLDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRRLAKEESDDTSNLK
 GFASAWKSLANDKEFRAAQDKVNDHLYYQPAMKRSDNAGLKTALARAVMYDTVIQHGDD
 DPDSFYALIKRTNKKAGGSPKDGIDEKKWLNKFLDVRYDDLMPANHDTREWRESVARV
 DVLRSIAKENNYNLNGPIHVRSEYGNFVIK

Figure S1. Cont.

>tr|D4FZ80|D4FZ80_BACNB Chitosanase OS=Bacillus subtilis subsp. natto (strain BEST195) GN=BSNT_03901 PE=4 SV=1

MKISMQKAAISLLVFTMFFTLMMSETVFAAGLNKDQKRAEQLTSLIFENGTTETIQYGYVE
RLDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRLLAKEESDDTSNLK
GFASAWKSLANDKEFRAAQDKVNDHLYYQPAMKRSDNAGLKTALARAVMYDTVIQHGDDG
DDEPSFYALIKRTNKKAGGSPKDGIDEKKWLNKFLDVRYYDDLMPANHDTRDEWRESVARV
DVLRSIAKENNYNLNGPIHVSNEYGNFVIK

>sp|O07921|CHIS_BACSU Chitosanase OS=Bacillus subtilis (strain 168) GN=BSU26890 PE=1 SV=1

MKISMQKADFWKAAISLLVFTMFFTLMMSETVFAAGLNKDQKRAEQLTSLIFENGTTETI
QYGYVERLDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRLLAKEESD
DTSNLKGFASAWKSLANDKEFRAAQDKVNDHLYYQPAMKRSDNAGLKTALARAVMYDTVI
QHGDDGDDPSFYALIKRTNKKAGGSPKDGIDEKKWLNKFLDVRYYDDLMPANHDTRDEWR
ESVARVDVLRSIAKENNYNLNGPIHVSNEYGNFVI

>tr|Q9XDS6|Q9XDS6_BURGA Chitosanase A OS=Burkholderia gladioli GN=csnA PE=4 SV=1

MATLPLAAADASARTVHAARAPQAAAAAASPEICSSPWTAARIYEAGNVVSFDGHDYTAA
YRSQGNAPATASGEAGSQPWVAGEACKPAKLSKAALDHDANFSPATLQFLKDNGLDGE
QWDNIMKLVNKPEQDSLWTKFYGYCEDIGDDRGYTMGIFGATTGGPNDGGPDGPALFKA
YDAASGASNPVQGLLARI GAHGS MQSILKITDSEKVF CGKVKGLQNDAAWREAMWRTF
YSVYIQYSVQQARSRGFGSALTIGSFVDTALNQGADGGSNLQGLLSRSGNSTDEKTFMT
SFYAQRTKVVDTHDFNQPPNGKNRVKQWSTLMSQGITSLKNCDADIVKVTSWTMK

>tr|E5AQ94|E5AQ94_BURRH Chitosanase (EC 3.2.1.132) OS=Burkholderia rhizoxinica (strain DSM 19002 / CIP 109453 / HKI 454) GN=RBRH_02856 PE=4 SV=1

MESALNQLISKTLTRMSCAVAMGLGLLAVSGMARAADDAQPNSEPCDAPWDAKTIYQGGEV
VSYAGRNYTAAAYWTQGNAPDLNSGDGNTGQPWRVVGQTCNAVANVASGNPDANFSPATLQF
LKKNTGLDGEQWNNIMKLINKPEQDSLQWTKYGYCQNIHDDRGTITIGIFGATTGGSNDT
GPDGPELEFKTFDALS GAKIPSAEGLTRIGAHGKMSG SILKISDSSKTFCSKVKAQQNNP
TWREAMWRTFYNVDIQYSVQQARQRFDSALTIGSFVDTALNQGADGGGDSLKALLARSG
DSPNEKTFMTRFYAERAKVVDTHDYNTSPNGKNRVKQWSKLMQGETDLKNADAAIVKVT
NWQLH

>tr|B3PI04|B3PI04_CELJU Chitosanase Csn46F OS=Cellvibrio japonicus (strain Ueda107) GN=CJA_3747 PE=4 SV=1

MLTPTQKKTAEAI VNI FETG SVLGDYSNVTLITGDTGHLTFGRSQTTLGSGLKLLQLY
CANPGARFRQQLTPFLARFAARDFSLDHEEHLKNILRATADHDHIMRETQDLFFDQAYWQP
AERAATQLGIKTPLVAVVYDSTVHGS AKLIRDKTNQSAGTLASLGEQKWIEAYVATRRY
WLANHPRKDLRPTVYRMDAFQRLIDLNLWGLELPLVVRGLEISNTTLNATPVNCYDGPQP
GTRTL SVQAPLLRGLDIRLVQLGLSKSGINLRADGIFNGSVSAVKTFQTKQNL PATGIV
DNALIAKLVS

>tr|Q7NR53|Q7NR53_CHRVO Probable chitosanase A OS=Chromobacterium violaceum (strain ATCC 12472 / DSM 30191 / JCM 1249 / NBRC 12614 / NCIMB 9131 / NCTC 9757) GN=CV_3931 PE=4 SV=1

MKYAAKRLSFPSAILAGLALLAAAGSAAAQGSTAGSPERQTVRLPDCAAWS PASPYQAG
SVASRGGVNYTAAFWTQGNPPEQQAWQAGKRCRPAAQAKAADHDANFP PATLKFLKANT
GLDGEQWDNIMKLVNKPEQDSLWTKFYGYCENIGDKRGYTTIGIFGATTGGPNDEGPDGP
TLFKEFDASSGASNPSIAGGLARAGVHGS MQGKILKISDSAKVFC DKIGSLQNNPAWRDA
MWNTFYKYVIQYSVQQARQRFSSALTIGSFVDTALNQGATGDSGTLQGLLSRSGNSGDE
KTFMTAFYAQRSKI VDTNDYNQPPNGKNRVKQWSTLLNMGETDLKNADAAVQKVTNWEMK

Figure S1. Cont.

>tr|A4KCU0|A4KCU0_9ENTR Putative uncharacterized protein OS=Citrobacter gillenii
 PE=4 SV=1
 MILTPAQKSLCERVLNAFETGSAEGDYSIAIAIFHDGPHGIRQITYGRSQTTEYSKLDLV
 TRYVNAVGRFSADLAPFVFKIGNIALVDNEPFKDLLRRAGRDDPIMKSVQDRFFDERYYL
 PALKWAQDFGFKDALSLLVLYDSFVHSGTILTLFLRAQFREVPASPAGGDEHTWISQYVEVR
 NSWLANHSNPELHPTVYRTRKCLRFEIDRGNWDL SQVPINANGIILVGPSPDISVPQNVDTS
 GGTTRQRSVFWRGSTRLLWFCG

>tr|Q47C49|Q47C49_DECAR Putative uncharacterized protein OS=Dechloromonas aromatica
 (strain RCB) GN=Daro_2852 PE=4 SV=1
 MSACYAEPHILQFFPSLPSTRRSWKMLTPTQKQTAQSIVNLFETGAVLGHYGNVTVIPGD
 TGHLTFGRSQVTTLCSGNLLDLLQRYCGNDGARFGDKLSAWLARFEAVDLSLDDDLHLHNL
 LRATADDPVMRETQDLFFDEVYQWPAKAADNFGIKSALGVAVVYDSVWHGSKWTLRDMT
 NQKSGPIATLGERAWIKAYVETRRSWLANHSRSDLRGTTTYRMDAFLRLIEQGYWGLALPL
 VVHGEEISLATLRAMPFGCYDGPQPGTRSIALQTPPLARGLDVRLQLGMSGRGVAIKADG
 IFGQTSVNLKRFQASQGLPANGIATPQQLSELVD

>tr|D0LGY9|D0LGY9_HALO1 Peptidoglycan-binding lysin domain protein OS=Haliangium
 ochraceum (strain DSM 14365 / JCM 11303 / SMP-2) GN=Hoch_2166 PE=4 SV=1
 MSNSGYQKRIQNRGSDAADPLQKQVAPGKVTRTRSRMASPVQRKQDGPQDPSAGRTEVANS
 TVFRHGGGAVDLGGSSAAEVAESGFGSGASDVPYRAEMERSFGTSFSDVQAYS GGDSQGA
 ATQLSAQAYTVGNRVAFRD SNPSRELVAHELAVVQGRGGEVQAKSEMSQPGDSLEREAD
 SVAARVASGESVQDHTARYD GAGG PALGSAPMRLVSDAAASETTAAPAADAADAAPAA
 LLTPEQVQSAIAFYGADRWPAPKIAQVSEKLGIAVREQVDEQFVQAAAAGFQQQRSLTVDG
 MLGSASIMPLFEGEEMDRAHTMANRI TAEYESSGNYGVVQNADVGIISYGAHQSTLHSGN
 LGRMLQDYLDRAVAAAEPTQASQTIQSYMGRINDRSQWESLRNEGPLLSALRAAGSEQIM
 QDAQNAFFSEDFWVPAVKAALNHGITSQLGYATLYDAKIQGGMEDSLQRATSAMGGIVGA
 TVERNRSQQVTEAEFLVAFNEAREGRLERIAVRRDQGGKRRDAEMLRNSKVRPQAFEEL
 ARDGNLDSLAVNDGENSLEFRITYGRRTEVDVPEEVGTTPATGDTEAGTGTAPPTTRPTP
 EVTPTPEALVPTTTPEVTPQPRNPQPSASEYTVKSGDLSAIAGRLLGDQRWREIATLN
 GITNPRALRVGQVLQVPSSESAAPEGGQSEPEAPAAAPVETAYVVRSGDTLGSIAARFL
 GSTNRWREIATLNGISDPRRLSVGQRLRIPTGGAQQATPEPEQQRPEQQRPEQQRPEQGG
 GGAGGGSSPEAGKPSWISVAEGELGVQEI VGSRHNPRVIEYHSTTGRFSDDETPWCASFV
 NWVLQQAGQSGTGSARALSFESYGTTLDRPAYGSIAVLAYGGGRGHCAFVVGKQGDRMLL
 LGGNQSNVNIKSFGTSQIVAYVPPGYQPPPSAFALDGATGEVGEVGGGLSDTR

>Cho4239-1_Jan
 MQKFAAHPGLPRAAVLFAIMTSLACGGGEQPGAMRKVLISSTVASQTDANAAAADGLSAP
 AKKEIAMQLVSSSENSSLDWKAQYAYIEDIAMGGYTGGIIGFTSGTSDMLELVTAYDQIK
 PGNLLSPYLPALRKVNGSASHAGLDPNFVSDWQKAAQDPVFQQSQNTLRDNAYFNPAVSL
 AKQDGLRRTLQGFIIYDAIVMHGPGTDSQRNSFQNSPQALKKALPPSKGGNETAYLNAFL
 DVRRALMQSEPDHAGNLDRLDTQQRLFLQNGNLDLNTPLNWKVNGDSYQIR

>tr|E4N811|E4N811_KITSK Putative chitosanase OS=Kitasatospora setae (strain ATCC
 33774 / GN=KSE_15150_csn1 PE=4 SV=1
 MHPSKHRTARTARTTRAALAILAGVLPPLTLAATGAGHAATPAAGGTSAVRAAAVAAGLD
 DPAKKEIAMKLVSSAENS SLDWKAQYKYIEDIGDGRGYTGGIIGFCSGTGDMLELVQHYT
 DLKAGNVLAKYLPALKKVN GTD SHTGLGS AFVNDWKTAAKD TVFQQA QNDERDRVYFNPA
 VKQAKADGLRALGQFAYYDAIVMHGPGTADS FGGIRAAALKKAKTPAQGGDEATYIKAF
 LDARKVVMKQEEAHADTSRVDTEQLVFLNAKNFDLNPPLKWKVYGDPEYINS

Figure S1. Cont.

>tr|E4NEU0|E4NEU0_KITSK Putative chitosanase OS=Kitasatospora setae (strain ATCC 33774 GN=KSE_40860_csn2 PE=4 SV=1

MSFIGSKRGTAVAVAVAGLLAVAIGGLSPATAAGTNLALNKPVTVSSTEASGYSGAKAV
DGGSTTRWSSSTEQVNDQWIRIDLGAGTAVDRVVLKWEAAYAKAYRVELSDDGSTWRQVYS
TSSGDGGTDDDLAVSGTGRYLRVYGTQRTAYGYSLWEVEAYGTGGGSTASPSPTASASAS
ASPTASPSASSSSTPPTGGVNLDDPAKKEIAMKLVSSAENSSLDWRAQFSYIEDIKDG
RGYTAGIIGFCSGTGMMLDLVEAYTAKKPGNVLASYLPAALRSVNGTDSHQGLDPGFPAAW
KQAATDPVFTQTQESERDRVYFNPAVAQAKKDGLRALGQFAYYDAAVMHGPDGLASIRSA
ALAKAKPPAQGGNETTWLNAFLDAREAEMRKEAAHSDTTRVSTEQRKFLNEGNFDLHTPL
TWSVYGEPEYSITS

>tr|E4NEU7|E4NEU7_KITSK Putative chitosanase OS=Kitasatospora setae (strain ATCC 33774 GN=KSE_40930_csn3 PE=4 SV=1

MRPARPLARRTLLAGAAAAAALVGLNLVPASAAGGGGLTTDQRRRADQLISVFENGTTV
IQYGYAENINDGRGVTAGRAGFTTNDGDALKVVRAYTEQVPDNPLAAFVPELERLAAAGS
GDTSGLEADYVTAWKRAADDPAFRRVQDAQVDERYFAPAMADADRLGLTTALARAELFD
ASVQHGNGSEYDALPALIARTSAKAGTPAAAGEDAWLDAFFDVRIDDLTHPANSATQAEW
SQSVDRVEALRRIARTGNRNLDPFTVTAFGATHTIS

>sp|P48846|CHIS_NOCSP Chitosanase OS=Nocardioides sp. (strain N106) GN=CsnN106 PE=3 SV=1

MRLKHPTARLALALVAVPRSVAAAGTVHAAPAPAGATRLAAVGLDDPHKKDIAMQLVS
SAENSSLDWKSQYKIEDIKDGRGYTAGIIGFCSGTGMMLDLVADYTDLKPGNILAKYLP
ALRKVNGTESHAGLASAFEKDWATAAKDSVFQQAQNDERDRSYFNPAVNQAKASLRALGQ
FAYYDAIVMHGPGDSSDSFGGIRKAAMKAKTPAQGRDEATYKAFLAARKTVMLKEEAH
SDTSRVDTQTVFLNAKNFDLNPPLKWKVYGDSYAINS

>tr|B2J4Y1|B2J4Y1_NOSP7 Peptidoglycan-binding LysM OS=Nostoc punctiforme (strain ATCC 29133 GN=Npun_R2009 PE=4 SV=1

MSCKIGSSYTIKPNDFLFEIAARELGDGRWREIMNPNGIPIFTEEEAENLQTGQEIPLPK
IDEPPQTQEVPGVEFFPPGTLNQNLTLTGLDAQQLTNILGMINGPEQANSKWWQTVDEEII
YGYAEDIEDGRGVTIGIYGATTKGKYNDADVIWKNYGQDYSNLPVDEIEKVVHAIANDQK
WWKAQWDAYISTYWQPTLKLKSKNYMKALTIQVGLIDTAMNAGMEDDENSENWGVEHLFTE
ASDDTDNEEDFVDRFEMELRLQFPTRDSGDMEERIGAWQKLLRDRKWMRVDLKNYVYIPIQ

>tr|O24825|O24825_9BACL Chitosanase OS=Paenibacillus ehimensis PE=4 SV=1

MHMSNAKPSKSPARKLLALLCCFTLLASLFGATALFQPSKAAAASPENFSPETLQFLRD
RTGLDGEQWNNIMKLINKPEQDDLNWIKYGYCEDINDERGYSIGIFGATTGGPRDTHPD
GPELFKAYDAAKAGNPSVEGALKRLGINGKMGSIILEIKDSEKVFCEGKIKKLQNDPAWR
KAMWETFYINVYIRYSVEQARQRFQTSALTIQSFVDTALNQGATGDSNTLQGLLARSQSST
NEKTFLLKFKHAKRTLVDVTNEYNQPPNGKNRVKQWDTLLDMGKMNLNKVNDAEIAQVTNWE
MK

>tr|E3E9P0|E3E9P0_PAEPS Probable chitosanase A OS=Paenibacillus polymyxa (strain SC2) GN=PPSC2_c3877 PE=4 SV=1

MFKISSKLI SHKTKFFMILLSFSVIASFSSLYGPLPAKALAEENGTPVQETNESNIQIIP
LTSEETTTSQTAVTAIDHDANFSPSTLQFLKDNLTGLDGEQWDNIMKLVNKPEQDSLKWTEF
YGAEDIGDDRGYTIIGIFGATTGGSNDTGPDPALFKAFDAASGASNPSIAGGLARAGLK
GKMSGSILKLSDSVVIKKIKALQNEAWREAMWRTFYDITYIEYSVQQARKRGFNALT
IGSFVDTALNQGATGDDGSLEGVLSRSGNSTNEKTFMTNFYAKRTLVDVTNDYNQPPNGK
NRVKQWSSILLASGETDLKNADVAIKVTNWEMK

Figure S1. Cont.

>tr|Q2PWA1|Q2PWA1_9BACL Secreted chitosanase OS=Paenibacillus sp. BH-2005 PE=4 SV=1
 MKFSLLLILLSFSMIFSGYLSNASQQKSYAAGNPDSNFFPATLNPLRDNTGLDGEQWNNI
 MMLVKNKPEQDDLNWIDFYGYCEDIDDDRGTIVLFGATTGGSNDTGPDPDLFKAYDAAK
 GASNPSVKGALARIGVKGSMKGKILEINESEESFCRKIGNLQDDPEWREAMWKTFYNIYI
 KYRGEQARNRGFNSALTIGSFVDTALNHGATGGSESLQGLLSKSGSSTDEKTFMTKFYKE
 RTKIVDTNEYNSPPNGKNRVKQWSNLLNMGETDLKNADSAVLQVTDWELQ

>tr|A7HRM9|A7HRM9_PARL1 Peptidoglycan-binding domain 1 protein OS=Parvibaculum
 lavamentivorans (strain DS-1 / DSM 13023 / NCIMB 13966) GN=Plav_0939 PE=4 SV=1
 MLSELQKEAAKAIVNI FETGAVRVDYARVTLLAGDSGHLTYGRAQTTLGSGNLHLLIKAY
 CEAPGAAYARACEPYLPRLADIDLRLDSDWTFRNLLKEAGADPVMRDTQDAFFDRVYWTP
 AAASAIRLGLTEALSLAIVYDSVVHGSWISLRDRTLAKAGQPSKAGERAWSLAYVRERRN
 WLAMHSNTLLRKTIVRMEAFEALIAAKNWSLALPMTVVRGIRIDEPALGYRPPVTASATDV
 TTRNLRLTSPRMTGNDVRALEAALVKEGYAINCDGVFDEGLEKALKSFQQDYGLIADGVA
 GPATRIMLG

>tr|Q8KZM5|Q8KZM5_9PSED Chitosanase OS=Pseudomonas sp. A-01 PE=4 SV=1
 MKIQRLVALAAAVSLSIGLSGCAASVEAAGTVDDLDPVQKDTAMSLVSSFENSSTDWQAQ
 YGYLEDIADGRGYTGGLIGFTSGTGMLELVRAYSASSPGNPLEQYIPALEAVNGTDSHA
 GLGQGFQAWADAAETSEFRAAQDAERDRVYFDPAVAQKADGLSALGQFVYDITLVVHG
 PGSQRDAFGGIRAEALSAALPPSQGGDETEYLEAFFDARNVIMREEPAHADTSRIDTAQR
 VFLQNGNFDLERPLTWSVYGDQFSLN

>tr|E1AXU1|E1AXU1_9PSED Chitosanase OS=Pseudomonas sp. LL2(2010) GN=CHI PE=4 SV=1
 MKIQRLVALAAAVSLSIGLSGCAASAE TAGTVDDLDPVQKDTAMSLVSSFENSSTDWQAQ
 YGYLEDIADGRGYTGGLIGFTSGTGMLELVRAHSASSPGNPLEQYIPALEAVNGTGS
 GLGQGFQAWADAAETSEFRAAQDAERDRVYFDPAVAQKADGLSALGQFAYYDITLVVHG
 PGSQRDAFGGIRAEALSAALPPSQGGDETEYLEAFFDARNVIMREEPAHADTSRIDTAQR
 VFLQNGNFDLERPLTWSVYGDQFSLN

>tr|A9WUI6|A9WUI6_RENSM Chitosanase OS=Renibacterium salmoninarum (strain ATCC 33209
 GN=RSa133209_3139 PE=4 SV=1
 MKLSCIRPQAGGAARNTRRKVLSLLELGAPVPNLRTRTVAVAVGGLLIASGAIAGTAAQA
 NAVSSLAPAITAVSAASTGDL SAPAKKEIAMQLVCSAENSSLDWKAQYGYIEDIDDDRGT
 TGGIIGFTSGTGMLELVQNYANTKPDNNVLPFLPVLKRVNGTKSHEGLGQKYVDAWHQ
 AAKDSVFLKEQDKLRDSMYFNPAVSQKSDGLSNLQGFMYDAIFMHGPGDSSDSFGGIR
 KSAMKNAKTPAQGGDEKTYLQAFATARKKIMQENAHSDTSRVDDAQLKFLNEGNYDLHT
 PLKWKVYGDPEYK

>tr|E3FVA7|E3FVA7_STIAD Uncharacterized protein OS=Stigmatella aurantiaca (strain
 DW4/3-1) GN=STAUR_1941 PE=4 SV=1
 MTTTLFACGAGGDESPGNPEAGPQELESCAYVVAASTHMGSSEWGSITFKNTGTNDIQNP
 QISFNVPSGVTCGEGPSGWTRLQGGTTCQYTSSSHLLTIGVDTSYTFSYFTDSSTSFTATQ
 VQVSAVRCASTADEKEGLTATQKTLAEALTRIWENNTPSLDVTLQAASHGALPYEAMSE
 AKKWGLTTALSKAALYDAFIQHGEAGVRELLQORTHASLVSGQAAPAVGPQGLSEDAWLR
 GFLEQRDTLAADPEGRYAIDRVATYEQRRRGNWELLTAVQNDVRARDWCWNVAYPDSGF
 TVRKLNPDGSWSTPASYLYSCR

Figure S1. Cont.

>tr|A0AD68|A0AD68_STRAM Putative secreted chitosanase OS=Streptomyces ambofaciens ATCC 23877 GN=SAMR0713 PE=4 SV=1

MRHPSDPARPAQTRRRRTVLAMANASLAAVPLLTPPRAAATPARAAGLDDPAKKEIAMRL
VSSAENSSLDWKAQYGYIEDIRDGRGYTAGIIGFCSGTGDMLDLVELYTARSPGNVLARH
LPALREVDGSDSHEGLDPGYPGDWQRAAGDPEFRRAQDDERDRVYFGPAVRQAQADGLRT
LGQFAYYDAIVMHGGGGDRLGFGSIRDRALGRARPPAGGGDEVAYLHAFLDARVWAMKQE
EAHSDTSRVDTAQRVFLREGNLDLDPPLDWQVYGDSYHVG

>SAV_6191

MKRAALLLLTAVPAAVA AATAVYLFGLAPSDPDRTESPKPPSRQSSASAQGLNAPAKKEL
AQEIVASAENSTLDWRSAYGYVEDIGDGRGYTAGVIGFCTGTHDLLTLVERYTKDHPGNG
LARFLPALRAVDGTDSHAGLGRAFTAAWRAEAGVAAFRAAQDAERDRVYFDPAVRLARHD
GLGALGQFIYYDAMVVHGPPTDPHSLQGIRARALREADTPAEGGSETRYLDAFLDVRRTV
MKSEKSHHDTTRIDTAQRFLYDGNLRLRTPLEWRVYGERYRVP

>SAV_2015

MHSPHIRTSTRLLALIGASLVAGPLVANQSATAAPVGLDDPAKKEIAMKLVSSAENSSL
DWKAQYKYIEDIGDGRGYTAGIIGFCSGTGDMLDLVELYTQRKPGNVLATYLPALRNNG
GDShQGLDPGFPGDWRRAAQDSAFQQAQNDERDRVYFDPAVRQKADGIGVLGQFTTYDA
IVMHGDDGSDSTSFSSIRGRALAKAEPQAQGGNEVTYLNALFLDARVWAMRQEEAHSDTSRV
DTAQRVFLTKGNLNLDPPLDWKVYGDSYHIG

>tr|D7BYX5|D7BYX5_STRBB Chitosanase OS=Streptomyces bingchenggensis (strain BCW-1) GN=SBI_02563 PE=4 SV=1

MPKRIGLVAAAVAVPAALVLTSSVIGIQAAEHTPSSAQAVKAEHRTAAVKNLDDPAKKEIA
MKLVSSAENSSLDWRAQFKYIEDIDDGRGYTAGIIGFCSGTGDMLLEVEYTTQVKPGNVL
AKYLPALREVDGSDSHDGLDPNFTKDWKKAEDKAFQDAQEHERDRVYFNPSVVSQKADG
IGTLGQFIYYDAIVMHGPGDSKYSFGGIRKTAMSKAKTPAQGGDEKTYLNALFLDARKKAM
QSEEPHSETSRVDTEQRKFLNEGNLGLPEPLKWKVYGDPEIN

>tr|Q9S2Z1|Q9S2Z1_STRCO Putative chitosanase (Putative secreted protein) OS=Streptomyces coelicolor (strain ATCC BAA-471 / A3(2) / M145) GN=SCO2024 PE=4 SV=1

MKRAGVLLLGALPVIAAGVYFVAVPDDSDPADTAAASSSSSATARSRRDDAKDRAERERE
ADDALIADLPPGLAAPAKKELAQQLVSSAENSTTKWRTAYGSIEDVGDGDGYTAGIIGFC
TGTHDLLMLVERYTEDHPDNGLAEYLPALREVDGSDSHEGLDPGFATAAWKAEAEVPAFRA
AQEAERDRVYFEPAVRLAKLDGLGTLGQFVYDAMVFHGPDTDAEGFYGLRERAMAEART
PGQGSEKAYLETFLDVRKQAMEAKRPGIDTSRVDTAQRRLTAGNLKLATPLVWEMYGD
TYRVP

>tr|Q9RJ88|Q9RJ88_STRCO Secreted chitosanase OS=Streptomyces coelicolor (strain ATCC BAA-471 / A3(2) / M145) GN=SCO0677 PE=4 SV=1

MRHPFRDPARPAQPSRRRTVLAMACASLATVPLFTPSHAAAASARATGLDDPAKKEIAMQL
VSSAENSSLDWKAQYRYIEDIGDGRGYTAGIIGFCSGTGDMLDLVELYGERSPGNVLAPY
LPALRRVDGSDSHEGLDPGFDDWRRAADQDPQFRRAQDDERDRVYFDPAVRRGKEDGLR
TLGQFAYYDAMVMHGDGGGLGFGSIRERARALGRARPPAQGGDEVAYLHAFLLDERVWAMKQE
QAHSDTSRVDTAQRVFLNEGNLDLEPPLDWHVYGDAYHIG

Figure S1. Cont.

>BN159_0394

MRRNTMLWQAITATVLTGGIVVGVGLTPASAAAGSNLALGKPV TASSVEGAGFEPKAVD
 GSTSTRWASLEGVDNQWIQIDLGSVTDVAQVVLKWEAAAYAKSYRVEVSDDGSTWRQVYST
 TTGNATDDLAVSGSGRYLRVYGTQRATAYGYSLEYEVEVYGDSP TSPDPTTTPDSTTNLL
 AGRPTSVSSVEGTGFEFGKAVDGSTSTRWASVEGVDPQWLRVDMGSKTKTVGRVVLKWEAA
 YAKSYRVEVSDDGSTWRQVYSTTTGDGGTDDLAVSGSGRYLRVYGTQRATAYGYSLEYEVE
 AYGTGSTTTPPSDTANLDDPAKKEIAMKLVSSSFENSSLDWRAQFAYIEDIGDGRGYTAGI
 IGFCSGTGDMLDLIVERYTQKKPSNPLAPYLPALREVN GTDSHAGLGTSFENAWRTAAQDS
 VFTKTQEEERDRVYFNPAVGGQAKTDGLKALGQFAYYDAAVMHGEEGFRAIRSRAVSRVKP
 PSQGGNEVTFLHAFLDEREEMRKEEAHSDTTRVSTAQRRLNEGNLHLNIPLYWSVYGE
 SFSITS

>BN159_0876

MRHPTRRLLAATVASTAAIPLLNAGHAGAVRAPGLDDPAKKEIAMKLVSSAENSSLDWK
 AQYKYIEDIGDGRGYTAGIIGFCSGTGDMLDLVELYTD RSPGNVLAKYLPALRAVDGTDS
 HAGLDPDYPRDWRRAAQDRAFQQAQNDERDRVYFDPAVRQ GKADGLRALGQFAYYDAIVM
 HGGGSDRLSFGSIRRRALQSARPPAQGGDEVRYLHAF LDARVWAMKQEEAHSDTSRVDTA
 QRVFLRQGNL DLNPPLDWKVYGDSYHIG

>BN159_6425

MRRTGILAFALAPVVLTAGAYFLMPTDSAESPAAPPVSAAQLAREEAKERAEETERAEDEK
 TIASLPPGLADPAKKE LALKITATAEFSSLNWRLGYGSVEDNGDGC GYTAGAIGFC TGTH
 DLLDLIVERYTETHPDNGLARYLPALRKVDGTDSHEGLD PGFPAAWKAEAKKPAFRAAQDE
 ERDEVYFNPAVRLAKLDGLGTLGQFVYDAMVLHGPDPSANGFYGLRERALHEARLP SRG
 GSEKAYLDIFLDLREQAMKRKSTADTSRVSTAQREWLYDGNLSLNTPLSWRMYGEPYQV
 S

>tr|C9E0E7|C9E0E7_STRLI_CsnA_OS=Streptomyces lividans TK24 GN=csnA PE=4 SV=1

MRHPFRDPAFAGPSRRITVLAMACASLATVPLFTPSRAAAASARATGLDDPAKKEIAMQL
 VSSAENSSLDWKAQYRYIEDIGDGRGYTAGIIGFCSGTGDMLDLVELY GQRS PGNVLAPY
 LPALRRVDGSDSHEGLD PGFPDDWRRAADQDPQFRRAQDDERDRVYFDPAVRRGKEDGLR
 TLGQFAYYDAMVMHGDGGGLGFGSIRERALGRARPPAQGGDEVAYLHAF LDERVWAMKQE
 EAHSDTSRVDTAQRVFLNEGNL DLEPPLDWHVYGDAYHIG

>tr|Q7WT07|Q7WT07_9ACTO_Putative secreted chitosanase OS=Streptomyces sp. AM-7161
 GN=csn_StrAM-7161 PE=4 SV=1

MPPTINGCAAAA SRPSSPPRPPASCSPPLAGLTSATAAAVNLAANRPVTTSSSTESGSFGG
 AKAVAGSASTRWASAEVDNQWIRIDLGSATTLGRVVLKWEAAAYARAYRVELSDDGSTWR
 QIHATTTGDGGTDDL SVNGTGRYL RVFGTQRGTPYGYSLYEVEAYGAGATTPPPSSGTGL
 DDPKKEVAMKLVSSSFENSSLDWRAQFAYIEDIDDGRGYTAGIIGFCSGTGDMLDLIVERY
 TAKKPGNPLARYLGALRAVDGTDSHEGLD PGFTNAWRQAAQDTVFQQTQEEERDRVYFTP
 AVSQAKQDGLRALGQFAYYDAAVMHGEGSFRD IRSVALS RALPPSRGGDEKAYLKAFLDA
 REEMRKEEAHSDTTRVSTAQRRLNEGNL DLNTP LHWVYGEAFSLTG

>CsnN174

MHSQHRTARIALAVVLTAI PASLATAGVGYASTQASTAVKAGAGLDDPHKKEIAMELVSS
 AENSSLDWKAQYKYIEDIGDGRGYTGGIIGFCSGTGDMLDLVQHYTDLEPGN ILAKYLP
 LKKVNGSASHSGLGTPFTKDWATAAKD TVFQQAQNDERDRVYFDPAVSQAKADGLRALGQ
 FAYYDAIVMHGPGNDPTSFGGIRKTAMKKARTPAQGGDETTYLNAFLDARKAAMLTEAAH
 DDTSRVDTQQRVFLKAGNLDLNPPLKWKTYGDPYVINS

Figure S1. Cont.

>SACTE_5457

MLHPHNRTARRTTRLTRTGGLAAAAALGLALMALPVTAHAGAPTQPAHHLEAAATGLDDP
AKKDIAMQLVSSAENSTLDWKAQYGYIEDIGDGRGYTAGIIGFCSGTGDMMLALVERYTDR
SPGNVLASYLPALREVDGTDSDHGLDPGFPRDWAEAAKDPVFQQAQNDERDRVYFDPAVR
QAKDDGLGTLGQFAYYDAIVMHGGGGDSTSFSGSIRQRALAEAEPPSRGGDEVAYLDAFLD
ARVWAMRQEEAHS DTSRVDTAQRVFLRDGNLNLDPPLDWQVYVYGDSEFHIG

>SACTE_5458

MKTKLLALAAALALTGCSGPGPDAPEKGLADPAKKDIAMRLVSSAENSTLDWKAQYGYIE
DIGDGRGYTAGVIGFCSGTGDMMLKVVERYTAARPGNALERFLPALRAVVGDSHEGLGDA
FTEAWARVADDTAFFRAQDAERDETYFDPAVARAEADGLSALGQFVYDAFVMHGYADAK
GTVGFRTIRAEAVEAADPPSRGGDEEAYLDAFLDARVRAIRQEP SHSDASRVETAQRV FV
HEGRLQLETPLVWKVYGESFRIDGP

>SACTE_5886

MPVKPVTRLVLFGAPIATAASILFSRGDSEPPKLTPOAAAPAPSDQEEPEAVARAEHAE
EIDALPPGLAAPEMKEIASRLVSSADASTLDWRSRYGAVEDLGDGNGYTAGIVGFCSGTN
DMLQLVEAYTEEQPDNPLAPYLPALREVDGSDSHEGLDPGFTRAWREAAEDETFRSAQDR
LRDRLYFDPAVRLAKLDGLGTLGQFVYDAMVLHGPGVEANGFYGIRDAAMAQADTAAEG
GDETAYLDAFLDAGRTAIRARTVQRDTSRIDTAQRVFLRDGNLELRTPLVWRMYGEDFRIP

>tr|D2AW85|D2AW85_STRD Chitosanase OS=Streptosporangium roseum (strain ATCC 12428 / DSM 43021 / JCM 3005 / NI 9100) GN=Sros_2053 PE=4 SV=1

MPISSDPRPGRRALRLLLPVVAAPALVLFAGPGHAAVPAPHL SAPSQGSVTAAAADLT
DPRKKDIAMQLVSSAENSSLNWKAQYKYIEDIDDGRGYTAGIIGFCSGTGDMLELVELYA
ERKPGNVLAKYLPALRKVDGTD SHKGLDPNYTKDWATAAKDKVFQQAQNDERDRVYFNPA
VQQA KDKGLRALGQFAYYDAIVMHGDGGDPESFSSIRRNALKKAKPPAQGGDEVKYLHAF
LDARKVAMKAEAEHEDTSRVDTAQRVFLNKGNDLNLTPLSWKVYGDPTTIG

>tr|C4KDK1|C4KDK1_THASP Peptidoglycan-binding domain 1 protein OS=Thauera sp. (strain MZ1T) GN=Tmz1t_3574 PE=4 SV=1

MERIIPCTDCAEKTLDVEQLGFRVTS CDPHPERPGFCVLR FEDRSATPAAGASLAAPAAA
ARAGRGAVAGGVTATQAAVAKAIVNLFETGEVLGQYGVKVTLIPGDTGHLTFGRSQTTLGS
GNLAKLLQQYCANPGARFAGRLASYLPRFLAIDESLDDD PRLHNVL RATADDPVMRD TQD
AFFDRTYWEPALRAAASLGVHTPLGVAVVYDSAVHGSWLAMRDR TTRAVGEPAAVGEQAW
IDAYVRTRRAWLEGHARADLRQTVYRMEAFGRLLIDQGFWGLEMPLVVVRGREISSVTLAAL
PPGCYDGPQPSRPLTLATPLARGLDVRLQLGLSDRGVDILADGIFGR TSNLLKAWQA
QHGLAATGIADPALIGELTA

>tr|A7K8G4|A7K8G4_9PHYC Putative uncharacterized protein Z204R OS=Acanthocystis turfacea Chlorella virus 1 GN=Z204R PE=4 SV=1

MFAISMSITPENKQSIADYVCRMALAEKELNAVGDYKGLGNRRVQFKTALGNALIK
GDPVPTPDPEQSHDTKPAPGNVPTPGPSTGTTRIPSSIVPQLATLGFSETDADTILSLIS
LPENSNTTEWKNYNFASRLGDGRGWTVTLYGACSGTGDLMILKDLQKINPHHKLVKYIP
AMEKTEGEDVVRGLENLGRDIKSLGDDKEWQAVWDIYIKLYWNFARNFSDKLINRPGAKL
TSPLTRGFMDTALNHGADLDSFGPI LKGMNNDKDEATWFLDFCESRRKLLKRGFQDL
DTSKTGDRCTLWANI FKSNTSLTRPIKCYRKYWGNTIS

Figure S1. Cont.

```

>tr|Q84608|Q84608_PBCV1 Chitosanase OS=Paramecium bursaria Chlorella virus 1
GN=A292L PE=4 SV=1
MSNKIEITDDNKMTIQNDFVSRMMKSIDQELVAMTNKYSGFGPGRQTNCKKALAKALGET
PVNPPVNPVPTPPVDTHIPSQVEAPLKKLGFNTTNADTILSLIALPENSTTQWWKNYNYA
SCLKDGRGWTVTIYGACSGTGDLLMVLESLSQKINPNHPLVKFIPAMRKTKGDDIRGLENL
GKVINGLGDDKEWQTAVWDIYVKLYWTFAADFSDKTGSAKNRPGPVMTSPLTRGFMVDVA
LNHGSNMESEFSDILKRMKNREEKDEAKWFDFCETRRKLLKAGFQDLDTSKTGDRCTLWA
NIFKEGNVGLKRPIKCYNGYWGKNIVIS

>tr|O12288|O12288_9PHYC VChta-1 protein OS=Chlorella virus GN=vChta-1 PE=4 SV=1
MSNKIEITDDNKMTIQNDFVSRMMKSIDQELVAMTNKYSGFGPGRQTNCKKALAKALGET
PVNPPVNPVPTPPVDTHIPSQVEAPLKKLGFNTTNADTILSLIALPENSTTQWWKNYNYA
SCLKDGRGWTVTIYGACSGTGDLLMVLESLSQKINPNHPLVKFIPAMRKTKGDDIRGLENL
GKVINGLGDDKEWQTAVWDIYVKLYWTFAADFSDKTGSAKNRPGPVMTSPLTRGFMVDVA
LNHGSNMESEFSDILKRMKNKDEKDEAKWFDFCETRRKLLKSGFQDLDTSKTGDRAILWS
ELEFKTGNVGLKRPIKCYNGYWGKNIVIS

>tr|A7RBE5|A7RBE5_PBCVA Putative uncharacterized protein C342L OS=Paramecium
bursaria Chlorella virus AR158 GN=C342L PE=4 SV=1
MFMTSRIEITEDNKSKIQNDFVSKMEDIEQELDAMTKKYSGFGPGRQTNFKRALSRALE
ETLVKPPIVPPVDSHIPSHVEASLKQFGFSIENADMILSLISLPENSTIEWWKNYNYAER
LHDGRGWTVTLYGACSGTGDLLMILEALQKINSKHPLVKFIPAMRKTKGDDIRGLENLKG
VINVLGDDKEWQEAVWEIYIKLYWTFATDFADKKNSAKNRPGAVLTSPLTHGFMIDVALN
HGANMESEFEPILKNMKNREEKDEAKWFDFCEARRKLLKSGFQQLDTSKTGDRCTLWANI
FKSGNVGLKRPIQCFDGYWGKNFIIA

>KUTG_00281
MRARCLVAATMLAVGVAVGGCGSDVPNGPTATSKAPRPGGGLTADQRRRADQLISIFENG
TATVQYGYAEDLHDGRGVTAGRAGYTTNDGDALRVIQAYTDLAPGNGLARFAPALRQLAD
TGSGDLPEADYIAAWRQAADDPAFRRVQDDQVEQNYFTPAMAAADRLGLITPLARAELYD
AAIQHGHGADPDGLPALIDRATARAGKVDEAAWLTAFLDVRADDLRNPANADTKEEWSKS
TDRVECFRRIAATGDYTLAGPLTVTAYGVTYSLA

```

Figure S1: FASTA-formatted sequences of all the proteins presented on the phylogenetic tree (Figure 1).

xxxxxxx α-helix
 xxxxxxx 3/10 helix
 xxxxxxx β-strand

```

Npun_R2009          -----TEEEAENLQTGQ-----EICLPKIDEPPTQEVPGVEFFPPGTLNQLNLTGLD
RBRH_02856         -----LNSGD--GNTGQPWRVGTCTNAVA---NVASGNPDANFSPATLQFLKKNLTGLD
PPSC2_c3877       -----QETNE--SNIQQI-PLTSEETTSQT--AVTAIDHDANFSPSTLQFLKDNLTGLD
CV_3931            -----QGQAWQAGKRCRPAAQ--AK-AADHDANFPPATLKFLLKANTGLD
Q2PWA1_9BACL      -----SQQK--SYAAGNPDNSNFFPATLNPLRDNLTGLD
Q9XDS6_BURGA      -----TASGE--AGSGQPWVAGEACKPAKL--SKAALDHDANFSPATLQFLKDNLTGLD
CHIS_BACCI        1 -----ASPDDNFSPETLQFLRNNTGLD 22
O24825_9BACL      -----ASPDENFSPETLQFLRDRNTGLD
A7K8G4_9PHYC      LIKGDV-----VTPDPPEQSHDTKPAPGNV--PTPGPSTGTTRIPSSIV-PQLATLGF
A7RBE5_PBCVA      LEETLV-----K---P-----PIVPPVDSHIPSHVE-ASLKQVGF
Q84608_PBCV1      LGETPV-----NPPVNP-----PVTTPVDTHIPSQVE-APLKKLGF
O12288_9PHYC      LGETPV-----NPPVNP-----PVTTPVDTHIPSQVE-APLKKLGF
Hoch_2166         -AGGP-----ALGSAPMRLVDSDA--AS--ETTAAP-AADAAA---D-AAPAALLT
STAUR_1941        GAGGDES--PG-----NPEAG--P-----
A4KCU0_9ENTR      -----MILT
Play_0939         -----MLS
Tmz1t_3574        DPH-----PE--RPGFCVLRFEEDRSATPAAG--ASLAAPAAAARAG---R-GAVAGGVT
CJA_3747          -----MLT
Daro_2852         YAE-----PH--ILQFFPS-----LPS---T-RRSWKMLT
Cho4239-1_Jan     -----adglsa
SCO2024           DSADPADTAAA--SSSSSATARSDDAKDRA--E--REREADDALI---A-DLPPGLAA
BN159_6425        dsaespaapp-----vsaaqlareekera--e--teraedekti---a-slppglad
SACTE_5886        --FSRGDSEPPKLTPOAAAPAPSDQE-EPEAV--A--RAEEHAEI---D-ALPPGLAA
KSE_40860_csn2    EAYGTGGGSTASPSPTASASASA----SPTAS--P--SPSASSSTTP---P-TGGVNLDD
BN159_0394        eaygtgst-----tpp---p-sdtanltd
Csn_StrAM-7161    EAYGAGAT-----TPP---P-SSGTGLDD
SAV_6191          -----l-----aps----dpdrt--e--spk-ppsrsqs---s-asaqglna
E1AXU1_9PSED     1 -----ET-AGTVDLDA 10
Q8KZM5_9PSED     -----AGTVDLDA
SACTE_5458        -----APEKGLAD
Amir_4331         DAWSPYQ-----RWT----TPGGG--T--NPT-TTTTSQ---V-PGRSDLAT
Sros_2053         -----A---VPAPH--L--S-A-PSQGSV---T-AAAADLTD
SCO0677           -----AAA---S-ARATGLDD
C9E0E7_STRLI     -----A---S-ARATGLDD
SACTE_5457       40 -----ga-p--t-q-paahhl---e-aaatgld 59
SAV_2015         -----APVGLDD
SAMR0713         -----T---P-ARAAGLDD
BN159_0876       -----vrapgldd
RSa133209_3139   -----A-----AQA---NA-VS--S--LAP-AITAVS---A-ASTGDLSA
AORI_1736        -----SI---F-ASGGDLAS
SBI_02563        -----AEH---TP-SS--A---QA-VKAEHR---T-AAVKNLDD
AMED_6991        -----A--P--APL-AASSAS---V-LATGDLSA
CsnN106          -----A---GT-VH--A--AP--APAGAT---R-LAAVGLDD
CsnN174          1 -----AGAGLDD 7
Cho_AmyCSO-2     -----AA--P--AA--PAAA--V---S-VASVGLDD
KSE_15150_csn1   -----A---TP-AA--G--GT--SAVRAA---A-VAAVGLDD
KSE_40930_csn3   -----AAGGGLTT
KUTG_00281       -----SDVPNGP-TAT-SKAPRPGGGLTA
Q9RC18_BacKFB    -----AGLNK
BAMF_3066        -----AGLNK
Q9ET84_BACAM     -----AGLNK
RBAM_029740      -----AGLNK
AOEQW7_9BACI    -----AGLNK
DOFHB6_BACIU     -----AGLNK
BSNT_03901       -----AGLNK
BSU26890         -----AGLNK
QOPVM7_BACAM     -----AGLNK
BATR1942_06760   -----AGLNK
  
```

Figure S2. Cont.

Npun_R2009	AQQLTNILG-MINGPEQANSKWWQT-----VDEEIIYGYAE-DIED----	G
RBRH_02856	GEQWNNIMK-LINKPEQDSLQWT-----KYYGYCQ-NIHD----	D
PPSC2_c3877	GEQWDNIMK-LVnkPEQDSLKWT-----EFYGYAE-DIGD----	D
CV_3931	GEQWDNIMK-LVnkPEQDSLQWT-----KfyGYCE-NIGD----	K
Q2PWA1_9BACL	GEQWNNIMM-LVnkPEQDDLNI-----DFYGYCE-DIDD----	D
Q9XDS6_BURGA	GEQWDNIMK-LVnkPEQDSLQWT-----KfyGYCE-DIGD----	D
CHIS_BACCI	23 GEQWNNIMK-LINKPEQDDLNI-----KYYGYCE-DIED----	E 56
O24825_9BACL	GEQWNNIMK-LINKPEQDDLNI-----KYYGYCE-DIND----	E
A7K8G4_9PHYC	SETDADTILSLISLPENSNTTEWWK-----NYNFAS-RLGD----	G
A7RBE5_PBCVA	SIENADMILSLISLPENSTIEWWK-----NYNYAE-RLHD----	G
Q84608_PBCV1	NTTNADTILSLIALPENSTTQWWK-----NYNYAS-CLKD----	G
O12288_9PHYC	NTTNADTILSLIALPENSTTQWWK-----NYNYAS-CLKD----	G
Hoch_2166	PEQVQSAIA-----FYGADRWPAKIAQVSEKLGIAVREQVDEQFVQAAAGFQQ----	Q
STAU_R1941	-QELESCAYVVAASTHMGSWSITFKNTGT-----NDIQNPQISFNVP-----	
A4KCU0_9ENTR	PAQKSLCER-VLNAFETGSAE--G-----DYSAIA-IFHDGPHGI	
Plav_0939	ELQKEAKA-IVNIFETGAVR--G-----DYARVT-LLAG--DS	
Tmz1t_3574	ATQAAVAKA-IVNLFETGEVL--G-----QYGVKVT-LIPG--DT	
CJA_3747	PTQKKTAEA-IVNIFETGSVL--G-----DYSNVT-LITG--DT	
Daro_2852	PTQKQTAQS-IVNLFETGAVL--G-----HYGNVT-VIPG--DT	
Cho4239-1_Jan	pakkeiamq-lvssenssldwka-----qyayie-diam--g	
SCO2024	PAKKELAQQ-LVSSAENSTTKWRT-----AYGSIE-DVGD--g	
BN159_6425	pakkelalk-itataefsslwrl-----gygsve-dngd--g	
SACTE_5886	PEMKEIASR-LVSSADASTLDWRS-----RYGAVE-DLGD--G	
KSE_40860_csn2	PAKKEIAMK-LVSSAENSSLDWRA-----QFSYIE-DIKD--G	
BN159_0394	pakkeiamk-lvssfenssldwra-----qfayie-digd--g	
Csn_StrAM-7161	PRKKEVAMK-LVSSFENSSLDWRA-----QFAYIE-DIDD--G	
SAV_6191	pakkelaqe-ivasaenstldwrs-----aygyve-digd--g	
E1AXU1_9PSED	11 PVQKDTAMS-LVSSFENSSTDWQA-----QYGYLE-DIAD--G 44	
Q8KZM5_9PSED	PVQKDTAMS-LVSSFENSSTDWQA-----QYGYLE-DIAD--G	
SACTE_5458	PAKKDIAmr-LVSSAENSTLDWKA-----QYGYIE-DIGD--G	
Amir_4331	PAKKDIAmr-LVSSAENSSLDWRA-----QFAYIE-DIGD--G	
Sros_2053	PRKKDIAMQ-LVSSAENSSLDWKA-----QYKYIE-DIDD--G	
SCO0677	PAKKEIAMQ-LVSSAENSSLDWKA-----QYRYIE-DIGD--G	
C9E0E7_STRLI	PAKKEIAMQ-LVSSAENSSLDWKA-----QYRYIE-DIGD--G	
SACTE_5457	60 pakkdiamq-lvssaenstldwka-----qygyie-digd--g 93	
SAV_2015	PAKKEIAMK-LVSSAENSSLDWKA-----QYKYIE-DIGD--G	
SAMR0713	PAKKEIAMR-LVSSAENSSLDWKA-----QYGYIE-DIRD--G	
BN159_0876	pakkeiamk-lvssaenstldwka-----qykyie-digd--g	
RSa133209_3139	PAKKEIAMQ-LVCSAENSSLDWKA-----QYGYIE-DIDD--D	
AORI_1736	PEKKEIAMQ-IVSSAENSSLDWKA-----QYSYIE-DIGD--G	
SBI_02563	PAKKEIAMK-LVSSAENSSLDWRA-----QFKYIE-DIDD--G	
AMED_6991	PAKKEIAMK-LVSSAENSSLDWKA-----QYKYIE-DIGD--G	
CsnN106	PHKKDIAMQ-LVSSAENSSLDWKS-----QYKYIE-DIKD--G	
CsnN174	8 PHKKEIAME-LVSSAENSSLDWKA-----QYKYIE-DIGD--G 41	
Cho_AmyCSO-2	PAKKEIAME-LVSSAENSSLDWKA-----QYKYIE-DIGD--G	
KSE_15150_csn1	PAKKEIAMK-LVSSAENSSLDWKA-----QYKYIE-DIGD--G	
KSE_40930_csn3	-DQRRRADQ-LISVFENGTTV--I-----QYGYAE-NIND--G	
KUTG_00281	-DQRRRADQ-LISIFENGTAT--V-----QYGYAE-DLHD--G	
Q9RC18_BacKFB	-DQKRRAEQ-LRRICEDGTTE--M-----RYPYVA-RLDD--A	
BAMF_3066	-DQKRRAEQ-LTSIFENKTE--I-----QYGYVE-ELDD--G	
Q9ET84_BACAM	-DQKRRAEQ-LTSIFENKTE--I-----QYGYVE-ALDD--G	
RBAM_029740	-DQKRRAEQ-LTSIFENKTE--I-----QYGYVE-ALDD--G	
A0EQW7_9BACI	-DQKRRAEQ-LTSIFENKTE--I-----QYGYVE-ALDD--G	
D0FHB6_BACIU	-DQKRRAEQ-LTSIFENGTTE--I-----QYGYVE-RLDD--G	
BSNT_03901	-DQKRRAEQ-LTSIFENGTTE--I-----QYGYVE-RLDD--G	
BSU26890	-DQKRRAEQ-LTSIFENGTTE--I-----QYGYVE-RLDD--G	
Q0PVM7_BACAM	-DQKRRAEQ-LTSIFENGMTE--I-----QYGYVE-HLPD--G	
BATR1942_06760	-DQKRRAEQ-LTSIFENGMTE--I-----QYGYVE-HLPD--G	

Figure S2. Cont.

Npun_R2009	RGVTIG--IYG-----	-----ATTGKGYN---DADVIWKNYQD	
RBRH_02856	RGYTIG--IFG-----	-----ATTGGSNDTGPDPPELKFKFDAL	
PPSC2_c3877	RGYTIG--IFG-----	-----ATTGGSNDTGPDPGPALEKAFDAA	
CV_3931	RGYTIG--IFG-----	-----ATTGGPNDEGPDGPTLFKEFDAS	
Q2PWA1_9BACL	RGYTIV--LFG-----	-----ATTGGSNDTGPDPDPLFKAYDAA	
Q9XDS6_BURGA	RGYTMG--IFG-----	-----ATTGGPNDDGGPDGPALEKAYDAA	
CHIS_BACCI	57 RGYTIG--LFG-----	-----ATTGGSRDTHPDGPDLEKAYDAA	88
O24825_9BACL	RGYSIG--IFG-----	-----ATTGGPRDTHPDGPELKFAYDAA	
A7K8G4_9PHYC	RGWTVT--LYG-----	-----ACSGTG----DLVMILKDLQK-	
A7RBE5_PBCVA	RGWTVT--LYG-----	-----ACSGTG----DLLMILEALQK-	
Q84608_PBCV1	RGWTVT--IYG-----	-----ACSGTG----DLLMVLESLOK-	
O12288_9PHYC	RGWTVT--IYG-----	-----ACSGTG----DLLMVLESLOK-	
Hoch_2166	RSLTVD--GMLGSASIMPLFEGEEMDRAHTMANRITAEYESSGNYGVVQNAVVGIIISYGAH		
STAU_R_1941	SGVTCGEGPSGWT---RL-----	-----QGGTTCQYTSSSH----LTIGVDTSYTF	
A4KCU0_9ENTR	RQITYG--RSQ-----	-----TTEYS-----KLQDLVTRYVN-	
Plav_0939	GHLTYG--RAQ-----	-----TTLGSG-----NLHLLIQYCE-	
Tmz1t_3574	GHLTFG--RSQ-----	-----TTLGSG-----NLAKLLQYCA-	
CJA_3747	GHLTFG--RSQ-----	-----TTLGSG-----NLGKLLQYCA-	
Daro_2852	GHLTFG--RSQ-----	-----TTLCSG-----NLLDLLQRYCG-	
Cho4239-1_Jan	-gytgg--iig-----	-----ftsqts-----dmlslvtaydq-	
SCO2024	DGYTAG--IIG-----	-----FCTGTH----DLLMLVERYTE-	
BN159_6425	cgyttag--aig-----	-----fctgth----dlldlveryte-	
SACTE_5886	NGYTAG--IVG-----	-----FCSGTN----DMLQLVEAYTE-	
KSE_40860_csn2	RGYTAG--IIG-----	-----FCSGTG----DMLDLVEAYTA-	
BN159_0394	rgyttag--iig-----	-----fcsqtg----dmlslverytq-	
Csn_StrAM-7161	RGYTAG--IIG-----	-----FCSGTG----DMLDLVERYTA-	
SAV_6191	rgyttag--vig-----	-----fctgth----dlldlverytk-	
E1AXU1_9PSED	45 RGYTGG--LIG-----	-----FTSGTG----DMLDLVRAHSA	70
Q8KZM5_9PSED	RGYTGG--LIG-----	-----FTSGTG----DMLDLVRAHSA-	
SACTE_5458	RGYTAG--VIG-----	-----FCSGTG----DMLKVVRYTA-	
Amir_4331	RGYTAG--IIG-----	-----FCSGTG----DMLDLVEAYTS-	
Sros_2053	RGYTAG--IIG-----	-----FCSGTG----DMLDLVELYAE-	
SCO0677	RGYTAG--IIG-----	-----FCSGTG----DMLDLVELYGE-	
C9E0E7_STRLI	RGYTAG--IIG-----	-----FCSGTG----DMLDLVELYGO-	
SACTE_5457	94 rgyttag--iig-----	-----fcsqtg----dmlslverytd-	119
SAV_2015	RGYTAG--IIG-----	-----FCSGTG----DMLDLVELYTD-	
SAMR0713	RGYTAG--IIG-----	-----FCSGTG----DMLDLVELYTA-	
BN159_0876	rgyttag--iig-----	-----fcsqtg----dmlslverytd-	
RSa133209_3139	RGYTGG--IIG-----	-----FTSGTG----DMLDLVQNYAN-	
AORI_1736	RGYTAG--IIG-----	-----FCSGTG----DMLDLVERYTN-	
SBI_02563	RGYTAG--IIG-----	-----FCSGTG----DMLDLVEYYTQ-	
AMED_6991	RGYTAG--IIG-----	-----FCSGTG----DMLDLVEAYTN-	
CsnN106	RGYTAG--IIG-----	-----FCSGTG----DMLDLVADYTD-	
CsnN174	42 RGYTGG--IIG-----	-----FCSGTG----DMLDLVQHYTD	67
Cho_AmyCSO-2	RGYTAG--IIG-----	-----FCSGTG----DMLDLVQHYTD-	
KSE_15150_csn1	RGYTGG--IIG-----	-----FCSGTG----DMLDLVQHYTD-	
KSE_40930_csn3	RGVTAG--RAG-----	-----FTTNDG----DALKVVRAYTE-	
KUTG_00281	RGVTAG--RAG-----	-----YTTNDG----DALRVIQAYTD-	
Q9RC18_BacKFB	RPSTCG--PAG-----	-----VTTATR----DGFEVVPVYKQ-	
BAMF_3066	RGYTCCG--RAG-----	-----FTTATG----DALEVVEVYTK-	
Q9ET84_BACAM	RGYTCCG--RAG-----	-----FTTATG----DALEVVEVYTK-	
RBAM_029740	RGYTCCG--RAG-----	-----FTTATG----DALEVVEVYTK-	
A0EQW7_9BACI	RGYTCCG--RAG-----	-----FTTATG----DALEVVEVYTK-	
D0FHB6_BACIU	RGYTCCG--RAG-----	-----FTTATG----DALEVVEVYTK-	
BSNT_03901	RGYTCCG--RAG-----	-----FTTATG----DALEVVEVYTK-	
BSU26890	RGYTCCG--RAG-----	-----FTTATG----DALEVVEVYTK-	
Q0PVM7_BACAM	RGYTCCG--RAG-----	-----FTTATG----DALEVVEVYTK-	
BATR1942_06760	RGYTCCG--RAG-----	-----FTTATG----DALEVVEVYTK-	

Figure S2. Cont.

Npun_R2009 IEKVH-AIANDQKWWKAQWDAYISTYWQPTLKLKSK-----NYMKALTIGVLIDT
RBRH_02856 CSKVK-AQQNPTWREAMWRTFYNDIQYSVQQARQR-----GFDSALTIGSFVDT
PPSC2_c3877 KKKIK-ALQNNPRAWREAWRTFYDITYEYSVQQARKR-----GFNTALTIGSFVDT
CV_3931 CDKIG-SLQNNPAWRDAMWNTFYKVYIQYSVQQARQR-----GFSSALTIGSFVDT
Q2PWA1_9BACL CRKIG-NLQDDPEWREAMWRTFYNYIYKYRGEQARNR-----GFNSALTIGSFVDT
Q9XDS6_BURGA CGKVK-GLQNDAAWREAMWRTFYSVYIQYSVQQARSR-----GFGSALTIGSFVDT
CHIS_BACCI 124 CGKIK-KLQNDAAWRKAMWETFYNYIYRYSVEQARQR-----GFTSALTIGSFVDT 173
O24825_9BACL CGKIK-KLQNDPAWRKAMWETFYNYIYRYSVEQARQR-----GFTSALTIGSFVDT
A7K8G4_9PHYC GRDI-KSLGDDKEWQAVVDIYIKLYWNFARNFSDKLD---INRPGAKLTSPLTRGFMVDT
A7RBE5_PBCVA GKVI-NVLGDDKEWQEAWEIYIKLYWTFATDFADKKNSAKNRPGAVLTSPLTRGFMIDV
Q84608_PBCV1 GKVI-NGLGDDKEWQTAVWDIYVKLYWTFADDFSDKTGSAKNRPGPVMTSPLTRGFMVDV
O12288_9PHYC GKVI-NGLGDDKEWQTAVWDIYVKLYWTFADDFSDKTGSAKNRPGPVMTSPLTRGFMVDV
Hoch_2166 LSALR-AAGSEQIMQDAQNAFFSEDFWVPAVKAALNH-----GITSALGYATLYDA
STAUR_1941 TRIWENNT--PSLDVTLQAASHGALPYEAAMSEAKKW-----GLTTALSKAALYDA
A4KCU0_9ENTR KDLLRRAGRDDPIMKSVQDRFFDERYYLPALKWAQDF-----GFKDALSLLVYDYS
Play_0939 RNLL-KEAGADPVMRDTQDAFFDRVYWPAAASAIRL-----GLTEALSIAIVYDS
Tmz1t_3574 HNVL-RATADDPVMRDTQDAFFDRTYWEPALRAAASL-----GVHTPLGVAVVYDS
CJA_3747 KNIL-RATADDHIMRETQDLFFDQAYWQPAERAATQL-----GIKTPLGVAVVYDS
Daro_2852 HNLL-RATADDPVMRETQDLFFDEVYWQPAAKAADNF-----GIKSALGVAVVYDS
Cho4239-1_Jan vsdwqkaa-qdpvfqqsqntlrdnayfnpavslakqd-----glr-tlgqfiyyda
SCO2024 TAAWKAEA-EVPAFRAAQEAERDRVYFPAVRLAKLD-----GLG-TLGQFVYYDA
BN159_6425 paawkaea-kkpafraaqdeerdevyfnpavrlakld-----glg-tlgqfvyyyda
SACTE_5886 TRAWREAA-EDETFRSAQDRLRDLRYFDPAVRLAKLD-----GLG-TLGQFVYYDA
KSE_40860_csn2 PAAWKQAA-TDPVFTQTQESERDRVYFNPAVAQAKKD-----GLR-ALGQFAYYDA
BN159_0394 enawrtaa-qdsvftktqeeerdrvyfnpavqgakt-----glk-algqfayyyda
Csn_StrAM-7161 TNAWRQAA-QDVFQQTQEEERDRVYFPAVSQAKQD-----GLR-ALGQFAYYDA
SAV_6191 taawraea-gvaafraaqdaerdrvyfdpavrlarhd-----glg-algqfiyyda
E1AXU1_9PSED 101 EQAWADAA-ETSEFRAAQDAERDRVYFDPAVAQGGKAD-----GLS-ALGQFAYYDT 149
Q8KZM5_9PSED EQAWADAA-ETSEFRAAQDAERDRVYFDPAVAQGGKAD-----GLS-ALGQFAYYDT
SACTE_5458 TEAWARVA-DDTAFRAAQDAERDETYFDPAVARAEAD-----GLS-ALGQFVYYDA
Amir_4331 QQAWRTAA-QDPVFQQAQETERDRVYFTPSVNQAKSD-----GLR-ALGQFAYYDA
Sros_2053 TKDWATAA-KDKVFQQAQNDERDRVYFNPAVQAKKD-----GLR-ALGQFAYYDA
SCO0677 PDDWRRADQDPQFRRAQDDERDRVYFDPAVRRGKED-----GLR-TLGQFAYYDA
C9E0E7_STRLI PDDWRRADQDPQFRRAQDDERDRVYFDPAVRRGKED-----GLR-TLGQFAYYDA
SACTE_5457 150 prdwaeaa-kdpvfqqaqnderdrvyfdpavrqqakd-----glg-tlgqfayyyda 198
SAV_2015 PGDWRRAA-QDSAFQQAQNDERDRVYFDPAVRQGGKAD-----GLS-VLQGFYYDA
SAMR0713 PGDWQRAA-GDPEFRAAQDDERDRVYFPAVRQAQAD-----GLR-TLGQFAYYDA
BN159_0876 prdwrraa-qdrafqqaqnderdrvyfdpavrqqkad-----glr-algqfayyyda
RSa133209_3139 VDAWHQAA-KDSVFLKEQDKLRDSMYFNPAVSQGGKSD-----GLS-NLQGFMYDA
AORI_1736 ENAWREAA-GREDFKAAQNSERDRVYFDPAVNQAKSD-----GLG-ALGQFIYYDA
SBI_02563 TKDWKAAA-EDKAFQDAQEHERDRVYFNPSVSQGGKAD-----GIG-TLGQFIYYDA
AMED_6991 ESAWKQAA-ATTAFTAQNNERDRVYFNPSVNQGGKAD-----GLS-NLQGFAYYDA
CsnN106 EKDWATAA-KDSVFQQAQNDERDRVYFNPAVNQAKA-----SLR-ALGQFAYYDA
CsnN174 98 TKDWATAA-KD TVFQQAQNDERDRVYFDPAVSQAKAD-----GLR-ALGQFAYYDA 146
Cho_AmyCSO-2 VNDWRATAA-KD TVFQQAQNDERDRVYFNPAVKQAKAE-----RLR-ALGQFVYYDA
KSE_15150_csn1 VNDWKATAA-KD TVFQQAQNDERDRVYFNPAVKQAKAD-----GLR-ALGQFAYYDA
KSE_40930_csn3 VTAWKRAA-DDPAFRRVQDAQVDERYFAPAMADADRL-----GLTTALARAEFLDA
KUTG_00281 IAAWRQAA-DDPAFRRVQDDQVEQNYFTPAMAAADRL-----GLITPLARAEYDA
Q9RC18_BacKFB ASAWKSLA-DDKAFRAAQDGVNDQVYYPAMERSDNA-----GLTTALARAVMYHT
BAMF_3066 ASAWRSLG-NDKAFRAAQDEVNDRLYYQAMKRSDQA-----GLKTALAKAVMYDT
Q9ET84_BACAM ASAWRSLG-NDKAFRAAQDKVNDSLYYPAMKRSENA-----GLKTALAKAVMYDT
RBAM_029740 ASAWRSLG-NDKAFRAAQDKVNDSLYYPAMKRSENA-----GLKTALAKAVMYDT
A0EQW7_9BACI ASAWRSLG-NDKAFRAAQDKVNDSLYYPAMKRSENA-----GLKTALAKAVMYDT
D0FHB6_BACIU ASAWKSLA-NDKEFRAAQDKVNDHLYYQAMKRSDNA-----GLKTALARAVMYDT
BSNT_03901 ASAWKSLA-NDKEFRAAQDKVNDHLYYQAMKRSDNA-----GLKTALARAVMYDT
BSU26890 ASAWKSLA-NDKEFRAAQDKVNDHLYYQAMKRSDNA-----GLKTALARAVMYDT
Q0PVM7_BACAM DSAWRSLG-KDKDFRAAQDVTNDRLYYQAMKQSDNI-----GLKTALAKAVMYDT
BATR1942_06760 DSAWRSLG-NDKDFRAAQDAVNDRLYYQAMKRSENA-----GLKTALARAVMYDT

Figure S2. Cont.

Npun_R2009	AMNAGMEDDN----	SENWGV EHLFTEA-----	SDDTDNEEDFVDRFMELRLQFP		
RBRH_02856	ALNQGADGGG----	DSLKAL----LAR-----	SGDSPNEKTFMTRFYAERAKVV		
PPSC2_c3877	ALNQGATGDD----	GSLEGV----LSR-----	SGNSTNEKTFMTNIFYAKRTL		
CV_3931	ALNQGATGDS----	GTLQGL----LSR-----	SGNSGDEKTFMTAFYAQRSKIV		
Q2PWA1_9BACL	ALNHGATGGS----	ESLQGL----LSK-----	SGSSTDEKTFMTKFKYKERTKIV		
Q9XDS6_BURGA	ALNQGADGGS----	NTLQGL----LSR-----	SGNSTDEKTFMTSFYAQRTKVV		
CHIS_BACCI	174	ALNQGATGGS----	DTLQGL----LAR-----	SGSSSNEKTFMKNFHAKRTL	214
024825_9BACL	ALNQGATGDS----	NTLQGL----LAR-----	SGSSTNEKTFMKNFHAKRTL		
A7K8G4_9PHYC	ALNHGADL-----	DSFGPILKGMNKN-----	DEQDEATWFLDFCESRRKLL		
A7RBE5_PBCVA	ALNHGANM-----	ESFEPILKNMKNR-----	EKDEAKWFLDFCEARRKLL		
Q84608_PBCV1	ALNHGSNM-----	ESFSDILKRMKNR-----	EKDEAKWFLDFCETRRKLL		
O12288_9PHYC	ALNHGSNM-----	ESFSDILKRMKNK-----	DEKDEAKWFLDFCETRRKLL		
Hoch_2166	KIQGGMEDSLQRATSAMGGIVGATVER		NGRSQQVTEAEFLVAFNEAREGR		
STAU_1941	FIQHGEAGV-----	RELLQTHASLGVSQAAPAVGPQGLSEDAWLRGLFLEQRTDL			
A4KCU0_9ENTR	FVH-----	SGTILTLFLRAQFR-----	E-VPPASGGDEHTWISQYVEVRNSWL		
Plav_0939	VVH-----	GSWISLRDRTL-A-----	K-AGQPSKAGERAWSLAYVRERRNWL		
Tmz1t_3574	AVH-----	GSWLAMRDRTT-R-----	A-VGEPAAVGEQAWIDAYVTRRAWL		
CJA_3747	TVH-----	GSAKLIRDKTN-Q-----	S-AGTLASLGEQKWIEAYVATRRYWL		
Daro_2852	WVH-----	GSWKTLLRDMTN-Q-----	K-SGPIATLGERAWIKAYVETRRSWL		
Cho4239-1_Jan	ivmhgpgtdsqr--	nsfqgnspqalkk-----	a-lppskggnetaaylnafldvrralm		
SCO2024	MVFGHPDTPD--	A-EGFYGLRERAMAE-----	A-RTPGQGGSEKAYLETFLDVRKQAM		
BN159_6425	mvlhgpdps--	a-ngfyglreralhe-----	a-rlpsrggsekayldifldlreqam		
SACTE_5886	MVLHGPVE--	A-NGFYGIRDAAMAQ-----	A-DTAAEGGDEATAYLDAFLDAGR		
KSE_40860_csn2	AVMHGP-----	DGLASIRSAALAK-----	A-KPPAQGGNETTWLNAFLDAREAM		
BN159_0394	avmhge-----	egfrairsravsr-----	v-kppsqqgnevtfllhafldereem		
Csn_StrAM-7161	AVMHGE-----	SGFRDIRSVALSR-----	A-LPPSRGGDEKAYLKAFLDAREEM		
SAV_6191	mvvhgpgtd--	p-hslqgiraralre-----	a-dtpaeggsetryldafldvrrtvm		
E1AXU1_9PSED	150	LVVHGPQSQ--	R-DAFGGIRAEALSA-----	A-LPPSQGGDETEYLEAFFDARNVIM	197
Q8KZM5_9PSED	LVVHGPQSQ--	R-DAFGGIRAEALSA-----	A-LPPSQGGDETEYLEAFFDARNVIM		
SACTE_5458	FVMHGYADAKGT--	VGFRGIRAEAVEA-----	A-DPPSRGGDEEAYLDAFLDARVRAI		
Amir_4331	IVMHGGGTD--	Q-ASFGTIRRKAMAR-----	A-KTPAQGGDEATAYLHAFLDERVVAM		
Sros_2053	IVMHGDGGD--	P-ESFSSIRRNALKK-----	A-KPPAQGGDEVKYLHAFLDARKVAM		
SCO0677	MVMHGDGG--	G-LGFGSIRERALGR-----	A-RPPAQGGDEVAYLHAFLDERVVAM		
C9E0E7_STRLI	MVMHGDGG--	G-LGFGSIRERALGR-----	A-RPPAQGGDEVAYLHAFLDERVVAM		
SACTE_5457	199	ivmhggggd--	s-tsfgsirgralae-----	a-eppsrggdevayldafldarvwam	246
SAV_2015	IVMHGDGGD--	S-TSFGSIRGRALAK-----	A-EPPAQGGNEVTYLNAFLDARVVAM		
SAMR0713	IVMHGGGGD--	R-LGFGSIRERALGR-----	A-RPPAQGGDEVAYLDAFLDARVVAM		
BN159_0876	ivmhgggsd--	r-lsfgsirrralqs-----	a-rppaqggdevrylhafldarvwam		
RSa133209_3139	IFMHGPGDS--	S-DSFGGIRKSAMKN-----	A-KTPAQGGDEKTYLQAFATARKKIM		
AORI_1736	IVMHGPGTS--	Q-DSFGGIRSAALQQ-----	A-KPPAQGGDEATAYLKAFLDARRVVM		
SBI_02563	IVMHGPGDS--	K-YSGFGIRKTAMSK-----	A-KTPAQGGDEKTYLNAFLDARKKAM		
AMED_6991	IVMHGPGDS--	S-DSFGGIRKAAMKK-----	A-KTPAQGGDEATYKAFFAARKVIM		
CsnN106	IVMHGPGDS--	S-DSFGGIRKAAMKK-----	A-KTPAQGRDEATYKAFLAARKTVM		
CsnN174	147	IVMHGPGND--	P-TSFGGIRKTAMKK-----	A-RTPAQGGDETTYLNAFLDARKAAM	194
Cho_AmyCSO-2	IVMHGPGSS--	S-DSFGGIRAAAMKK-----	A-KTPAQGGDEATYLNFLDARKVIM		
KSE_15150_csn1	IVMHGPGTS--	A-DSFGGIRAAALKK-----	A-KTPAQGGDEATYKAFLDARKVVM		
KSE_40930_csn3	SVQHNGNSE--	Y-DALPALIARTSAK-----	AGTPAAA--GEDAWLDAFFDVRIDDL		
KUTG_00281	AIQHGHDGAD--	P-DGLPALIDRATAR-----	AGK---V--DEAAWLTAFLDVRADDL		
Q9RC18_BacKFB	VRQRGDGDD--	G-DSRYALIKRTPKG-----	AGGSPKEGIDEQKCLNKFSHVRYDDL		
BAMF_3066	VIQHDGDD--	P-DSFYALIKRTNKK-----	MGGSPKDGIDEKKWLNKFLDVRYYDDL		
Q9ET84_BACAM	VIQHDGDD--	P-DSFYALIKRTNKK-----	MGGSPKDGIDEKKWLNKFLDVRYYDDL		
RBAM_029740	VIQHDGDD--	P-DSFYALIKRTNKK-----	MGGSPKDGIDEKKWLNKFLDVRYYDDL		
AOEQW7_9BACI	VIQHDGDD--	P-DSFYALIKRTNKK-----	MGGSPKDGIDEKKWLNKFLDVRYYDDL		
DOFHB6_BACIU	VIQHDGDD--	P-DSFYALIKRTNKK-----	AGGSPKDGIDEKKWLNKFLDVRYYDDL		
BSNT_03901	VIQHDGDD--	P-DSFYALIKRTNKK-----	AGGSPKDGIDEKKWLNKFLDVRYYDDL		
BSU26890	VIQHDGDD--	P-DSFYALIKRTNKK-----	AGGSPKDGIDEKKWLNKFLDVRYYDDL		
QOPVM7_BACAM	IIQHGGGDD--	P-DSLNSLIKRTNKK-----	AGGSPKNGVDEKKWLNKFLDVRYYDDL		
BATR1942_06760	VIQHDGDD--	P-DSFYSLIKRTNKK-----	AGGSPKDGIDEKKWLNKFLDVRYYDDL		

* :

Figure S2. Cont.

Npun_R2009	TR----	DS-----	GDMEERIG-AWQKLLRDRKWD	MRVDLKNYVYIPQ-----	
RBRH_02856	DTHDYNTS-----	PNGKNRVK-QWSKLMDOGETDL	KNADAAIVKVTNWQLH----		
PPSC2_c3877	DTNDYNQP-----	PNGKNRVK-QWSSLLASGETDL	KNADVAIKVTNWEMK----		
CV_3931	DTNDYNQP-----	PNGKNRVK-QWSTLLNMGETDL	KNADAAVQKVTNWEMK----		
Q2PWA1_9BACL	DTNEYNSP-----	PNGKNRVK-QWSNLLNMGETDL	KNADSAVLQVTDWELQ----		
Q9XDS6_BURGA	215 DTHDFNQP-----	PNGKNRVK-QWSTLMSQGITSL	KNCDADIVKVTSWTMK----		259
CHIS_BACCI	DTNEYNQP-----	PNGKNRVK-QWDTLVDMGKM	NLKNVDSEIAQVTDWEMK----		
O24825_9BACL	KRGFQDLDTSKT-----	GDRC-TLWANI FKSGNTSLTR	-PIKC-----YRGYWGN		
A7K8G4_9PHYC	KSGFQQDLDTSKT-----	GDRC-TLWANI FKSGNVGLKR	-PIQC-----FDGYWGK		
A7RBE5_PBCVA	KAGFQDLDTSKT-----	GDRC-TLWANI FKEGNVGLKR	-PIKC-----YNGYWGK		
Q84608_PBCV1	KSGFQDLDTSKT-----	GDRA-ILWSELFKTGNVGLKR	-PIKC-----YNGYWGK		
O12288_9PHYC	ERIAVRRDGQKRRDAEMLRNS	KVRPQ-AFEELARDGNLDLSA	-NVDGENSLEFRTYIGGR		
Hoch_2166	AADPEG-----	RYAIDRVA-TYEKQRRRGNW	LLT-AVQ-----NDVRARD		
STAUR_1941	ANHSN-----	PELHPTVYRTR-CLRFEDRGN	WDLSQVPIIN-----ANGI---		
A4KCU0_9ENTR	AMHSN-----	TLLRKTIRME-AFEALIAAKN	WSLAL-PMT-----VRGI---		
Plav_0939	EGHAR-----	ADLRQTVYRME-AFGRLIDQ	GFWGLEM-PLV-----VRGR---		
Tmz1t_3574	ANHPR-----	KDLRPTVYRMD-AFQRLIDL	NLWGLEL-PLV-----VRGL---		
CJA_3747	ANHSR-----	SDLRGTTYRMD-AFLRLIEQ	GYWGLAL-PLV-----VHGE---		
Daro_2852	Cho4239-1_Jan	qsepdh-----	ag-----nldrldtqqrflq	ngnldlnt-pln-----wkv	
SCO2024	EAKRPG-----	ID-----TSRVDTAQR	RFLTAGNLKLAT-PLV-----	WEMY---	
BN159_6425	krkkst-----	ad-----tsrvstaqrewly	dgnlslnt-pls-----wrm		
SACTE_5886	RARTVQ-----	RD-----TSRIDTAQR	VFLRDGNLELRT-PLV-----	WRMY---	
KSE_40860_csn2	RKEAAH-----	SD-----TRRVSTEQR	KFLNEGNFDLHT-PLT-----	WSVY---	
BN159_0394	rkeeah-----	sd-----ttrvstaqrrfl	negnlhlni-ply-----w	svy---	
Csn_StrAM-7161	RKEAAH-----	SD-----TRRVSTAQR	RFLNEGNLDLNT-PLH-----	WSVY---	
SAY_6191	kseksh-----	hd-----ttridtaqrrfl	lydgnldlrlt-ple-----	wrvy---	
E1AXU1_9PSED	218 REEPAH-----	AD-----TSRIDTAQR	VFLQNGNFDLER-PLT-----	WSVY---	233
Q8KZM5_9PSED	REEPAH-----	AD-----TSRIDTAQR	VFLQNGNFDLER-PLT-----	WSVY---	
SACTE_5458	RQEPSH-----	SD-----ASRVETAQR	VFVHEGRLQLET-PLV-----	WKVY---	
Amir_4331	KLEEAH-----	ED-----TSRVDTAQR	VFLNNGNLDLNT-PLA-----	WKVY---	
Sros_2053	KAEEAH-----	ED-----TSRVDTAQR	VFLNKGNDLNT-PLS-----	WKVY---	
SCO0677	KQEQAH-----	SD-----TSRVDTAQR	VFLNEGNLDLEP-PLD-----	WHVY---	
C9E0E7_STRLI	KQEEAH-----	SD-----TSRVDTAQR	VFLNEGNLDLEP-PLD-----	WHVY---	
SACTE_5457	247 rkeeah-----	sd-----tsrvdtaqr	vflrdgnlndp-pld-----	wqvy---	282
SAV_2015	RQEEAH-----	SD-----TSRVDTAQR	VFLTKGNLNDP-PLD-----	WKVY---	
SAMR0713	KQEEAH-----	SD-----TSRVDTAQR	VFLREGNLDLDP-PLD-----	WQVY---	
BN159_0876	kqeeah-----	sd-----tsrvdtaqr	vflrqgnlndp-pld-----	wkvy---	
RSa133209_3139	KQENAH-----	SD-----TSRVDDAQL	KFLNEGNLDLHT-PLK-----	WKVY---	
AORI_1736	KQEEAH-----	AD-----TSRIDTAQL	VWLNDGNLDLHT-PLK-----	WAVY---	
SBI_02563	QSEEPH-----	SE-----TSRVDTAQR	VFLNEGNLLEP-PLK-----	WKVY---	
AMED_6991	KQEEAH-----	AD-----TSRVDTAQR	VFLSEGNDLHT-PLK-----	WKVY---	
CsnN106	LKEEAH-----	SD-----TSRVDTAQR	VFLNAKNFDLNP-PLK-----	WKVY---	
CsnN174	219 LTEAAH-----	DD-----TSRVDTAQR	VFLKGNLDLNP-PLK-----	WKTY---	230
Cho_AmyCSO-2	KQEEAH-----	AD-----TSRVDTAQR	VFLNAKNFDLNP-PLK-----	WKVY---	
KSE_15150_csn1	KQEEAH-----	AD-----TSRVDTAQR	VFLNAKNFDLNP-PLK-----	WKVY---	
KSE_40930_csn3	THPANS-----	ATQAEWSQSVDRVE-ALRRI	ARTGNRNLGDG-PFT-----	VTAF---	
KUTG_00281	RNPANA-----	DTKEEWSKSTDRVE-CFRRI	AATGDYTLAG-PLT-----	VTAY---	
Q9RC18_BacKFB	MNGANH-----	DRRDEWRESVGRVH-VLRSI	ANQNNYNLNG-GIH-----	VRSH---	
BAMF_3066	MNPSDE-----	DTQDEWRESVARVD-VFRD	IVKAKNYNLDG-PIH-----	VRSS---	
Q9ET84_BACAM	MNPADE-----	DTQDEWRESVARVD-VFRD	IVKEKYNLNG-PIH-----	VRSS---	
RBAM_029740	MNPSDE-----	DTQDEWRESVARVD-VFRD	IVKEKYNLNG-PIH-----	VRSS---	
A0EQW7_9BACI	MNPSDE-----	DTQDEWRESVARVD-VFRD	IVKEKYNLNG-PIH-----	VRSS---	
D0FHB6_BACIU	MNPANH-----	DTRDEWRESVARVD-VLRSI	AKENNYNNG-PIH-----	VRSN---	
BSNT_03901	MNPANH-----	DTRDEWRESVARVD-VLRSI	AKENNYNNG-PIH-----	VRSN---	
BSU26890	MNPANH-----	DTRDEWRESVARVD-VLRSI	AKENNYNNG-PIH-----	VRSN---	
Q0PVM7_BACAM	MNPADP-----	DTRDEWRESVARVD-VLRSI	AKANNYNLNG-PIN-----	VYSE---	
BATR1942_06760	MNPSDP-----	DTRDEWRESVARVD-VLRSI	AKANNYNLNG-PIN-----	VYSE---	

Figure S2. Cont.

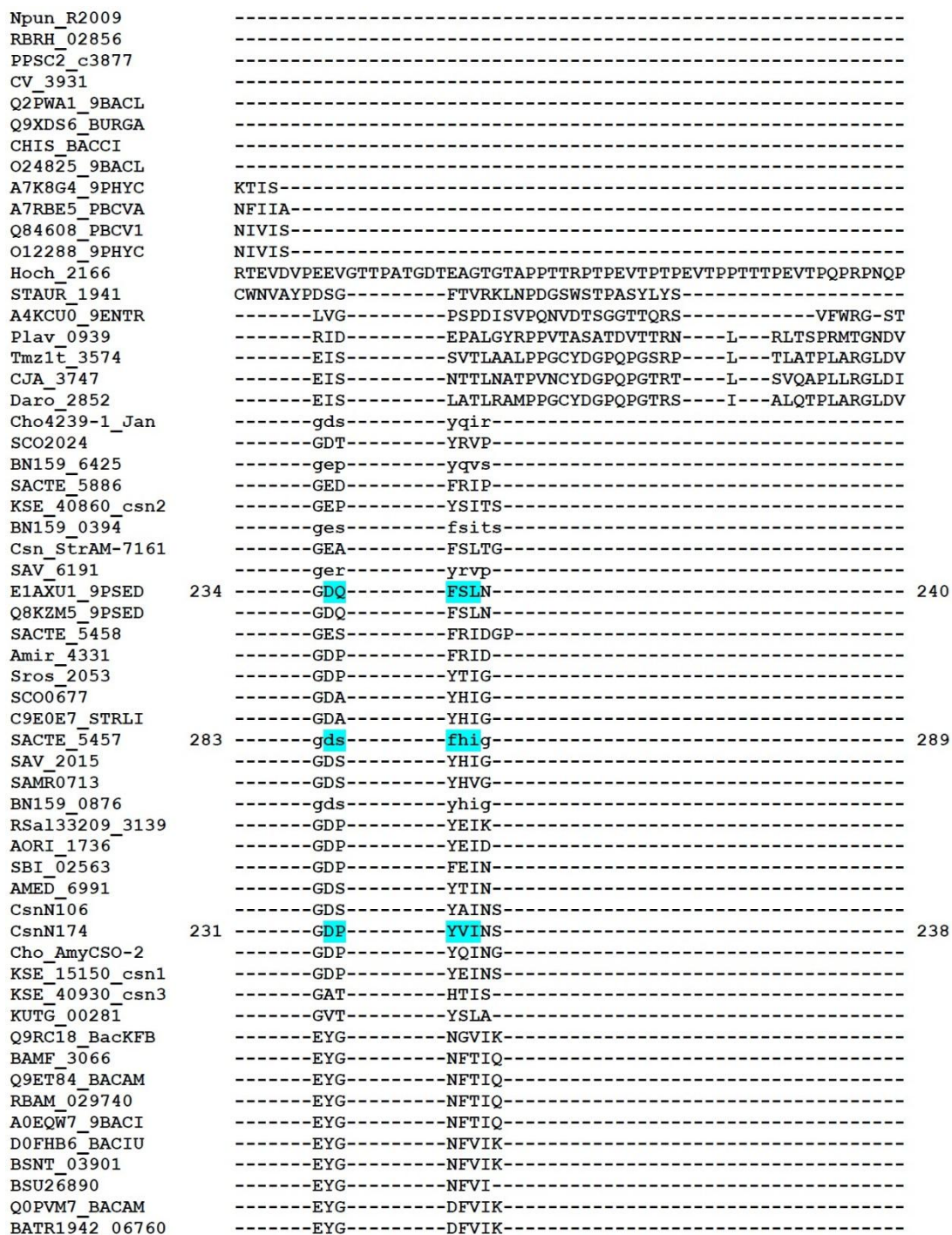


Figure S2: Sequence alignment obtained with Clustal Omega program after submission of FASTA-formatted sequences from which the predicted *N*-terminal signal peptides have been deleted. Residue numbering is shown for the four proteins for which 3D-structures are available. Color-coded secondary structures, derived from the respective Protein Data Bank files are also shown for these four proteins. (▼): catalytic residues.

>Amir_4331

DGFTGFPVVLGGKCLDVAGASSANGTAVQLYTCNGTSAQSWTVDTGAKTIKALGKCLDVDDGRHDGALLQIWDCANVPEQR
 WTVDQTRVVGQSGKCVDPAGNTADGTRLQVWTCDAWSPYQRWTTGGGTNPTTTTTSQVPGRS
 DLATPAKDIAMRLVS
 SAENSSLDWRAQFAYIEDIGDGRGYTAGIIGFCSGTGDMLELVEHYTSVKPNNALAGYLPALRRVNGTPSHEGLDPGFQQAW
 RTAAQDPVFQQAQETERDRVYFTPSVNQAKSDGLRALGQFAYYDAIVMHGGGTDQASFGTIRRKAMARAKTPAQGGDE
 TAYL
 HAFDERVVMKLEEAHEDTSRVDTAQRVFLNNGNLDLNTPLAWKVYGD PFRID

Ricin_B_lectin domain;ricin-type beta-trefoil; present in chitinases and several other types of GH

> AMED_6991

APAPLAASSASVLTGDLAPAKKEIAMKLVSSAENSSLDWKAQYKYIEDIGDGRGYTAGIIGFCSGTGDMLELVEAYTNSV
 PNNPLAKYLPALRKVNGTASHSGLGSASFESAWKQAAATTAFOQAQNNERDRVYFNPSVNQKADGLSNLQGFAYYDAIVMHG
 PGDSSDSFGGIRKAAMKAKTPAQGGDEATYKKAFFAARKVIMKQEEAHADTSRVDTEQAKFLSEGNDLHTPLKWKVYGD
 S
 YTIN

> AORI_1736

SIFASGGDLASPEKKEIAMQIVSSAENSSLDWKAQYSYIEDIGDGRGYTAGIIGFCSGTGDMLELVERYTNSVDPNPLAEYL
 PALREVDGTESHEGLTGFENAWREAAGREDFKAAQNSERDRVYFDPVAVQAKSDGLGALGQFIYYDAIVMHGPGTSQDSFG
 GIRSAALQQAQPPAQGGDEATYKKAFLDARRVVMKQEEAHADTSRIDTAQLVWLNLDGNLDLHTPLKWAVYGD
 P
 YEID

>Cho_AmyCSO-2

AAPAAPAAAASVSVASVGLDDPAKKEIAMELVSSAENSSLDWKAQYKYIEDIGDGRGYTAGIIGFCSGTGDMLELVQHYTDLK
 PGNVLAKYLPALKKVNVTDSHSLGSAFVNDWRATAKDTVVFQRAQNDERDRVYFNPAVKQAKAERLRALGQFVYYDAIVMHG
 PGSSSDSFGGIRAAAAMKAKTPAQGGDEATYLNFLDARKVIMKQEEAHADTSRVDTEQRVFLNAKNFDLNPPLKWKVYGD
 P
 YQING

> BAME_3066

AGLNKDQKRRAEQLTSIFENGKTEIQYGYVEELDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRLLAKD
 ESDDISNLKGFASAWRSLGNDKAFRAAQDEVNDRLYYQPAMKRSQAGLKTALAKAVMYDTVIQHGDDPDSFYALIKRTN
 KKMGGSPKDGTDKWKLNKFLDVRYDDLMPDSDTQDEWRESVARVDVFRDIVKAKNYNLGPIHVRSSSEYGNFTIQ

>Q0PVM7_BACAM

AGLNKDQKRRAEQLTSIFENGMTEIQYGYVEHLDPGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRLLAKE
 ESDDISNLKGFDSAWRSLGKDKDFRAAQDVTNDRLYYQPAMKQSDNIGLKTALAKAVMYDTIIQHGDDPDSLNSLIKRTN
 KKAGGSPKNGVDEKKWLNKFLDVRYDDLMPADPDTRDEWRESVARVDVLRSAKANNYNLNGPINVYSEYGFVIK

> RBAM_029740

AGLNKDQKRRAEQLTSIFENGKTEIQYGYVEALDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRLLAKD
 ESDDISNLKGFASAWRSLGNDKAFRAAQDKVNDSLYYQPAMKRSNENAGLKTALAKAVMYDTVIQHGDDPDSFYALIKRTN
 KKMGGSPKDGTDKWKLNKFLDVRYDDLMPDSDTQDEWRESVARVDVFRDIVKEKNYNLGPIHVRSSSEYGNFTIQ

>Q9ET84_BACAM

AGLNKDQKRRAEQLTSIFENGKTEIQYGYVEALDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKLLKYLPELRRLLAKD
 ESDDISNLKGFASAWRSLGNDKAFRAAQDKVNDSLYYQPAMKRSNENAGLKTALAKAVMYDTVIQHGDDPDSFYALIKRTN
 KKMGGSPKDGTDKWKLNKFLDVRYDDLMPADED TQDEWRESVARVDVFRDIVKEKNYNLGPIHVRSSSEYGNFTIQ

Figure S3. Cont.

> BATR1942_06760

AGLNKDQKRRAEQLTSIFENGMTEIQYGYVEHLPDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKPKYLPELRRLAKE
ESDDISNLKGFSAWRS LGNDKDFRAAQDAVNDRLYYQPAMKRSENAGLKTALAKAVMYD'TVIQHGDGDDPDSFYSLIKRTN
KKAGGSPKDGIDEKKWLNKFLDVRYDDLMPSPDPTREWRESVARVDVLRSAKANNYNLNGPINVYSEEYGFVIK

>CHIS_BACCI

ASPDDNFSPETLQFLRNNTGLDGEQWNNIMKLINKPEQDDLNIWIKYYGYCEDIEDERGYTIGLFGATTGGSRDTHPDGPDFL
KAYDAAKGASNP SADGALKRLGINGKMGSI LEIKDSEKVF CGKIKKLQNDAAWRKAMWETFYNVYIRYSVEQARQGF TSA
VTIGSFVDTALNQGATGGSDTLQGLLARS GSSSNEKT'FMKNFHAKRTL'LVDTNKYNKPPNGKNRVKQWDTLVDMGKMN LKNV
DSEIAQVTDWEMK

>A0EQW7_9BACI

AGLNKDQKRRAEQLTSIFENKTEIQYGYVEALDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKPKYLPELRRLAKD
ESDDISNLKGFASAWRS LGNDKAFRAAQDKVND'SLYYQPAMKRSENAGLKTALAKAVMYD'TVIQHGDGDDPDSFYALIKRTN
KKMGGSPKDGTD EKKWLNKFLDVRYDDLMP SDED'TQDEWRESVARVDVFRDIVKEKNYNLNGPIHVR SSEYGNFTIQ

>Q9RC18_BacKFB

AGLNKDQKRRAEQLRRICEDGTTEMRYPYVARLDDARPSTCGPAGVTTATRDGFEVVPVYKQAVANKKLPNYLAGLRLEKE
ASDDTSKLGFPASAWKSLADDKAFRAAQDGVNDQVYYQPAMERSDNAGLTTALARAVMYHTVQRQGDGDDGDSRYALIKRTP
KGAGGSPKEGIDEQKCLNKFSHVRYDDL MNGANHDRRDEWRESVGRVHVLRSIANQNNYNLNGGIHVR SHEYGNVVIK

>D0FHB6_BACIU

AGLNKDQKRRAEQLTSIFENGTTEIQYGYVERLDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKPKYLPELRRLAKE
ESDDTSNLKGFASAWKSLANDKEFRAAQDKVNDHLYYQPAMKRSDNAGLKTALARAVMYD'TVIQHGDGDDPDSFYALIKRTN
KKAGGSPKDGIDEKKWLNKFLDVRYDDLMPANHDTREWRESVARVDVLRSAKENNYNLNGPIHVR SNEYGNFVIK

>BSNT_03901

AGLNKDQKRRAEQLTSIFENGTTEIQYGYVERLDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKPKYLPELRRLAKE
ESDDTSNLKGFASAWKSLANDKEFRAAQDKVNDHLYYQPAMKRSDNAGLKTALARAVMYD'TVIQHGDGDDPDSFYALIKRTN
KKAGGSPKDGIDEKKWLNKFLDVRYDDLMPANHDTREWRESVARVDVLRSAKENNYNLNGPIHVR SNEYGNFVIK

>BSU26890

AGLNKDQKRRAEQLTSIFENGTTEIQYGYVERLDDGRGYTCGRAGFTTATGDALEVVEVYTKAVPNNKPKYLPELRRLAKE
ESDDTSNLKGFASAWKSLANDKEFRAAQDKVNDHLYYQPAMKRSDNAGLKTALARAVMYD'TVIQHGDGDDPDSFYALIKRTN
KKAGGSPKDGIDEKKWLNKFLDVRYDDLMPANHDTREWRESVARVDVLRSAKENNYNLNGPIHVR SNEYGNFVI

>Q9XDS6_BURGA

MATLPLAAADASARTVHAARAPQAAAAAASPEICSSPWTAARIYEAGNVVSFDGHDYTAAYRSQGNAPATASGEAGSGQPWV
AGEACKPAKLSKAALDHDANFSPATLQFLKDNTGLDGEQWDNIMKLVNKPEQDSDLWTKFYGYCEDIGDDRGYTMGIFGATT
GGPNDGGPDGPALFKAYDAASGASNP SVQGGLARIGAHGSMQGSILKITDSEKVF CGKVKGLQNDAAWREAMWRTFY SVYIQ
YSVQQARSRGFGSALTIGSFVDTALNQGADGGSNTLQGLLSRSGNSTDEKT'FMTSFYAQR'TKVVD'THDFNQPPNGKNRVKQW
STLMSQGITS LKNCDADIVKVT SWTMK

ChtBD3 superfamily domain; found in chitinases; discovered in ChiC of *Streptomyces griseus*

Figure S3. Cont.

>RBRH_02856

DDAQPNSEPCDAPWDAKTIYQGGEVVS YAGRNYTAA YWTQGNAPDLNSGDGNTGQPWRVGT CNAVANVASGNPDANFSPAT
LQFLKKN TGLDGEQWNNIMKLINKPEQDSLQWTKYGYCQNIHDDRGTITIGIFGATTGGSNDTGPDPGPELFKTFDALSGAKI
PSAEGGLTRIGAHGKMSG SILKISDSSKTFCSKVKAQQNNPTWREAMWRFTFYNVDIQYSVQQARQGRGFSALTIGSFVDTAL
NQGADGGGDSLKALLARSGDSPNEKTFMTRFYAERAKVVDTHDYNTSPNGKNRVKQWSKLMDOGETDLKNADAAIVKVTNWQ
LH

ChtBD3 superfamily domain; found in chitinases; discovered in ChiC of *Streptomyces griseus*

>CJA_3747

MLTPTQKKTAEAI VNI FETG SVLGDYSNVTLITGDTGHLTFGRSQTTLGSGNLGKLLQLYCANPGARFRQQLTPFLARFAAR
DFSLDHEEHLKNI LRATAD DHIMRETQDLFFDQAYWQPAERAATQLGIKTP LGVAVVYDSTVHGS AKLIRDKTNQSAGTLAS
LGEQKWIEAYVATRRYWL ANHPRKDLRPTVYRMDAFQRLIDLNLWGLELPLVVRGLEISNTTLNATPVNICYDGPQPGTRTLS
VQAPLLRGLDIRLVQLGLSKSGINLRADGIFGNGSVSAVKTFQTKQNL PATGIVDNALIAKLVS

C-terminal extension: PG_binding_1 domain; Putative peptidoglycan binding domain. This domain is composed of three alpha helices. This domain is found at the N or C terminus of a variety of enzymes involved in bacterial cell wall degradation.

>CV_3931

QGSTAGSPERQTVRLPDCAAWS PASPYQAGSVASRGGVNYTAAFWTQGNPPEQGOAWQAGKRCRPAQAQAKAADHDANFPPA
TLKFLKANTGLDGEQWNIKLVNKPEQDSLEWTKFYGYCENIGDKRGYITIGIFGATTGGPNDEGPDGPTLFKEFDASSGAS
NPSIAGGLARAGVHGSMQ GKILKISDSAKVFCDKIGSLQNNPAWRDAMWNTFYKVYIQYSVQQARQGRFSSALTIGSFVDTA
LNQGATGDSGTLQGLLSRSGNSGDEKTFMTAFYAQRSKI VDTNDYNQPPNGKNRVKQWSTLLNMGETDLKNADAAVQKVTNW
EMK

ChtBD3 superfamily domain; found in chitinases; discovered in ChiC of *Streptomyces griseus*

>A4KCU0_9ENTR

MILTPAQKSLCERVLNAFETGSAEGDYSAIAIFHDGPHGIRQITYGRSQTTEYSKLDLVTRYVNANGRF SADLAPFVPKIG
NIALVDNEPFKDLLRRAGRDDPIMKSVQDRFFDERYYLPALKWAQDFGFKDALSLLVYDYSFVHSGTILTF LRAQFREVP
SGGDEHTWISQYVEVRNSWLANHSNPELHP TVYR TKCLRFEIDRGNWDL SQVPINANGI LVGSPDISVPQNVDTSGGTTQR
SVFWRGSTRLLWFCG

C-terminal extension of unknown function

>Daro_2852

MSACYAEPHILQFFPSLPSTRRSWKMLTPTQKQTAQSIVNLFETGAVLGHYGNVTVIPGDTGHLTFGRSQTTLCSGNLLDLL
QRYCGNDGARFGDKLSAWLARFEAVDLSLDDDLHLHNL LRATADDPVMRETQDLFFDEVYWQPAKAADNFGIKSALGVAVV
YDSVWHGSKWTLRDMTNQKSGPIATLGERAWIKAYVETRRSWLANHSRSDLRGTTYRMDAFLRLIEQGYWGLALPLVHGE
ISLATLRAMP PG CYDGPQPGTRSI ALQTPLARGLDVRLQLGMSGRGVAIKADGIFGQTSVNL LKRFQASQGLPANGIATPQ
QLSELVD

N-terminal extension of unknown function

C-terminal extension: PG_binding_1 domain; Putative peptidoglycan binding domain. This domain is composed of three alpha helices. This domain is found at the N or C terminus of a variety of enzymes involved in bacterial cell wall degradation.

Figure S3. Cont.

>CsnN106

AGTVHAAPAPAGATRLAAVGLDDPHKDIAMQLVSSAENSSLDWKSQYKYIEDIKDGRGYTAGIIGFCSGTGDMLDLVADYT
DLKPGNILAKYLPALRKVNGTESHAGLASAFEKDWATAAKDSVFQQAQNDERDRSYFNPAVNQAKASLRALGQFAYYDAIVM
HGPGDSSDSFGGIRKAAMKKAKTPAQGRDEATYKLAFLAARKTVMLKEEAHSDTSRVDTEQTVFLNAKNFDLNPPLKWKVYG
DSYAINS

>Npun_R2009

MSCKIGSSYTIKPNDTLFEIAARELGDGRWREIMNPNGIPTTEEEAENLQTGQEICLPKIDEPPTQEVPGVEFFPPGTLNQ
LNTLTGLDAQQLTNILGMINGPEQANSKWWQTVDEEIIYGYAEDIEDGRGVTIGIYGATTGKGYNDADVIWKNYGQDYSNLP
VDEIIEKVHAIANDQKWWKAQWDAYISTYWQPTLKLKSKNYMKALTIGVLIDTAMNAGMEDDENSENWGVEHLFTEASDDTD
NEEDFVDRFMEELRLQFPTRDSGDMEERIGAWQKLLRDRKWDMRVDLKNYVYIPQ

LysM superfamily domain; "Lysine motif" - small motif involved in peptidoglycan binding

>O24825_9BACL

ASPDENFSPETLQFLRDRGTGLDGEQWNNIMKLINKPEQDDLNIWKYGYCEDINDERGYSIGIFGATTGGPRDTHPD
GPELFKAYDAAKAGNPSVEGALKRLGINGKMKGSILEIKDSEKVFCEGKIKKLQNDPAWRKAMWETFYNVYIRYSVEQARQR
GFTSALTIGSFVDTALNQGATGDSNTLQGLLARSGSSTNEKTFLLKFKHAKRTLVDVTNEYNQPPNGKNRVKQWDTLDDMGKM
NLKNVDAEIAQVTNWEMK

>PPSC2_c3877

LAEENGTVPQETNESNIQQIPLTSEETTSQTAVTAIDHDANFSPSTLQFLKDNTEGLDGEQWNNIMKLVNKPEQDSLKWTEFY
GYAEDIGDDRGTIGIFGATTGGSNDTGPDPALFKAFDAASGASNPSIAGGLARAGLKGKMSGKILKLSDSDSVIKKKIKI
LQNEAWREAMWRFTFYDYIEYSVQQARKRGFNTALTIGSFVDTALNQGATGDDGSLEGVLSRSGNSTNEKTFMTNFYAKRT
LVVDTNDYNQPPNGKNRVKQWSSLLASGETDLKNADVAIKVTNWEMK

N-terminal domain; shared with some chemotaxis proteins and chitinases

>Q2PWA1_9BACL

SQQKSYAAGNPDNSFFPATLNPLRDNTEGLDGEQWNNIMMLVKNKPEQDDLNIWIDFYGYCEDIDDDRGTIVLFGATTGGSNDT
GPDGPDLEFKAYDAAKAGSNPSVKGALARIQVKGSMKKGKILEINESEESFCRIGNLQDDPEWREAMWKTFYNIYIKYRGEQA
RNRGFNSALTIGSFVDTALNHGATGGSESLQGLLSKSGSSTDEKTFMTKFKYKERTKIVDTNEYNQPPNGKNRVKQWSNLLNM
GETDLKNADSAVLQVTDWELQ

>Plav_0939

MLSELQKEAAKAIVNI FETGAVRGDYARVTLLAGDSGHLTYGRAQTTLGSGNLHLLIKAYCEAPGAAYARACEPYLPRLADI
DLRLDSDWTFRNLLKEAGADPVMRDTQDAFFDRVYWTAAAASAIRLGLTEALS LAIVYDSVVHGSWISLRDRTLAKAGQPSK
AGERAWSLAYVRERRNWLAMHSNTLLRKT IYRMEAFEALIAAKNWSLALPMTVRGI RIDEPALGYRPPVTASATDVTTRNLR
LTSRPMTGNDVRALEAALVKEGYAINCDGVFDEGLEKALKSFQQDYGLIADGVAGPATRIMLG

C-terminal extension: PG_binding_1 domain; Putative peptidoglycan binding domain. This domain is composed of three alpha helices. This domain is found at the N or C terminus of a variety of enzymes involved in bacterial cell wall degradation.

>E1AXU1_9PSED

ETAGTVDDLDAVPQKDTAMSLVSSFENSSTDWQAQYGYLEDIADGRGYTGGLIGFTSGTGDMLELVRAHSASSPGNPLEQYIP
ALEAVNGTGSAGLGGQFEQAWADAAETSEFRAAQDAERDRVYFDPAVAQKADGLSALGQFAYYDTLVVHG
PGSQRDAFGGIRAEALSAAALPPSQGGDETEYLEAFFDARNVIMREEPAHADTSRIDTAQRVFLQNGNFDLERPLTWSVYGDQ
FSLN

Figure S3. Cont.

>Q8KZM5_9PSED

AGTVDLDAVPQKDTAMSLVSSFENSSTDWQAQYGYLEDIADGRGYTGGLIGFTSGTGMLELVRAYSASSPGNPLEQYIPAL
EAVNGTDSHAGLGQGFQAWADAAETSEFRAAQDAERDRVYFDPAVAQKADGLSALGQFVYYDTLVVHGPGSQRDAFGGIR
AEALSAALPPSQGGDETEYLEAFFDARNVIMREEPAHADTSRIDTAQRVFLQNGNFDLERPLTWSVYGDQFSLN

>RSa133209_3139

GTAAQANAVSSSLAPAITAVSAASTGDLAPAKKEIAMQLVCSAENSSLDWKAQYGYIEDIDDDRGYTGGIIGFTSGTGMLE
LVQNYANTKPDNNVLPFLPVLRLKVNKTSHEGLGQKYVDQAWHQAADSVFLKEQDKLRDSMYFNPAVSQKSDGLSNLQGF
MYYDAIFMHGPGDSSDSFGGIRKSAMKNAKTPAQGGDEKTYLQAFATARKKIMKQENAHSDTSRVDDAQLKFLNEGNYDLHT
PLKWKVYGDPEYK

>STaur_1941

MTTTLFACGAGDESPGNPEAGPQELESCAYVVAASTHMGSSEWGSITFKNTGTNDIQNPQISFNVPSTGEGPSGWTRL
QGGTTCQYTSSSHLLTIGVDTSYTFSYFTDSSSTFTATQVQVSAVRCASTADEKEGLTATQKTLAEALTRIENNTPSLDVTL
QAASHGALPYEAMSEAKKWLTTALSKAALYDAFIQHGEAGVRELLQRTHASLVSGQAAPAVGPQGLSEDAWLRGFLEQR
RDTLAADPEGRYAIDRVATYEQRRRGNWELLTAVQNDVRRDCWNVAYPDSGFTVRKLNPDGSWSTPASYLYSCR
(No conserved domains detected; similarity with other GH46 due essentially to a
segment homologous to the C-terminal part of other chitosanases)

>SAMR0713

TPARAAGLDDPAKKEIAMRLVSSAENSSLDWKAQYGYIEDIRDGRGYTAGIIGFCSGTGMMLDLVELYTARSPGNVLARHLP
ALREVDGSDSHEGLDPGYPGDWQRAAGDPEFRAAQDDEDRVYFGPAVRQAQADGLRTLQGFAYYDAIVMHGGGGDRLGFGS
IRDALGRARPPAGGGDEVAYLHAFLDARVWAMKQEEAHSSTSRVDTAQRVFLREGNLDLDPPLDWQVYGD SYHVG

>SAV_6191

AATAVYLEFGLAPSDPDRTESPKPPSRQSSASAAQGLNAPAKKELAQEIVASAENSTLDWRSAYGYVEDIGDGRGYTAGVIGFC
TGTHDLLTLVERYTKDHPGNGLARFLPALRAVDGTD SHAGLGRAFTAAWRAEAGVAAFRAAQDAERDRVYFDPAVRLARHDG
LGALGQFIYYDAMVVHGPPTDPHSLQGIRARALREADT PAEGGSETRYLDAFLDVRRTVMKSEKSHHDTTRIDTAQRFLYD
GNLDRTPLEWRVYGERYRVP

>SAV_2015

APVGLDDPAKKEIAMKLVSSAENSSLDWKAQYKYIEDIGDGRGYTAGIIGFCSGTGMMLDLVELYTQRKPGNVLATYLPALR
NVNGGDSHQGLDPGFPGDWRRAAQDSAFQQAQNDERDRVYFDPAVRQKADGIGVLGQFTYYDAIVMHGDGGDSTSFSSIRG
RALAKAEPPAQGGNEVTYLNALFLDARVWAMRQEEAHSSTSRVDTAQRVFLTKGNLNLDPPLDWKVYGD SYHIG

>SBI_02563

AEHTPSSAQAVKAEHRATAAVKNLDDPAKKEIAMKLVSSAENSSLDWRAQFKYIEDIDDGRGYTAGIIGFCSGTGMMLLVEY
YTQVKPGNVLAKYLPALREVDGSDSHDGLDPNFTKDWKKAEDKAFQDAQEHERDRVYFNPSVSQKADGIGTLGQFIYYDA
IVMHGPGDSKYSFGGIRKTAMSKAKTPAQGGDEKTYLNALFLDARKKAMQSEEPHSETSRVDTEQRKFLNEGNGLEPPLKWK
VYGDPEIN

>SCO2024

YFAVPDDSDADPADTAAASSSSSATARSRRDDAKDRAEREREADDALIADLPPGLAAPAKKELAQQLVSSAENSTTKWRTAYG
SIEDVGDGDGYTAGIIGFCTGTHDLLMLVERYTEDHPDNGLAEYLPALREVDGSDSHEGLDPGFATAAWKAEAEVPAFRAAQE
AERDRVYFEPAVRLAKLDGLGLGQFVYYDAMVFHGPDTDAEGFYGLRERAMAEARTPGQGGSEKAYLETFLDVRKQAMEAK
RPGIDTSRVDTAQRRELTAGNLKLATPLVWEMYGDYRVP
N-terminal extension of unknown function

Figure S3. Cont.

>SCO0677

AAASARATGLDDPAKKEIAMQLVSSAENSSLDWKAQYRYIEDIGDGRGYTAGIIGFCSGTGDMMLDLVELYGERSPGNVLAPY
LPALRRVDGSDSHEGLDPGFDDWRRADQDPQFRRAQDDERDRVYFDPVRRGKEDGLRRTLQGFAYYDAMVMHGDGGGLGF
GSIRERALGRARPPAQGGDEVAYLHAFLDERVWAMKQEQAHS DTSRVDTAQRVFLNEGNLDLEPPLDWHVYGDAYHIG

>BN159_0394

AGSNLALGKPV TASSVEGAGFEPAKAVD GSTSTRWASLEGV DNQW IQIDLGSVTDVAQVVLKWEAAYAKSYRVEVSDDGSTW
RQVYSTTTGNGATDDLAVSGSGRYLRVYGTQRATAYGYSLYEVEVYGDSP TSPDPTTPPDSTTNLLAGRPTSVSSVEGTGFE
GGKAVD GSTSTRWASVEGVDPQWLRVDMGSKTVGRVVLKWEAAYAKSYRVEVSDDGSTWRQVYSTTTGDGGTDDLAVSGSG
RYLRVYGTQRATAYGYSLYEVEAYGTGSTTPPPSDTANLDDPAKKEIAMKLVSSFENSSLDW
RAQFAYIEDIGDGRGYTAGIIGFCSGTGDMMLDLVERYTQKPSNPLAPYLPALREVNGTDSHAGLGT SFENAWRTAAQDSVF
TKTQEEERDRVYFNPAVGQAKTDGLKALGQFAYYDAAVMHGEEGFRAIRSRAVSRVKPPSQGGNEVTF LHAFLDEREEEMRK
EEAHS DTRVSTAQRFLNEGNLHLNIPLYWSVYGESFSITS

Two N-terminal F5_F8_C domains; discoidin domain; Cell surface-attached carbohydrate-binding domain, present in eukaryotes and assumed to have horizontally transferred to eubacterial genomes.

>BN159_0876

VRAPGLDDPAKKEIAMKLVSSAENSSLDWKAQYKYIEDIGDGRGYTAGIIGFCSGTGDMMLDLVELYTDRSPGNVLAKYLPAL
RAVDGTD SHAGLDPDYPRDWRRAAQDRAFQQAQNDERDRVYFDPVRRGKADGLRALGQFAYYDAIVMHGGSDRLSFGSIR
RRALQSARPPAQGGDEVRYLHAFLDARVWAMKQEEAHS DTSRVDTAQRVFLRQGNLDLNPPLDWKVYGD SYHIG

>BN159_6425

YFLMPTDSAESPAPPVSAQAQLAREEAKERAETERAEDEKTIASLPPGLADPAKKE LALKITATAEFSSLNWRLGYGSVEDN
GDGCGYTAGAIGFCTGTHDLLDLVERYTETHPDNGLARYLPALRKVDGTD SHEGLDPGFPAAWKAEAKKPAFRAAQDEERDE
VYFNPAVRLAKLDGLGTLGQFVYDAMVLHGPDPSANGFYGLRERALHEARLPSRGGSEKAYLDIF
LDLREQAMKRKSTADTSRVSTAQREWLYDGNLSLNTPLSWRMYGEPYQVS

N-terminal extension of unknown function

>C9E0E7_STRLI

ASARATGLDDPAKKEIAMQLVSSAENSSLDWKAQYRYIEDIGDGRGYTAGIIGFCSGTGDMMLDLVELYQORSPGNVLAPYLP
ALRRVDGSDSHEGLDPGFDDWRRADQDPQFRRAQDDERDRVYFDPVRRGKEDGLRRTLQGFAYYDAMVMHGDGGGLGFGS
IRERALGRARPPAQGGDEVAYLHAFLDERVWAMKQEEAHS DTSRVDTAQRVFLNEGNLDLEPPLDWHVYGDAYHIG

>Csn_StrAM-7161

SCSPPLAGLTSATAAAVNLANRPVTTSSSTESGSGGAKAVAGSASTRWASAEGVDNQWIRIDLGSATT LGRVVLKWEAAYA
RAYVELSDDGSTWRQIHATTTGDGGTDDLSVNGTGRYLRVFGTQRGTTPYGYSLYEVEAYGAGATTPPPSSGTGLDDPRKKE
VAMKLVSSFENSSLDWRAQFAYIEDIDDGRGYTAGIIGFCSGTGDMLELVERYTAKKPGNPLARYLGALRAVDGTD SHEGLD
PGFTNAWRQAAQDTVFQQTQEEERDRVYFTP AVSQAQDGLRALGQFAYYDAAVMHGESGFRDIRSVALSRALPPSRGGDEK
AYLKAFLDAREEEMRKEEAHS DTRVSTAQRFLNEGNLDLNTPLHWSVYGEAFSLTG

F5_F8_C domain; discoidin domain; Cell surface-attached carbohydrate-binding domain, present in eukaryotes and assumed to have horizontally transferred to eubacterial genomes.

>CsnN174

AGAGLDDPHKKEIAMELVSSAENSSLDWKAQYKYIEDIGDGRGYTGGIIGFCSGTGDMLELVQHYTDLEPGNILAKYLPALK
KVNGSASHSGLGTPFTKDWATAAKD TVFQQAQNDERDRVYFDP AVSQAQADGLRALGQFAYYDAIVMHGPGNDPTSF GGIRK
TAMKKARTPAQGGDETTYLNAFLDARKAAMLTEAAHDDTSRVDTEQRVFLKAGNLDLNPPLKWKTYGDPYVINS

Figure S3. Cont.

>SACTE_5457

GAPTQFAAHHLEAAATGLDDPAKKDIAMQLVSSAENSTLDWKAQYGYIEDIGDGRGYTAGIIGFCSGTGDMMLALVERYTDRS
PGNVLASYL PALREVDGTDSDHGLDPGFPRDWAEAAKDPVFQQAQNDERDRVYFDPAVRQAKDDGLGTLGQFAYYDAIVMHG
GGGDSTSFSGIRQALAEAEPPSRGGDEVAYLDAFLDARVWAMRQEEAHS DTSRVDTAQRVFLRDGN
LNLDPPLDWQVYGDSFHIG

>SACTE_5458

APEKGLADPAKKDIAMRLVSSAENSTLDWKAQYGYIEDIGDGRGYTAGVIGFCSGTGDMMLKVERYTAARPGNALERFLPAL
RAVVGGDSHEGLGDAFTEAWARVADDTAFFRAQDAERDETYFDPAVARAEADGLSALGQFVYYDAFVMHGYADAKGTVGFRT
IRAEAVEAADPPSRGGDEEAYLDAFLDARVRAIRQEP SHSDASRVETAQRVVFVHEGRLQLETPLVWKVYGESFRIDGP

>SACTE_5886

ASILFSRGDSEPPKLTPOAAAPAPSDQEEPEAVARAEHAEIDALPPGLAAPEMKEIASRLVSSADASTLDWRSRYGAVED
LGDNGYTAGIVGFCSTNDMLQLVEAYTEEQPDNPLAPYLPALREVDGSDSHEGLDPGFTRAWREAAEDETFRSAQDRLRD
RLYFDPAVRLAKLDGLGTLGQFVYYDAMVLHGPGVEANGFYGIRDAAMAQADTAAEGGDETAFLDAGRTAIRARTVQR
DTSRIDTAQRVFLRDGNLELRTPLVWRMYGEDFRIP

N-terminal extension of unknown function

>Sros_2053

AVPAPHL SAPSQGSVTAAAADLTDPRKKDIAMQLVSSAENSSLNWKAYKYIEDIDDGRGYTAGIIGFCSGTGDMLELVELY
AERKPGNVLAKYLPALRKVDGTD SHKGLDPNYTKDWATAAKDKVFQQAQNDERDRVYFNPAVQQA KDKGLRALGQFAYYDAI
VMHGDGGDPESFSSIRRNALKKAKPPAQGGDEVKYLHAFDARKVAMKAEAEHEDTSRVDTAQRVFLNKGNDLNTPLSWKV
YGDPTYIG

>Tmz1t_3574

MERII PCTDCAEKTLDVEQLGFRVTS CDPHPERPGFCVLR FEDRSATPAAGASLAAPAAAARAGRGAVAGGVTATQAAVAKA
IVNLEFETGEVLGQYGVKVTLIPGDTGH LTFGRSQTTLGSGNLAKLLQOYCANPGARFAGRLASYLPRFLAIDESLDDD PRLHN
VLRATADDPVMRD TQDAFFDRTYWEPALRAAASLGVHTPLGVAVVYDSAVHGSWLMAMRDRTTRAVGEPAAVGEQAWIDAYVR
TRRAWLEGHARADLRQTVYRMEAFGR LIDQGFWGLEMLVVRGREISSVTLAALPPGCGYDGPQPGSRPLTLATPLARGLDVR
LLQLGLSDRGVDILADGIFGR TSNLLKAWQAQHGLAATGIADPALIGELTA

N-terminal extension of unknown function

C-terminal extension: PG_binding_1 domain; Putative peptidoglycan binding domain.
This domain is composed of three alpha helices. This domain is found at the N or C
terminus of a variety of enzymes involved in bacterial cell wall degradation.

>A7K8G4_9PHYC

MFAISMSITPENKQSIADYVKRMLALAEKELNAVGDYKGLGPNRRVQFKTALGNALIKGDPVPTPDPEQSHDTKPPAGNV
PTPGPSTGTRIPSSIVPQLATLGFSETDADTILSLISLPENSNTTEWWKNYNFASRLGDGRGWTVTLYGACSGTGDLVMILK
DLQKINPHHKLVKYIPAMEKTEGEDVRGLENLGRDIKSLGDDKEWQAVWDIYIKLYWNFARNFSDKLINRPGAKLTSPLTR
GFMVDTALNHGADLDSFGPILKGMNNDKDEATWFLDFCESRRKLLKRGFQDLDTSKTGDRCTLWANI FKSGNTSLTRPIK
CYRGYWGNTIS

PRK09752 - adhesin domain (provisional)

>Q84608_PBCV1

MSNKIEITDDNKMTIQNDFVSRMMKSIDQELVAMTNKYSFGFGPGRQTNCKKALAKALGETPVNPPVNPVTPPVVDTHIPSQV
EAPLKKLGFNTTNADTILSLIALPENSTTQWWKNYNFASCLKDRGWTVTIYGACSGTGDLMLVLESLOKINPNHPLVKFIP
AMRKTGKDDIRGLENLGVINGLDDKEWQAVWDIYVKLYWTF AADFSDKTGSAKNRPGPVMTSPLTRGFMVDVALNHGSN
MESFSDILKRMKNREEKDEAKWFLDFCETRRLKLLKAGFQDLDTSKTGDRCTLWA

NIFKEGNVGLKRPKICYNGYWGKNIVIS

N-terminal domain of unknown function, shared with other chitosanases from *Chlorella*
viruses

Figure S3. Cont.

>O12288_9PHYC

MSNKIEITDDNKMTIQNDFVSRMMKSIDQELVAMTNKYSGFPGRQTNCKKALAKALGETPVNPPVNPVTPPVVDTHIPSQV
 EAPLKKLGFNTTNADTILSLIALPENSTTQWWKNYNYASCLKDGRTVTVIYGACSGTGDLLMVLESLOKINPNHPLVKFIP
 AMRKTGDDIRGLENLGKVINGLGDDKEWQTAVWDIYVKLYWTFAADFSDKTGSAKNRPGPVMTSPLTRGFMVDVALNHGSN
 MESFSDILKRMKNKDEKDEAKWFLDFCETRRLKLLKSGFQQLDTSKTGDRAILWSELFKTGNVGLKRPIKCYNGYWGKNIVIS
 N-terminal domain of unknown function, shared with other chitosanases from *Chlorella*
 viruses

>A7RBE5_PBCVA

MFMTSRIEITEDNKSKIQNDFVSKMMEDIEQELDAMTKKYSGFPGRQTNFKRALSRALEETLVKPPIVPPVDSHIPSHVEA
 SLKQFGFSIENADMILSLISLPENSTIEWWKNYNYAERLHDGRGWTVTLYGACSGTGDLLMILEALQKINSKHPLVKFIPAM
 RKTGDDIRGLENLGKVINVLGDDKEWQEAVWEIYIKLYWTFATDFADKKNRPGAVLTSPLTHGFMIDVALNHGANME
 SFEPILKNMKNREEKDEAKWFLDFCEARRKLLKSGFQQLDTSKTGDRC TLWANI FKSGNVGLKRPIQCFDGYWGKNFIIA
 N-terminal domain of unknown function, shared with other chitosanases from *Chlorella*
 viruses

>KUTG_00281

SDVPNGPTATSKAPRPGGGLTADQRRRADQLISIFENGTATVQYGYAEDLHDGRGVTAGRAGYTTNDGDALRVIQAYTDLAP
 GNLARFAPALRQLADTGSGLPEADYIAAWRQAADDPAFRRVQDDQVEQNYFTPAMAAADRLGLITPLARAELYDAAIQHG
 HGADPDGLPALIDRATARAGKVDEAAWLTAFLDVRADDLRNPANADTKEEWSKSTDRVECFRRIAATGDYTLAGPLTVTAYG
 VTYSLA

Figure S3: Identification of non-catalytic modules in GH46 proteins. FASTA-formatted sequences are shown without the predicted *N*-terminal signal peptides.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).