SUPPLEMENT

The E3-ubiquitin ligase adaptor protein Skp1 Contains a Hydroxyproline-linked Pentasaccharide Assembled by Glycosyltransferases that are Important for *Toxoplasma* Proliferation*

Kazi Rahman[1,2], Peng Zhao[3], Msano Mandalasi[1], Hanke van der Wel[1], Lance Wells[3], Ira J. Blader[4], Christopher M. West[1]

Departments of [1]Biochemistry & Molecular Biology and [2]Microbiology & Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104 USA, and Department of Biochemistry & Molecular Biology, University of Georgia, GA 30602; [3]Complex Carbohydrate Research Center, Department of Biochemistry & Molecular Biology, University of Georgia, GA 30602; [4]Department of Microbiology & Immunology, University at Buffalo, Buffalo, NY 14214

Contents

Table S1: Oligonucleotide sequences employed

Supplemental TABLE S1: Oligonucleotide sequences employed

Targeting sequence amplification:
Tg*phyA*-disruption 2
```
a)   PhyAF1   5'-flank   5'-end   5'-GGGGTACCCGTTGGTGGCAGGAGAAT (Kpn1)
a')  PhyAR1   5'-flank   3'-end   5'-GGAAGCTTTGGCAAGGAAAGCCACAA (HindIII)
b)   PhyAF2   3'-flank   5'-end   5'-GCTCTAGAGTTCGGTCAAATCTGTCGTTATTT(XbaI)
b')  PhyAR2   3'-flank   3'-end   5'-GCGCGGCCGCGTCGAGGTTCAGCAGACTTT (NotI)
```
Tg*gnt1*-disruption and complementation
```
c)   GntF1    5'-flank   5'-end   5'-GGGGTACCCAGCCTCACACAGACGAAA (KpnI)
c')  GntR1    5'-flank   3'-end   5'-GGAAGCTTCTGTTGAATCGCCGAGAAATG (HindIII)
d)   GntF2    3'-flank   5'-end   5'-GCTCTAGAGTTCGGTCAAATCTGTCGTTATTT (XbaI)
d')  GntR2    3'-flank   3'-end   5'-GCGCGGCCGCGTCGAGGTTCAGCAGACTTT (NotI)
```
Tg*pgtA*-disruption and complementation
```
e)   PgtF1    5'-flank   5'-end   5'-GGGGTACCGTTAACAGCAGACCGCATTTC (KpnI)
e')  PgtR1    5'-flank   3'-end   5'-GGAAGCTTAGAGGAGAAAGGTGAGGAGAA (HindIII)
f)   PgtF2    3'-flank   5'-end   5'-GCTCTAGAAAGGATGTCTTGTCGCTGTATC (XbaI)
f')  PgtR2    3'-flank   3'-end   5'-GCGCGGCCGCCTGGCATTCCGTTAATCTCTCT (NotI)
```

Primers used for PCR confirmation of disruption and complementation strains (see Fig. 3)
Tg*phyA*-disruption 2
PCR 1:
```
Forward:      5'- TGCAGGGTCACTTCTGTT
Reverse:      5'- GCAAATGATGTGCAACTCCTC
```
PCR 2:
```
Forward:      5'- CTTCGTCGGCTTCAGCATT
Reverse:      5'- TGTGCTGTTGTCCTGGTACGA
```
PCR 3:
```
Forward:      5'- AGGACAGAGTTGAGAAGTTGGCGT
Reverse:      5'- TGGTTACCTATGGCAACGG
```

Tg*gnt1*-disruption
PCR 1:
```
Forward:      5'- TGAGAGTAGCGAGGTAGATGAG
Reverse:      5'- GAAGGAAGAGCAGGAAGTACAA
```
PCR 2:
```
Forward:      5'- GATCGGAAACACACGACAAAC
Reverse:      5'- TGTGCTGTTGTCCTGGTACGA
```
PCR 3:
```
Forward:      5'- AGGACAGAGTTGAGAAGTTGGCGT
Reverse:      5'- TCCACGTGGTTGAGAGATTG
```

Tg*pgtA*-disruption
PCR 1:
```
Forward:      5'- CTCCGTTCTCCGAGTTCTTTC
Reverse:      5'- TTCCTCTCTTCACCTCTCTCT
```
PCR 2:
```
Forward:      5'-TGGAGAGAAACATGCAGGAA
Reverse:      5'-TGTGCTGTTGTCCTGGTACGA
```
PCR 3:
```
Forward:      5'- AGGACAGAGTTGAGAAGTTGGCGT
Reverse:      5'- CTGTCTCGACTGCCTTCAAATA
```

SF tagging of TgSkp1
Amplification primers
```
g) SKP.LIC.F  5'-TACTTCCAATCCAATTTAGCGCGTGCCTTTGCTTTCACTTGTCT (LIC)
g')SKP.LIC.R  5'-TCCTCCACTTCCAATTTTAGCTGCATCTTCGCACCACTTGTTCTC (LIC)
```

Tagging confirmation
```
l) Forward:   5'-TCGCCTCCATGAGTGCGTTGATAA
l')Reverse:   5'-ATTTACAACTCAGTGCTCCAGCCG
```

2

FIGURE S1. **Alignment of Gnt1-like sequences.** The amino acid sequence of *Dictyostelium discoideum* Gnt1 is aligned with the amino acid sequences of predicted Gnt1-like proteins from 3 other amoebozoa, 4 coccidian apicomplexa, and 2 chromerids that also possess *phyA*- and *pgtA*-like sequences. Amino acids are color-coded with respect to chemical similarities that guided the alignments, giving preference to the registration of hydrophobic residues: green, hydrophobic; blue, acidic; dark red, basic; black, polar; bright red, secondary structure breaking (P or G). To facilitate the aligning, positions occupied by identical amino acids across all the organisms are bolded and similar amino acids are highlighted when they have majority representation at a given position. Numbers in parentheses refer to the position of the amino acids to the right. Origins of the sequences are given at the bottom. Sequences whose expression has been confirmed at the transcriptional (EST) or proteomic (MS) level, as reported by http://toxodb.org/toxo/, are annotated at the right.


```
Dd (1)      MNENSIFVSIISYRDSECQWTIKNLIELAKYKENIFIGVCLQYSM-NDDSDNKCFQFNFEEEYGKNQIRI-----
Dp (1)      MENGNIFISIISYRDSECQWTIKNLIENSNQPNSLYIGVCLQYDMGNGEGDKHCFEYQVDKEYESKHIRY-----
Pp (1)        MSLFVAIPSYRDPECKHTIIDLFQKAENPDLIRVGVCNQIHPTEDLDLTLHDLPRKSQ--------------
Ac (1)        MEDTIFVSIPSYRDPELVHTVRDLLSKAHHPELVYVGICWQWRPDEDQPYVEEFDRFLARHECEDEAEAVASST
Tg (1)  MDACGRTKAHQR-VSCSFAPIGSEIEKELPVRISWSVASYRDNQLASTLLSAFHFAHFPSRLFPVVLWQGELQAPCCLCRTAKVPETGPGTGGKETNFTKAD
Hh (1)  MDACVRTKAHQR-GSCSFAPIGSEVEKELPVRISWSVASYRDNQLASTLLSAFHSAHFPSRLFPVVLWQGELQAPCCLCRTAKVPETGRGTGGKETNFIKPE
Nc (1)  MDAYRTKKLPQRRVSCSSAPVAPAVENEFPVRISWSVASYRDGELADTLLSAFHFAHFPGRLFPVVLWQGELHASCCLCRTEQAAETRAGAGRNKACTEPNT
Sn (1)            MQQHQQPLIAWSVTSYRDSQLLPTLQSGLYQAHFTSRLLPVVCWQGDLKAPCCSCYSISCRVDVVGGTVSSCDGLSRH
Cv (1)    MEIASGDGDGSDRVSGDSVFVSVAAYKDPQLMDTLKSLLHHAANPLSLSIGVALQDNFEEAID--LRSLHPFFCLPGKAQAGQEENFV
Vb (1)        MRKILVLIPSYRDDELPLTIQSAIHQADAPREISFGCVWQGDLRNKDKTAADDLITSFTTFNPALRAKWLGSHR

Tg  AEES-STS-FPHCQTKCTKSVSDVAQVEKGDAHQGKRSQRNNTLESSEVDEEN-GDAVESVSTDGYRGHGKEPLTSSKSGREGTAAAPDTRFEPVSRSVSS-----
Hh  ADES-SLS-FPHFPTKCTKGVSDAAQVEKGDAHQEKRSQRNGTLESSEVDEGN-GDAVGSVSRDGYRGHGREPVTGSKSEREGTAAAPDTRLEPVPRFVSS-----
Nc  EQSCLSRSHFVPRFS-APVQRNDVAHAEMRHGRRGGDGQ-QNGAEASRGDEGNPGDVKRSRRERREEPRG-ERSTNSKCESGGNADASDA--ESLSLAVSS-----
Sn  GKHLQKQYSSACVATGDCSTTSLPIIRGDGAATQTADVASNQAAATTCPIATAASTVALCSVAAASLAPASSAADVAASSAAASSSDSLELRTPQKHERNISCIHP

Tg  ----------------------------CAQKQRRLDGCHCWCHTFKETVSSFASQSNLLEAHPVGRDSRHPRAADSAEGRDGQSASYERIWRWQVERGKRLTE
Hh  ----------------------------CPQKQRRPDGCRCWCHTFKETLSSFSSQSNLLEAHPVRRDPRHARGADLAKGRDGQNASYERIWKYRVERGRRLTE
Nc  ---------------------------RPKTKRHPPRRRYCRCWCHKAVECLSSR-PRTSIGEEDDSKQRGPSGGTSSSQMVQTNAHPLPASSPVH--------CE
Sn  SSSVRQPHLCRTNRESPIFSSCYSAAVITHDAPRKAAGDKCACWCHSL-TAARSRVSAAVLQEPQTIHAGDSRVLSETNIFEQTDDSYGRGITSATVPQNTPPSAL

Tg  RKETAKYFLVPCDNASETKPPAVEPNPATAVGEPPTKSNLNEESFDLWELALFIWRRRIGADASLSHRQRILARKKRGGKHFPNTSAFSGENACDIKRVCGEKQN-
Hh  KKGTAQYFPVPCENESETQPPAVEPNSATAAGQLPTESNMNEESFDLWELALFIWRRRIGADASLSHRQRILVRKKRGEKHFAKSSACSGENACDIKRLYGEKKNK
Nc  RHETL---GVRRPQAPSPKPP------ATAV-EDPAGSCVNEESFDLWEVARFIWHRMNGPKSCLAQRRKISTRKTLD-KKLGNLPGRSGEIVRDGENRRG-KTE-
Sn  LQRHCRRRWRRAADDTAPTGLNRRATATVVEGGTATKEATPRAAAAPSSAAAASSCAAGKPPPDLGNEGVVTKGAAEDACVAAGKAGELAAADADSRSLTATNNCR

Sn  ARLISENEANASPSERELLNEPRSSICSGVLSAASAEQGLSYHQREAVLRVDEIPADTEDFHSTTEECLASQRRCWDGGDVTFFPPSASRQNGAVVRAPKRGGTQE

Sn  VCIAGIREGSLDCRDGQQKGTESLNREAFPLLELYRTCWAKPRTLQHPKWGTGFTRHFRKEELSSVRSKAGNSCCDREVTPESCATGGSRRGCSTCCSGRPCPVKP

Sn  LDKSRHLEASSCSSGTNCLWRSTCSSRARVTRSGTRTRRVRPGSEGGVRQTSRSVGSDGNSSSCVYVSSTRSTSVCSVAMTSSSCSTKPSRRFRHARVNPASSYFS

Tg  --------------------------------------------EGTQRRKTSLGAPEMSERDTREAQPTQPHPQTAQERNGRASASLSSFSASFPDSVTTALPMLPPG
Hh  ----GIQRKSTSLGAERSMAQTPSGKAGISEMGDTSVCCEETPACDRHNGAIAHELSERDTQEAQPTQPHPQTAQERNGRASASVSSFSASFSDSVTTALPLLPPR
Nc  ----------------------------TGDWAGKRFRLNKQSCERPSGKSAHDLINRTNALAPTQPRAQTPPATTGVAAPA---SRSTSCASSST--PV-TSP
Sn  SASSFSSATCHCSGHTSGVPPLSTGRSDHSCDCWCTVGRRLRRPWSSPPLSFLPQNNTSPWPVAVPRGNCSADAGQISHKTKAPCVRRSSRSGRFGAASSAAAGAA

Tg  FSLSLKRVAMRQIIDLLPRWLQMSLLAR--KPPFS-LDVTEGVEESHDEEDA-HEAATARHLPLRCYLLSVVLPALPSEEDKARFSFVVEDSRKGEDQDIGQQTRQ
Hh  FSLSLKRVAMREIIDLLPRWLQLSLLAR--QPPFS-LEAAEGGEESQDVGDA-HEEATGRHLPLRCFLLSVVLPALPSEEDKACFSFVVGDSRTGGDEDVGQRTRN
Nc  FAPSLKRAALREIADLLPRWLQLSLRSRCTHPEVSSFDNEEGVWAKNEGMEGDHERTRSGNLPLRCFLLSVALPSFPPSQEDEESSSVLKDAGNGDPSLARPLAD
Sn  -TVFLQNQPLCSVQHLLPLWVQQQLLSS----TAT-KVVEEPRRQTSACCSSTFRLSAEPPVTACLHLLSLLLPPTSSTCPTPAVKHEEKQEHQQERKESIGTLSA

Tg  REHKQETAHQLLRIPVLELLMPKIKERRTTSRIDATSRAPVSSCSSTSFE-SSVHAPSWRPP-LSRSSPLPA----SSPSESEVPPASK-----------------
Hh  REHKQETAAQLLRIPVLELLMPKINEERTTGRFAATNRESLSSCSPTFFQ-SSLHASSWRSP-LSRSSPLPA----SSPSESEAPTASE-----------------
Nc  AEHTHEGANRVLRIPVLEILLPQPNKNPFAAALAAGGQSPSAPCSRTASEPPSVVASSRSCFQLSRSSDASAWARDASPSESG----------------------
Sn  AARATRKLQPHPSHEEVPEGSLRGGSQRVRSEKCCTSSDSSNGCSVGSEKKRTSKSRSCGPSCTSYTRSTSGCRPPKCRHCSGEDNRHRSRRRRYACSGDESYSST

Sn  CCRRPSSTSTCSISNDSNCSVASAPCVAQCYSNSLCKSIISFLPPRPYAGSRSLNSLHSASCRRLSRNRSARKSSYSHSSSNSSVGSSRRKASGCNGGINLLHKGY

Dd (70)   ---------------------------------------------IRMNHTEAK---GPCYARALVQQQ-LF-------------KGE----KY
Dp (71)   ---------------------------------------------LRMDYRDAK---GPCYARALVQQQ-LY-------------RDE----EY
Pp (59)   ---------------------------------------------VRIKEVDCRTATGPCYARSI-TQS-LW-------------EGE----EF
Ac (75)   -------------------------------TTTSTTTNTAHPRVPQRVRIAWVDWREARGPCHARRV-AQA-LY-------------ADQ----RY
Tg (670)  -QVSEREDIRRASGKGEEEDVDAPDRGDDRRSVRIEEEAREDEEEDERACMRIAFLDWRESRGPCLARAI-CEW-LLPVSP---SREARTSGEETRLEL
Hh (669)  -QVSEREEIRCAAGKGEEEDVDASEKGDDRRSVRTEEEAREDEEEDERACMRIAFLDWRESRGPCLARAI-CEW-LLPVSP---SREARPSGEKTRLEL
Nc (621)  -------QGSEGEDNAATGEEAHDAGRNVEDELGEEAWETDEDGDETAVMCLAFMDWRDSRGPCFARAL-CEW-LLPTSPLREATAVRKPGE-TRLEL
Sn (1132) -----------AAGEGERLQLAVVDLKGPCTAADAAAVVGRSEAAENLRIRLVFLDWRSTRGPCFARYI-CEQ-MLPAA----SSLSADRGE----LL
Cv (87)   -------------------------PTRHPLQFGEEAFESLSFFADRLRVICVPADQAKGPCWARSL-CQS-LL-------------GSE----EF
Vb (74)   PSHRTAGPGRQDDRASGGSSTVQDDPQHGPAACRPRARAPHPLFRFECGYLKVLVLDKGAARGPAWARHL-AQL-LW-------------EGE----PY

Dd  YLQIDSHMRFVKDWDIEMINQL------------------------------------------------------------------
Dp  YLQIDSHMRFVKDWDKILIEQL------------------------------------------------------------------
Pp  YLQIDSHMRFVKGWDAMLKKYL------------------------------------------------------------------
Ac  HLALDSHMRFIPGWDSVLVGLL------------------------------------------------------------------
Tg  FLQTDSHMRFAPHFDCFLLRQLKLAAALSAERRRAQ-ASSQST--HSSFSPHRGASSSG-----------------------------
```

3

```
Hh  FLQTDSHMRFAPHFDCFLLRQLKLAAALSAERRRGQ-ASSLSP--HSSFSPYRSASSSG----------------------------------------------
Nc  LLQTDSHMRFAPHFDCFLLNQLTLAAALSAETKRARSPSTLSPPSHA-FS--LSANASP----------------------------------------------
Sn  LLQTDSHMRFAPHYDCFLVRQLYLAAAAATAQERNVDKIVATTYGSAVAGGFPRPTTEQQERRLTAGPRRKKATETAAGSDLGHSAAITSESHLGEGGGGVAAVQS
Cv  FLQIDSHMRFAPGWDHILLEDL-----------------------------------------------------------------------------------
Vb  LLAIDSHMRFRPKWDTLLLEDL-----------------------------------------------------------------------------------


Sn  EGDEKLCCDHESNTRREKGQLAFQEQGKKTAEGKRIEAAQKRRERPRKRGQVALSAEEEGAGKQAISLFGRPQHSDVYSRSAACRRHSKFFSDSQKKHVSFLSCCV

Sn  ACGSCCDCPCSSCSCCCSVPQRRVTCSAKKQLPFRKQLSLLQRKNGTHRCTPKRQIWPRRNQQRLLRRACKPLRAYLVSKKSAQTPDKKLHLLQKTVLSFYRKRRR


Dd  (120)  -----------------------LQCKKPNDDNGGMVIDEKAILTCYPMGYKLPNLIPT-----------------------------------------
Dp  (121)  -----------------------QMCKVN------GSVDTNAILTCYPMGYTLPNKIPV-----------------------------------------
Pp  (111)  -----------------------SQTTNP----------DKSIITSYPVGYEQPNLVPS-----------------------------------------
Ac  (144)  ------------------------ARCP-----------SPKPLLTAYPAPYLRPNSVTTD---------------------------------------
Tg  (780)  -------------------TSFDLLRSACSSLLL----TEKVILTGYPPGYEEGTPFFEYPRPPQTSGQEAFSSFSNCLSEACAAPSLISLLPPAQTL
Hh  (818)  -------------------SSADLLRSACSSLLL----TEKVILTGYPPGYEEEMPFFEYPRPPHTNGQEAFSSFSNRPSETCAAPSFIPHLPPAQTL
Nc  (765)  -------------------ARSSAPPSACSSLLV----TEKVVLTGYPPGYEEGLPFFEFPQPPHTSGTTESPSVGACPSDDT---RHLPPRPSSQTF
Sn  (1527) EHRAEKSQRAVAVRQQKQPLHTELPQRQAREHMLLQSCRLPRVILTAYPGYDEGLQFYEWPDS-LT-GVASQSHGGHSVREKSARAAAAGVEAYEQ--
Cv  (161)  -----------------------RTCLKDT--------PRAVLTAYPPGYSEQDSFDSVP----------------------------------------
Vb  (176)  -----------------------GRCPS----------SKPILTSYPPAYPHGVGWDYVS----------------------------------------


Dd  ----------------------------------------------------------------------------------------------------
Dp  ----------------------------------------------------------------------------------------------------
Pp  ----------------------------------------------------------------------------------------------------
Ac  ----------------------------------------------------------------------------------------------------
Tg  SSLPPVSSSTSSS-FASSACLSSSTSSSSASSAFPTSTSSFSSSPSSSSTSYLSPFTSSASSSSPTLTSCFTASSCDSWCSSSEAERSGVVDGPATLCPQQV EST
Hh  SSLPPLSPSTPSSSFASSSCLSSTSSSSTAS-ASPTSTSSFTSSP-SSSTSYFSFFTS-ASSSPPTFTSCFTASSCASWCSSGKVERSGMVKGPAALCPQQV
Nc  -SLPCVSSSSSSSRSCEREGFATPHAPSASRSRRLCASGLEAASA----------------------------------------------------------
Sn  --------RKQEQNAAVGARVVPTDIHSNDKKQCAPAAA-----------------------------------------------------------------
Cv  ----------------------------------------------------------------------------------------------------
Vb  ----------------------------------------------------------------------------------------------------


Dd  (156)  ------------------------------------------------------------HRFP-ILLVASGFGENDGFLRLGGKIVS
Dp  (152)  ------------------------------------------------------------HRYP-ILLVATQFG-DDGFLRLGGKIIS
Pp  (137)  ------------------------------------------------------------YKNA-CFLVAKGFG-DDGMLRLDGKLLK
Ac  (171)  ------------------------------------------------------------PRPP--FLCAREFGADDGMLRTCGKLLH
Tg  (956)  CSSSSAEVSPVARGIQDSRDLVLPLQAKINEGKEKKILTVGRDETGEGEKNAEYETLNEIPRLSFPETYFPGILLCAGHFD-RNGLLRTKGRMLR EST
Hh  (992)  CSSSSAEVAAVARGIQDSGDVVLPLQAKFGERKEKEVLTVGTDEASEGEASREYEKLNEIPRLSFPETYFPGILLCAGHFD-RNGLLRTKGRMLR
Nc  (880)  -------LP-AHGGQEPRGQSEEGRSEGTEEERPTGRRNEPGEGGQDLDLAEW-T-KEKTGQRFPEVYFPGILLCAGHFD-RNGLLRTKGRALR
Sn  (1653) -------VLAGSAADESPDYEPCVGAGTASEASRVHSLAVAAIPATGQRERKELASATRSASFCGHALLPPVLLCAARFD-EQGVLRTTGRVLR
Cv  (190)  ------------------------------------------------------------AERRPVLLCGWKFD-AKGMLRTKGRLLK
Vb  (203)  ------------------------------------------------------------DETRPVLLCASHFD-ENGLLRTKGRMLT


Dd  KKLIERE---------------------------------------------------------------------------------------------
Dp  KKL-------------------------------------------------------------------------------------------------
Pp  QCL-------------------------------------------------------------------------------------------------
Ac  LPLPATTAATAAPTSGAQQGQSQARAQGKPG--------------------------------------------------------------------
Tg  VPGPTLKAACMGSGGAGQGSTRKEHIIDDGCSSLLASSPPIPSPSTSLSCASSSFSDASTVPLSFP-SSESSGLGVDARDVPEKTTHRLNQVTACSCPCLHC EST
Hh  MPGPTPKAASVVSGGSGQGSTRKQHIIIDGCSSLLAASPPTPSPSTSLSCASSSFSDASTVPLSFP-SSESSGLGVDARDAPEKAIHPLNQVTVCSCHCLHC
Nc  FDGSASKAEPRGSGGDSQGATRNRC---SYAGSSSASAPPLASASFSAAADPVPSESSGFSLEEGS-REGNLSMKMTHLVRPP--------TACACSCLHC
Sn  QAPETTRREQPCILQQRDQQRHSREVHPWRERFEDCVLLEDQERQGRQRQEVDLRQQHGEQQGPPQ-QQGLAQQQKQNEQLQEEEHHQPQHDAWKQERQPRE
Cv  SPL-------------------------------------------------------------------------------------------------
Vb  SHL-------------------------------------------------------------------------------------------------


Dd  (191)  -------------------------------------------------------------------------NNPCSSLFWV
Dp  (186)  -------------------------------------------------------------------------DSPCKSLFWV
Pp  (166)  -------------------------------------------------------------------------KEPKPSSFWV
Ac  (228)  -------------------------------------------------------------------------VGPLPSLFWA
Tg  (1151) CGFAEAACGASEGGPSKPAESAPPSFLASSDLSAARSSPSFFAPPSSS-GSLNC--------------------------FRPLESLFWA
Hh  (1188) CGFEEAACGASEGGHSKPAESALPSFLASSDLSSARSSPSSFAPPSSSSGSLNC--------------------------FFPVKSLFWA
Nc  (1052) CEL-EEAPGTCEQERASLAESPASSVSSSALSLSPSFSPS-----------TC--------------------------LFPLKSLFWA
Sn  (1840) VSEVCGPATLLNSVVGGSAVAGCLCSPSVAKRDAKEATTTRSGRPNDAAVSVTEHEARVHQACCCCCYKRALREKQGRKSAAAEGNLASVQSVTPSLFWA
Cv  (220)  -------------------------------------------------------------------------EVPILSLFWA
Vb  (233)  -------------------------------------------------------------------------PSPCPSLFWA


Dd  SGFSFSRSDIINSVPYDPNLQYLFFGEEISMSARLFTHGYNFY-SPTKTLIFHLWNRDYRSTFRENNSLEIQKLEENSKKRLLILFNQNNNNINDNDDNNNNNNNN
Dp  SGFSFSRSNVIKEVPYDPNLQHLFFGEEISMSARLYTHGFNFY-SPTKTIIFHLWNRDYRPTFRENKSDETIKIENHSKKRLLKLFGLEN----------------
Pp  SGFAFSSSKVIQEVPYDPHLHYLFFGEEMLMGARLWTHGWDFY-CPGESIIYHLWTRSYRKTFRETTNPERDALEQKSKERVCKIMGMKP----------------
Ac  SGFSFSRAEVLQEVPYDPHLPFLFFGEESSMAARLWTHGWDFF-SPPHHVIYHLWSRSYRPTFWEVEDKEKLKERSLRRVRCLLGA--------------------
Tg  AGFSFGPARVTREVGYDPRLHFVFFGEEQTMTLRLFTHGWSFY-APRFSVVFHLWTRARRPFFKADLFRLLSEDGDKRNERTTGRKRRCTQQAKSVLSAHVPLHKT
Hh  AGFSFGPARVTREVGYDPRLHFVFFGEEQTMTLRLFTHGWSFY-APRFSVVFHLWTRAHRPFFKTDLFRLLSEDGDKRNERTTGRKRRCTQQEKTVLSARVSLHKT
Nc  AGFSFAPARVIREVGYDPRLQFVFFGEEQTMTLRLFTHGWSFY-APRFSVVFHLWTRARRPFFKTDLRHLFSEKEKPLGDTCLAVSSSISSSVCSPFSSSVCSSLS
Sn  AGFSFGPARVKEVGYDCQLPFVFFGEEPTMAMRLFTHGWQFF-SPAASVVFHLWSRRHRPFFRQELHQLHIGSNKCGSTSSCRTSTRCTNEPGSSAPAAAVEPQS
Cv  AGLSLSPSSFVREVPYDPSLEFVFFGEEPSMLCRMASWGWRCF-CPSQSVVFHLWERGGRPSVPGSGQGEETEKKEKRERGEERIRQLIGLKKRNDTPGPPQQTVS
Vb  AGFWFGRSDVLREVPYVP-LPDIFFGEEQVMTLLLFRHGWDFY-CPTKSVVSHLWKREHRPFFSSDQGAAPPAMCSRQASLGAV---------------------
```

```
Dd   NNNNNNNNNNNNNNNNNNNNNNNSSSSSSSSNNNNNNNNNNNNNNSSSTNNNNNNNNNNNDD--------------------------------------------
Dp   ---------------------------------------------------------------------------------------------------------
Pp   ---------------------------------------------------------------------------------------------------------
Ac   ---------------------------------------------------------------------------------------------------------
Tg   RDTTETDSWRTVHENETGDTP-TTTRDIREETLPRPIDAALCFSHVAVSKSPL------SFTEE------------------------------------- MS
Hh   GDTTETDAWRTVHEKQTGDTPATTSRDINEETLPRPIDGSLCFSHVAVFKSPLCPSQSASFTKE-----------------------------------------
Nc   SISSSVSSSVCSSVSSRALPPAGSTQKRVEAVETRREITSTAERKRQRREERKAKGEKEECRDGWKATQGRTEAESPSGSESQGPAPRIATTRKGLIGDEQK
Sn   AAGV----SRLEQHDVRFGEVAARAKHDINGEAHEKQVRTYGTVKASVISGQETKALRNQFENCTEEATKVEGVSSAPSVSCCSMCRAHAAAKLRM------
Cv   VPRGPWSV-------------------------------------------------------------------------------------------------
Vb   ---------------------------------------------------------------------------------------------------------


Nc   SSRASYASHSFDLSAVARPPSCATTQEGKEGRSAGTCGELEADRNGEEETKDTRDACKGTRAREAERLPARLGAYVSEDGDWKERKSGNKLQWDQTEGSKEAATAG
Nc


Dd   (360)-------------------------------------------------------------------------------------GIK-IELKYNLGKIK-
Dp   (278)------------------------------------------------------------------------------------DNQIGEIESKYGLGATR-
Pp   (265)---------------------------------------------------------------------------------LTSPVSTDPEIEIELSKYGLGEKR-
Ac   (323)--------------------------------------------------------------------------------TPRDQVEAECLSELDLYGLGRAR-
Tg   (1383)--------------------LPTPAPP-LSSPSSSS--LSLFFPPRLSSSSSRPSL-----AEMLSPGEARVHRILSATAESTTAADLCSLGLGSQRR
Hh   (1420)--------------------LPTPAPSSLSSPSSSP--LSLFFSPPLSPSSSLPSL-----AEMLSPGEALVHRILSATAESTTAADLCSLGLGAERR
Nc   (1416)LETNEAKKTDAWRNADTHETRDSPWGGSRGLPSPASSSASLPPFWAGSLSPASS-PSLPRSAFASMLSPGEARVRRILSATA-SSCPAELRSLGLGSERP
Sn   (2137)-------------------------------------------------------------------------QCIFAADSGLLSQQELRQLGFGRKRA
Cv   (343)--------------------------------------------------------------------PSADLDTGDGQGGSGHSEVPLLGPEA
Vb   (325)------------------------------------------------------------------RAILAGRDDDRPDAPCER


Dd   SLDDYSNYCGVDFKNKTINNKGKF------GGYYEERETFFMNEIMEYVIKSQIGI*(423)
Dp   TLENYSDFSGVDFKNKSLNNNAKF------GGYFEEKDTFFMNEIMEFVIKSQIGI*  (345)
Pp   SLDDYMNVCGVDFQNRKIASKAKL------GNLHES---DFADPVLDMVMKSSPLVPLT*(338)
Ac   SLEEFQQHTGIEFKTRQIGDKAKW------GGLPST---MFVEGMAQFAL-SLAGLSLDLSPPSSPSSSTEKFT*  (409)
Tg   P-ESFWREIDVDAAKKVISRRARCPSQSANGGYSED---IFQVTAEEHRAQREGVQLLLSLLTERKKDM*(1510)
Hh   P-ESFWREIDVDVAKKMISWRARN------GGYSED---IFQVTAEEHRVQREGVQLLLSLLTQRQKDTCN*(1551)
Nc   P-EAFWREIGVNVETRFISKRARD------GGYSED---IFQVTAKEHQTQREGVQLLLSLLREHEKETRC*(1574)
Sn   P-EDFWKRGGIDWRQRRISPLAYR------GGLPHDE--AFEGGGPTAAAAINLVLELLRERPLQPENNNRKTGDSCSSPRIASSGRSCCS*(2244)
Cv   TPRSLWAERGIDIDRTEVLPLAAR------GGFSHER--SFAETEHKENVAEVLHLLA-QKGLFGPG*(426)
Vb   PAQAYWDMIRVDPQTQRISDWALQ------GGMESEE--AFVLTRAAEEKLSRLMSIFTQRGIIAPVDRPS*(405)
```

MS = sequence supported by mass spectrometry at ToxoDB
EST = sequence supported by EST sequences at ToxoDB

Origin of Sequences:

Amoebozoa–
Dd: Dictyostelium discoideum, GI:60471644
Dp: Dictyostelium purpureum, jgi|Dicpu1|89733|fgeneshDP_pm.C_scaffold_328000004
Pp: Physarum polycephalum, contig assembled from multiple overlapping raw genomic fragments
    accessed at Physarum database (http://genome.fli-leibniz.de/ggl_blast/blast_ggl.pl)
Ac: Acanthamoeba castellani (Neff strain), GI:470526279
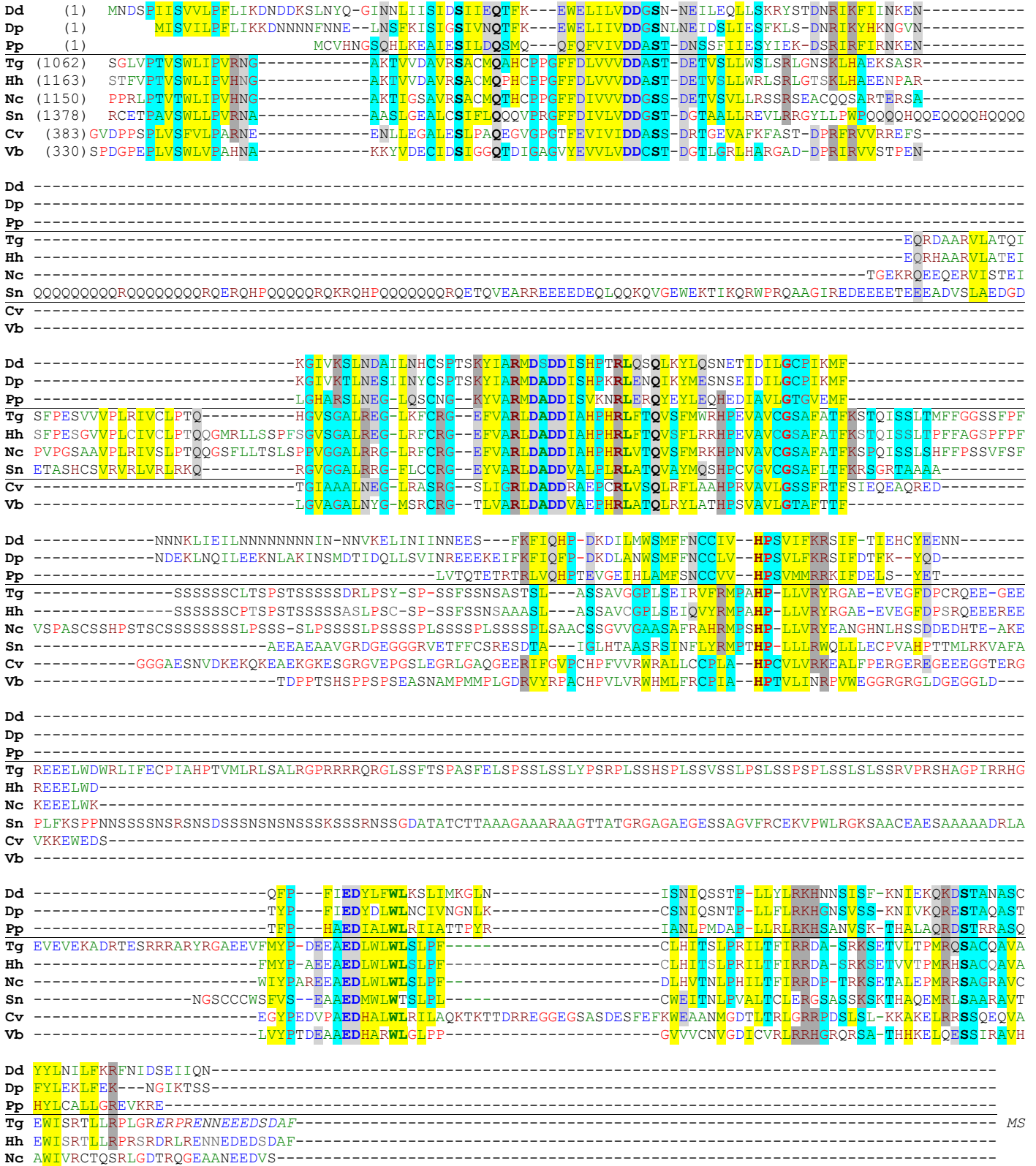
Apicomplexa–
Tg: Toxoplasma gondii, From ToxoDB 7.3: TGGT1_315885
Hh: Hammondia hammondi, From ToxoDB 7.3: HHA_315885
Nc: Neospora caninum, From ToxoDB 7.3: NCLIV_058510
Sc: Sarcocystis neurona, From ToxoDB 7.3:SN3_00900395

Chromerida–
Cv: Chromera velia, From CryptoDB, Cvel_16021
Vb: Vitrella brassicaformis, from CryptoDB Vbra_13029

FIGURE S2. **Alignment of PgtA-like sequences.** Sequences related to DdPgtA are aligned as in Fig. S1. Since the order of the two predicted GT domains of the apicomplexan and chromerid *pgtA*-like sequences is reversed relative that of the amoebozoa, N-terminal regions of the amoebozoan proteins are aligned with the C-terminal regions of the other sequences, and the C-terminal regions of amoebozoan PgtA sequences are aligned with N-terminal regions of the other sequences. Sequence origins are listed at the bottom. Sequences whose expression has been confirmed at the transcriptional (EST) or proteomic (MS) level are annotated at the right.

β3-GalT domain

```
Dd    (1)   MNDSPIISVVLPFLIKDNDDKSLNYQ-GINNLIISIDSIIEQTFK---EWELILVDDGSN-NEILEQLLSKRYSTDNRIKFIINKEN-----------
Dp    (1)      MISVILPFLIKKDNNNNFNNE--LNSFKISIGSIVNQTFK---EWELIIVDDGSNLNEIDSLIESFKLS-DNRIKYHKNGVN-----------
Pp    (1)              MCVHNGSQHLKEAIESILDQSMQ---QFQFVIVDDAST-DNSSFIIESYIEK-DSRIRFIRNKEN-----------
Tg (1062)   SGLVPTVSWLIPVRNG------------AKTVVDAVRSACMQAHCPPGFFDLVVVDDAST-DETVSLLWSLSRLGNSKLHAEKSASR-----------
Hh (1163)   STFVPTVSWLIPVHNG------------AKTVVDAVRSACMQPHCPPGFFDLVVVDDAST-DETVSLLWRLSRLGTSKLHAEENPAR-----------
Nc (1150)   PPRLPTVTWLIPVHNG------------AKTIGSAVRSACMQTHCPPGFDIVVVDDGSS-DETVSVLLRSSRSEACQQSARTERSA-----------
Sn (1378)   RCETPAVSWLLPVRNA------------AASLGEALCSIFLQQQVPRGFFDIVLVDDGST-DGTAALLREVLRRGYLLPWPQQQQHQQEQQQQHQQQQ
Cv  (383)   GVDPPSPLVSFVLPARNE------------ENLLEGALESLPAQEGVGPGTFEVIVIDDASS-DRTGEVAFKFAST-DPRFRVVRREFS-----------
Vb  (330)   SPDGPEPLVSWLVPAHNA------------KKYVDECIDSIGGQTDIGAGVYEVVLVDDCST-DGTLGRLHARGAD-DPRIRVVSTPEN-----------

Dd    -------------------------------------------------------------------------------------------------
Dp    -------------------------------------------------------------------------------------------------
Pp    -------------------------------------------------------------------------------------------------
Tg    ----------------------------------------------------------------------------EQRDAARVLATQI
Hh    ----------------------------------------------------------------------------EQRHAARVLATEI
Nc    ------------------------------------------------------------------------TGEKRQEEQERVISTEI
Sn    QQQQQQQQQRQQQQQQQQRQERQHPQQQQQRQKRQHPQQQQQQQRQETQVEARREEEEDEQLQQKQVGEWEKTIKQRWPRQAAGIREDEEEETEEEADVSLAEDGD
Cv    -------------------------------------------------------------------------------------------------
Vb    -------------------------------------------------------------------------------------------------

Dd    -------------------------KGIVKSLNDAILNHCSPTSKYIARMDSDDISHPTRLQSQLKYLQSNETIDILGCPIKMF--------------------
Dp    -------------------------KGIVKTLNESIINYCSPTSKYIARMDADDISHPKRLENQIKYMESNSEIDILGCPIKMF--------------------
Pp    -------------------------LGHARSLNEG-LQSCNG--KYVARMDADDISVKNRLERQYEYLEQHEDIAVLGTGVEMF--------------------
Tg    SFPESVVVPLRIVCLPTQ----------HGVSGALREG-LKFCRG--EFVARLDADDIAHPHRLFTQVSFMWRHPEVAVCGSAFATFKSTQISSLTMFFGGSSFPF
Hh    SFPESGVVPLCIVCLPTQQGMRLLSSPFSGVSGALREG-LRFCRG--EFVARLDADDIAHPHRLFTQVSFLRRHPEVAVCGSAFATFKSTQISSLTPFFAGSPFPF
Nc    PVPGSAAVPLRIVSLPTQQGSFLLTSLSPPVGGALRRG-LRFCRG--EFVARLDADDIAHPHRLVTQVSFMRKHPNVAVCGSAFATFKSPQISSLSHFFPSSVFSF
Sn    ETASHCSVRVRLVRLRKQ---------RGVGGALRRG-FLCCRG--EYVARLDADDVALPLRLATQVAYMQSHPCVGVCGSAFLTFKRSGRTAAAA---------
Cv    -------------------------TGIAAALNEG-LRASRG--SLIGRLDADDRAEPCRLVSQLRFLAAHPRVAVLGSSFRTFSIEQEAQRED---------
Vb    -------------------------LGVAGALNYG-MSRCRG--TLVARLDADDVAEPHRLATQLRYLATHPSVAVLGTAFTTF-------------------

Dd    -------------NNNKLIEILNNNNNNNNNIN-NNVKELINIINNEES---FKFIQHP-DKDILMWSMFFNCCIV--HPSVIFKRSIF-TIEHCYEENN-------
Dp    ------------NDEKLNQILEEKNLAKINSMDTIDQLLSVINREEEKEIFKFIQFP-DKDLANWSMFFNCCLV--HPSVLFKRSIFDTFK--YQD---------
Pp    ------------------------------------LVTQTETRTRLVQHPTEVGEIHLAMFSNCCVV--HPSVMMRRKIFDELS--YET---------
Tg    ---------------SSSSSSCLTSPSTSSSSSDRLPSY-SP-SSFSSNSASTSL---ASSAVGGPLSEIRVFRMPAHP-LLVRYRGAE-EVEGFDPCRQEE-GEE
Hh    ---------------SSSSSSCPTSPSTSSSSSASLPSC-SP-SSFSSNSAAASL---ASSAVCGPLSEIQVYRMPAHP-LLVRYRGAE-EVEGFDPSRQEEEREE
Nc    VSPASCSSHPSTSCSSSSSSSSSLPSSS-SLPSSSSLPSSSSPLSSSSPLSSSSPLSAACSSGVVGAASAFRAHRMPSHP-LLVRYEANGHNLHSSDDEDHTE-AKE
Sn    ------------------------AEEAEAAVGRDGEGGGRVETFFCSRESDTA---IGLHTAASRSINFLYRMPTHP-LLLRWQLLLECPVAHPTTMLRKVAFA
Cv    ----------GGGAESNVDKEKQKEAEKGKESGRGVEPGSLEGRLGAQGEERIFGVPCHPFVVRWRALLCCPLA--HPCVLVRKEALFPERGEREGEEEGGTERG
Vb    ------------------------TDPPTSHSPPSPSEASNAMPMMPLGDRVYRPACHPVLVRWHMLFRCPIA--HPTVLINRPVWEGGRGRGLDGEGGLD---

Dd    -------------------------------------------------------------------------------------------------
Dp    -------------------------------------------------------------------------------------------------
Pp    -------------------------------------------------------------------------------------------------
Tg    REEELWDWRLIFECPIAHPTVMLRLSALRGPRRRRQRGLSSFTSPASFELSPSSLSSLYPSRPLSSHSPLSSVSSLPSLSSPSPLSSLSLSSRVPRSHAGPIRRHG
Hh    REEELWD------------------------------------------------------------------------------------------
Nc    KEEELWK------------------------------------------------------------------------------------------
Sn    PLFKSPPNNSSSSSNSRSNSDSSSNSNSNSSSKSSSRNSSGDATATCTTAAAGAAARAAGTTATGRGAGAEGESSAGVFRCEKVPWLRGKSAACEAESAAAAADRLA
Cv    VKKEWEDS-----------------------------------------------------------------------------------------
Vb    -------------------------------------------------------------------------------------------------

Dd    -------------------------QFP---FIEDYLFWLKSLIMKGLN------------------ISNIQSSTP-LLYLRKHNNSISF-KNIEKQKDSTANASC
Dp    -------------------------TYP---FIEDYDLWLNCIVNGNLK------------------CSNIQSNTP-LLFLRKHGNSVSS-KNIVKQRESTAQAST
Pp    -------------------------TFP---HAEDIALWLRIIATTPYR------------------IANLPMDAP-LLRLRKHSANVSK-THALAQRDSTRRASQ
Tg    EVEVEKADRTESRRRARYRGAEEVFMYP-DEEAEDLWLWLSLPF----------------------CLHITSLPRILTFIRRDA-SRKSETVLTPMRQSACQAVA
Hh    --------------------FMYP-AEEAEDLWLWLSLPF----------------------CLHITSLPRILTFIRRDA-SRKSETVVTPMRHSACQAVA
Nc    --------------------WIYPAREEAEDLWLWLSLPF----------------------DLHVTNLPHILTFIRRDP-TRKSETALEPMRRSAGRAVC
Sn    ------------------NGSCCCWSFVS--EAAEDMWLWTSLPL----------------------CWEITNLPVALTCLERGSASSKSKTHAQEMRLSAARAVT
Cv    ----------------------EGYPEDVPAEDHALWLRILAQKTKTTDRREGGEGSASDESFEFKWEAANMGDTLTRLGRRPDSLSL-KKAKELRRSSQEQVA
Vb    ----------------------LVYPTDEAAEDHARWLGLPP----------------------GVVVCNVGDICVRLRRHGRQRSA-THHKELQESSIRAVH

Dd    YYLNILFKRFNIDSEIIQN--------------------------------------------------------------------------------
Dp    FYLEKLFEK---NGIKTSS--------------------------------------------------------------------------------
Pp    HYLCALLGREVKRE-------------------------------------------------------------------------------------
Tg    EWISRTLLRPLGRERPRENNEEEDSDAF------------------------------------------------------------------- MS
Hh    EWISRTLLRPRSRDRLRENNEDEDSDAF-----------------------------------------------------------------------
Nc    AWIVRCTQSRLGDTRQGEAANEEDVS-------------------------------------------------------------------------
```

```
Sn  RWIYCHVPLPSSSQQQQQQQSCCSSTSNTSRSSGTSRSWPMKETRSSTNSTTCRSSTSGKSSSSSSAGPLLCPLLLAGVVTQQQQQGQEDQQKQQFQQQRQQS
Cv  LFMEKASSTSNTSRSSGTS-----------------------------------------------------------------------------------
Vb  RYMQRHVDTSVTVEEVQVL-----------------------------------------------------------------------------------


Dd  ------------NSSLS-MKEIIQFFQLLSPSSLSKINN--ISIELFEFA-FKYLELIEKSCTKQQP-NYS----NSIK-DAANEKMGELVSLCL-SNYPNNQKS-
Dp  ------------SSNAS-LDSEVVRFLIISPVKYIAENNDTLSEDTFEKS-IEYLEAIESSFLNQENQKYS----TNIK-DATNERIGEIASLCL-SKFPNNQKS-
Pp  ----------------------VVDGLVDAQSSAKVEAQIVELLEGLEQALC-AKP-HFTEAEKKQIQTEAT-KRIGELTTLAM-SRDPGNNYQL
Tg  ------------------LREVKRA-ETQARSEKRLPRGEAQRSACVARKTTAAVGGEFRRTEAVGESGEAAERTNV----------------------
Hh  ------------------LREVKRA-EMQTRNEKRLPREEAQRSTCFALKTKAAEGGEYIRTEAVGESGEAAKRKNVNRC------------------------
Nc  ----------------------LPTDETSVECRNRVAERLPNAEPETRSFVEFDSTNQ--SEPRRQERKKTAERTRA------------------------
Sn  QQQQQSPQQLQQQPAQQQLQQHSAQQQQQAQQVHCG-GGMVDIETELPRANSPKGTERSQEEQQRRAHQQQTLQQSQEAPQQQEDGKQQQHQAKYSQQQQQRRG
Cv  --SSSCMGLGPVDPELVALLWDTGKAGGQREKGAAAGGGRGRPIREAIQLLEKMLKFFENMLSPERPGRGEKEKHRNGNAEEHLRGHSDTCLAASLDRCGLSPWSF
Vb  --LWKQRRIDGSTTGRRIDDLLKALENHFVQLIDDLAGQA----REAPQQQQMGSDWWADGRTAALKQHIHCATERIRGEMCLVESSFASEGDEEERRAGALRFIK


Dd  -----------------------------------------------------------------------------------------------------
Dp  -----------------------------------------------------------------------------------------------------
Pp  -----------------------------------------------------------------------------------------------------
Tg  --------------------------------------RNAGEFETRQLQRGEEDKDKQDEDGGNGEGAATQGEGRRGEDETDLREELANAMTA EST
Hh  --------------------------------------RNAGEFENREVERGEQDKDAEDGGNDEGAATQGEGRRGEDEMDLREELAHAMTT
Nc  ------------------------------EN-GRGEGHSKSKTEEE--AQD-AGGGEEERIQEEER-GEEGTRLSEEVKKKVKN
Sn  NWQQQQQEVLQCQHQQEENVRQRGQGEQQVQQQQKDCIQQQLLQKTCCPPSTTQKPKGKNEFE-DVVV--AGVPLLHLVQ-CMR-GDRVTETVEEAAAYE--
Cv  R----------------------------------------------------------------------------------------------------
Vb  -----------------------------------------------------------------------------------------------------


Dd  ----------------------------------------------------------SLLWEKWLSRNPTSQ-LLSLLSNLNVKSSTTIINNNINN
Dp  ----------------------------------------------------------EAMWKKWISRNPTSQ-LLSLLSNL--KTNET---------
Pp  ----------------------------------------------------------FAKWVQKRS--PKQQ-LFSLISNF--GTSSSTPNSTGET
Tg  AS-------------PQANGDLHVPF-EESPLLPSVVA-----------CLRGYRTPATTA-EGEAAFRIKRCLEALESFFCVAVDKPQSLFGQEDVSASPA EST
Hh  S-------------RQANKGLRVPF-EESPLLRNVVA-----------CLRGYRTPAT-AEEGEAAFRILE---ALESFFCVAVDKPQSLSGQEDVVASPA
Nc  ATSSSRESGATMQGFSQEKRGRSVGLGEDSRLLPRVVA-----------CLRGHRTPAT-AEEGEEAFKVLK---ALEDFFCAAAEQFRRRSRHRENL----
Sn  ---------------------------RLLQDVTAFFLQLTQFFYSPRQGCICCASSAAAGERAALSEQQQVQQRLRAWQQPELQRRNVQQQTVQPKQQ
Cv  -----------------------------------------------------------------------------------------------------
Vb  -----------------------------------------------------------------------------------------------------


Dd  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSILNFISGINSNKINTPK--SNNNKFKEN (467)
Dp  --------------NNNNTNKTSDSNASLLNFLSNSST—NNKPTINSNNNNFKDN (439)
Pp  PQ (343)
Tg  CDQGNRLTVGV-CSCPCGQGDTDSGCSP------------------------------------------------------------EALKKFF EST
Hh  CERGKPRTVDV-CSCPCGQGDADSGCSP------------------------------------------------------------EALKKFF
Nc  ---FFRRPVKETCPCPCGEGGTAVGCSA------------------------------------------------------------KALKKTF
Sn  QTHPRWLHHEQKQQQQAKTHLWDAQQQQQEVTSVGTLQQRRQEAGEVQQQARAEACRTSAADCLCRSCCCTAAHTWENESSKRKGKNCFYDTIPAEAFEKFF
Cv  ------------------------------------------------------------------------------------------------KDI
Vb  -----------------------------------------------------------------------------------------------------


Tg  KREAARMRGELATRRLGECGASDLFLLWLAKDKPVETSPEILAGLL* EST (1801)
Hh  TQEAARKRGELATRRLGECGTADLFLSWLAKDKPVETSPEILAGLL* (1918)
Nc  QHEAARLRGELATRRLGAFGAGDLFLSWLANDNPAQTSPDLLRSLM* (1803)
Sn  VSQATQLQGDVALRRLATSGEPDLFLRWL--EST--TCG-ALQLDR* (2317)
Cv  DIRGTLSLEDIAKGNLTS-GR-DFAQ-WLSSSSD---AKERLKQLLSGHN* (880)
Vb  ----------LSQGNME--GSLPLLLEWLKNPS----SQQQLKRLTAGS* (723)
```

α2-FucT domain:

```
Nc  (1)  MALFLSRACLTPPKTPRNVFFPWQVFSVHRTGAFLERRRIALKVPVNLTKFEIQSYLEKIYGARVLKVSTLIVVPRRRRDIFGPRPSMRYYRVGSTFKKAIV
Nc       TLEDGVPDAVKMLRSSIELAKNPDITKHNLTYSGRVREFKPPSQAQRWEMGESKYAWRLPLPNLLAGDDLQLNPALRIDESTTQMFPDFSKPFFHNAHFKFS
Nc       WKPEEVPFQPTVKLDITPWRRRVERLQGQPPPSLGDAETVALPKFQTPSNAHASPPSGRNG


Dd  (468) -----------GIRIICFSKDRAFQLKEYLRTFFKYLKN-----------------------------------------
Dp  (440) -----------GIKIICFSKDRAFQLREHIRTFFKYLSE-----------------------------------------
Pp  (344) -----------KLSIICFSKDRAFQLRELLRSISLFLSP-----------------------------------------
Tg  (1)     MDQPSRAPSILVLVFSKDRAAQLFLCLHSFFSSYQSFQASSSPPLSSTSPLSSAVLISSAPRASREG--LTTAAPLSPSEDASSSVFVSSLSPSPPS
Hh  (1)     MDQPSRAPSILVLVFSKDRAAQLFLCLHSFFSSCGFFQASSSPPLSSTS----GVLISSAPRASREGPSPSTAAPPSPSDDASPSVLVSSLSPSPPS
Nc  (266)  STMAESPQASPSILVLVFSKDRAAQLFLSFFSACGFFRSPHTPPVS---------------PGESQASRPQQRAPSPSRDVGLSVSSPPSAASPPS
Sn# (1)     CRLQQAGFSVLVLVFSKDRPSQLFLSLHSFLHALQVLDNHQPLQHQQQQQRQQQQHHHQAQQQQQQKPDEVQQQQQQQKQEEVQRQQQQAQQQPVQQ
Cv  (1)     MRRVKVLIPSKDRPLQLLACLLSFCRSVKG--------------------------------------------------
Vb  (1)     MSAAVVIFSKDRPLQLCGCLSSLTTCIRD---------------------------------------------------


Dd  -------------------------------------------DDNGNDKFEIIVDVLFTYSN-----------------------------------------
Dp  ----------------------------------------TDSSFKIEFSVLYTYSN-----------------------------------------
Pp  ------------------------------------HIETSISVLYKATNPESTQNPTTDPPANPPANPPANPPANPPTNPFAIPPTDSLTNLLI
Tg  SSSSSAPAPQNASSSLPSFLPSEDQSFPHVRGASSAPAP---------VEVYVQVIYRASS-----------------------------------------
Hh  SSSPPAPSPQNSFSSLPSSLPSEDQSFADVRGASSAPAP---------VEVYVQVIYRASS-----------------------------------------
Nc  SSPLLSSSSLSSPRSSPVPSPALCSSESRGSGPHAL-----------VEVDVQVIYRASS-----------------------------------------
Sn  QQAQQQQVQQQQVQQQQAQQQQVQQQQQVRQETESGEEHVEFVVGECVALHIHIMYKASS-----------------------------------------
Cv  ----------------------------------------------DADKNLRVSVSVLWKASD-----------------------------------------
Vb  ---------------------------------------SSDAQIHVIYKASD-----------------------------------------
```

```
Pp  NPPANRPTNPPAKTPTNPPAKTPTNPPTDPPTNPPTNPSSDPPTNPPTNPHTKT

Dd  EKFKNSYQLVIESFPQVNFIKEENFTDQLIN-----------------------------------------------------------
Dp  LKFKSSYQAVEKEFPNVKFIKEENFTEQLIQ-----------------------------------------------------------
Pp  HNFDQSYAKVKSLFPQVNFVPETNFSSQVKD-----------------------------------------------------------
Tg  PDFEASYKLVELLLMRLLAA-PRPTGAEMGSVKLAADATRGANSSTSTLKRSRHSERPFSTSSLAFSSEASCRPLSVSQFSEHPASAPRSSASSPSPLAVSSPPLY
Hh  PEFEASYKLVESLLLCLLAATPRKTCAEMESAKPAAEASRGANSSTSTQARSRHSQSTLSASSLAFSSESPCGPLSVSQFADHPPSVPRSCPSSPSALAVSSPPLY
Nc  PEFQTSYRLVEALLSHLLEATSA---------------------------------TQKDRGVSQDPPTRRFDSNASLSCSLSRVSPSRPSASVL------
Sn  SAFAASYKLVEALLMRLLAAHSFACLLQRLVLQQCCSCSLPAATRGQQQQQVRQETESGEEHVEFVVGECVALHIHIMYKASSSAFAASYKRVEKLLSLLLLDAEK
Cv  SNFQEAYNRLPNLLSSVLSAL-PWIDLLWVNEDGEGEGAPSTFGDGLQK-------------------------------------------
Vb  ASFAAAYARLAER-------WTDARVHWSREAPGCAISSALEGLE-------------------------------------------

Tg  RLNLLAFHSERDTESSAETRRERVLTRSECSGTHKGEEASLSDCSEKASCDENEGSTTSPVSPFQSLLLRCLTSVHTPGGNSSSSRWQDVFSSPSASLTKKDGWRT
Hh  RLNLLAFHSERDTESSGDTQRDRAHTRSECSDTQRGAEASLSKCSEKASCQETAGNRTSVSPFQSLLLRCLTSVHTAGRHTSSSRWQTVFSTPSAS-TEKDGWRT
Sn  RRAKQPYGVRTLQFHRESGDLTSSSSSSKEHSSGSRSRSSRRRRSSSRSRGREESSGIKTEGSSSRLLAKNTGGKAHSFACLLQRLVLQQCCSCSL----------

Dd  ------------------------------------------------------------LNKLEYVMFSVDDI-LYYNEFNLK
Dp  -----------------------------------------------------EAQKSNTMEFTLFSVDDI-LYFHQFNIS
Pp  ------------------------------------------------------------FVTRAN--AYVMFCVDDV-LFYNPAPVK
Tg  SCPSLSPCKSSVSSSRTASESSQVTPLEPSSSAA-PGDELKSKRNVPQGDHRGNKELGGSATNRREAKRPQGSSDATREAATRCMYTHVLLMVDDCVWLYPPSEFF
Hh  SFPSLSPFKSSVSSSRTASESSQFTPLEPSSSAASCGDGLERKCDVLEGAQRGNKAPGGSATNWREAKRPQRSSDTTREPGTRCMYTHVLLMVDDCVWLYPPSECF
Nc  --ESLSEPSWDKLANPSASEEPPNVSLE----GKKAFANEMTAQSACREVGEKEKNRTGAATPEHSNTGFQP--DAVINSS--CMYTHVLLMVDDCVWLYPPSQLH
Sn  SPAASSTSSGRTTNSGRRRRSSSSDR-GFNCSSGRRSVSSSNRGDGSSSRVLAKNTGGKATAAAAFPYNQQRQGRRTAGGGGGRDYTHVLLFVDDCVWLYPHAAAE
Cv  ---------------------------------------ILEETAKAQSGGESLADDGEEEDFVLFAVDDC-LFYSDLRLE
Vb  ----------------------------------------------------------ASHVLFTVDDA-LFLRPLAPS

Dd  EY-CLSLNSEPLAL---------------------------------------------------------------------------
Dp  NY-CKILKKETSSL---------------------------------------------------------------------------
Pp  E--GLGVLREDPSVF---------------------------------------------------------------------------
Tg  ER-SKEVDLT--GVSEKARLADPELSKNERFSSSSTHLSSSSSSSISSSPSVSS------AFSRSSRSSLPGFPSSALPEMCSAVSSPSTASSLSSCVSSSRPPSEA
Hh  ER-SKEVDLT--GVSD-ARSTEPGLSKTERFSSSSTQLSSSSSSLSSSSSVSSSSSVSSAFSRSSLSSPPGVSSSALPETCSAVSPPSAASCLSSSVSSLCPPSEA
Nc  S--GETLDIARVSGLQE---FDTPIPK-----YETGPSESALSSL--SPSSSS-------HSPSSSSCSPSFCPNLAPSFASSLPSGSSFRP-SSTSPPFPPVSEA
Sn  NR--QLLND---GMPQPLQEEQPSHRQRKRKKKKREKKQQLPSQPQQPVLQQEPVLPQEPVLQHEPVQQQQPVLQQQPVLQQQPVLQQEALALQLQAAIGPVLPLQR
Cv  QS-CQ-ILSERPEVF---------------------------------------------------------------------------
Vb  LSRCVALLEENDALL---------------------------------------------------------------------------

Dd  ------------------------------------------------------------------------------------------------
Dp  ------------------------------------------------------------------------------------------------
Pp  ------------------------------------------------------------------------------------------------
Tg  SCSSRASPHPAFFTLALAA----------------------------------------------RLLDVCPSLLHVSPRLHFGVNYSQPSDSVIPL
Hh  SCPSRASPHPAFFTLAHAA----------------------------------------------GLLDFCPSLLHISPRLHFGVNYSQPSDSVLPL
Nc  SRLSGASGPPVVSTLAHAV----------------------------------------------HLLDSCPALLHVSPRLHFGLDYSQPSDSVLAL
Sn  QLRLGQPQEQQAQQQPSCGQGFVAVKEKKTRQQDGGLQQTDGKNCPHRLRWRQQQLLRCSRGNTDSHLLLLHALQLLQQRADVLLVTPRLHLGVEFSQPSAFLLPL
Cv  ------------------------------------------------------------------------------------------------
Vb  ------------------------------------------------------------------------------------------------

Dd  ------------------------------------------------------------------------------------------------
Dp  ------------------------------------------------------------------------------------------------
Pp  ------------------------------------------------------------------------------------------------
Tg  ARAAVLPPLS-LPRAVSSATA----RDTGSEEAPAAQDSQETPSDAIFSVGT-GKTAVEAG-----------------------------------
Hh  SRAAVLSPLS-LLRLESSATA----RDTGSEEATAAKASQEKPGDAISSVGT-GTTTVEAGG----------------------------------
Nc  PSAAVLSPLSPLFPPWRLLPASPVERDTVRGETPARENQPRKRRRDAI-SGTPGQGSSETEKFRNGEAEAGQKNKEEEEGEGDREEKEEEGEEGEEGEEGEGDREL
Sn  LPRCLLPYFPSPTPS-SCSS----------------------------------------------------------------------------
Cv  ------------------------------------------------------------------------------------------------
Vb  ------------------------------------------------------------------------------------------------

Nc  EKEEEGEEGEEGEEGEGDREEKEEEGEEGEGGRNGVGGSDGGREGGARNGEETRSGRERK

Dd  --------------------------------------------------------GFYMKLNKNITYCHTCNQDITIPLNSNTISR-------TENNFKYLKWNRN-DNDCKK
Dp  --------------------------------------------------------GFYMKMTSNITYCHTADEEITIPLSLSELSYE------ETGGEKYLKWNRS-ESDCKR
Pp  --------------------------------------------------------GYYLKLYPGLNFCHTANSPMRAPQNFAKT----------TG-NGILKWEMN-DKYVSH
Tg  EQDSQEEGER-EGERKDDDEGERQEGNDAELEKGGQRRGTTLSAQPKPSAFGERLASYLDSRWHMLLPQKG-IQEETD-----------ASQSFSFLAVARKTSFG
Hh  EHVSQEGKRGEGERKDDDEGEGEGNEAELQEGGQRSRTATSAQPKPSAFRERLASCMDSRWHTLLPQNL-IKEESE----------TSQNFSFLAVARKTSFG
Nc  EKNQERDRDELTKGYKLEREAEFHRGH-TVHTDEAQRTKRWRRSTRDRENPSDRVSAMKKQKLQR--TQCGKMSAEDT-----------GSCSIEADCS-RKRCQG
Sn  ----------------------------------------------------------------------------NVLQSFFLLLKREGRLK
Cv  --------------------------------------------SVQFRLNPRVTYSHPRSRFGYRPMLYPLQMDKPSSSKET---TPFLLYDR--SKTSFA
Vb  --------------------------------------------SVCLRLNPRVEYSHPAGSFFPPPVLHWTRERDALLLRRLREGDCSGGCGGDAGGTGAD

Dd  DWNYPWDLCSTIYRCNDIDSIINGIVKYY-GIRNGINHPNRFEFNGNRP-IIQKQIYQNK----------------------------------
Dp  DWNYPFDLCSTIYKTSDVDSILNGIVKYY-GIRNGINHPNRLEFNGNRA-IIQKQCYQNK----------------------------------
Pp  DWNYPFDLCATIYRKSDMLKLLGCLENMF--GNSGYSHPNTFEANGNRPQILN--CFSSS----------------------------------
Tg  DFGMALDLSCTLYRSADLLCLLQTLLKVE---PDSLNHPNRLEVAGNRH-FEKCTSSNGA----------------------------------
Hh  DFAMALDLSCTLFRYADLLRLLQTLLKVE---PDSLNHPNRLEVAGNRH-FERCTSSNGA----------------------------------
Nc  DFEMALDLSCTLYRQADLVLLLQTLLKVE---PDSLCHPNRLEVAGNRH-FERSRLPSGS----------------------------------
Sn  DYGMPFDVSCSIYRYDDIMNTCQELQRDE---PAALTHPNLLEVSGCRY-MARAFLAAAGDTGATRGSQEQQQQQASQQQKQQEERESQRQQLRGQAERDSSRE--
Cv  DWSMAFDLSCSLYRRNETAYLIQTLRKTK--GESAVGHPNSLEASANVW-IFNETPKGRG----------------------------------
Vb  EWSMPFDLSGSIYRTSDLRRLFTCIEESPVLGREWLDHPNRLEAAGNRA-IQELM----------------------------------
```

```
Dd  ------------PYCLCLSDHYSPMSVVTINRVQDVYDNPIY-DQTLSLDD-------------------------------------------------------------
Dp  ------------PYCLC--PTTPIMSVVTINRVQDVYDNPIY-DQTLSLED-------------------------------------------------------------
Pp  ------------TFGIC--PSKPVLSVLTVNRVQDIYKNRVYEDQELSVEE-------------------------------------------------------------
Tg  -----CRLSLLRPLVAF--PVHPRAVVLTINRVQSLYLNPVYKTASARSPPK-RLREAEDEKEGKESADEESAEESNVEEQS-----EEGDEEKKGEENSWSCE--
Hh  -----CLLSLLRPLVAF--PLHPRAVVLTINRVQSLYLNPVYTTPSARGPPK-RLRESEDEKEGKESAEEETAQEEKAEEGNAEGHVEEGDE-KEGKEDSWSCE--
Nc  -----CSLSALRPLIAF--PLHPRAVVLTINRVQRVYSNPVYLAPSPTTRSKVRGGNARAGDAWSCASLDVYFRRGVEMWRPRLSQWARAHESERATEKQEPSE--
Sn  VQEQQLPLATLRPYVAF--PQLPVAVILTINRVQDIYKNPVFSPTRTALQPQQEQQQQQERKQQWQLQRTKKKQQRRRPEKQQPQQQEKTQQQQQERRRKKARR--
Cv  -EGADSDSPWEKTINAC--PSSPVCSVLTINRVQEVYKNPVFEGTQLAPESQDNSPSALLDLFNFLFPPSPDTFPDD--------------------------------
Vb  -SGLDLNAT--RPFTAA--PTQPLLTLITANRVQDVYSNAIYADAPQWDIRSLLATFHQIDRTHTSH-----------------------------------------


Dd  ------------------------------------------------------------------------------------------------LDQLLYSNKSLNDEKYK
Dp  ------------------------------------------------------------------------------------------------LDILLQQNDQLDDLEYY
Pp  ------------------------------------------------------------------------------------------------LDKFLWDEREIDVTHYQ
Tg  ALDQYFRSGLARWLSTLAGTGGEKGEKETGEDNGGRSSIGTAFPCHVSPCALPLGW---------------------------------------FFPFLLPTAPQFPFDTN
Hh  ALDQYFRSGLARWLSTLAGTGGEKGEKETEEDNDGGSSIKTAFARRVSPCALPLEW---------------------------------------FFPFLLPTAPQFPFDTK
Nc  ALEQYFRSGVARWLSTLSGSGSEKGEKET----------------------------------------------------------------LFPFSLPTAPRFPYEME
Sn  AVPLESQRQKIKMKHKTQQRGEKHLS DRAVVDWSCRSLDKFFTSRAQAWLEQQRQRQQKMQHQPGTQQRQQERERAGAALASFSQVQQFFPFGLPRQCCMSCCSC
Cv  -----------------------------------------------------------------------------------------------PLEVLKERGVMDLKAYW
Vb  -----------------------------------------------------------------------------------------------TDFLTAVDGFFAAPFYT


Dd  ENSL----SLNFKSVHIGELFIS*(768)
Dp  NN------SKSYKSVHIGNLYVLKK*(736)
Pp  KN--------QFTSVHIGDFVLKPT*(727
Tg  SCF--------FPSVHLPPPSLAALLALCRSQGSSASTELKSTSASLPQQSSFP----------------------------
Hh  SCF--------FPSVHLPPPSLASPLAFCRSPGSSALTELNSSSALLPQQSSFPLPSFLPSSFLPSSSFSSSSLPSSFLPSSF
Nc  GCC--------FPSVHVPPPPLLSPVCTIQASPSPGLADAGALPEAPSVSFAPAAIRSS-----------------------  (1149)
Sn  NCSSGCCCFNSFAAVHMAPPALPSPAAAAALRHSTSDSSGSDSSRSDSSGCDSSGSDSSRSDSSGSDSSGSHSSRSDS
Cv  PQ------NAYFDSVHIAPVPPSGTSHSLTPALPLLR----------------------------------------------  (382)
Vb  ST--------YWDCLHTPPPPLSASAAVPR----------------------------------------------------  (329


Tg  ----------------------------------------PSPSLSSSLPSS-LSSSFSSS----------ISASAFSTSAFSS (1061)
Hh  SSSFSSSSSLPSSFLPSSFSSSFSSSSLPSSFLPSSSFSSSFSSSSLPSSSLPASSLPASAFSTSAFSS (1162)
Sn  SRSDSSRSDRSRSDRSNRRSSSSINRRNDIRRDSGSSWKCSSSSKLTRNNNRRSSDESTNSRERRKRRTGELRRFRSRRRGGCSSSLDSRFSCHSSVSLSAGGHFSS


Sn  SSSRSGRIPRNNSSKGTSRRRREKSGRNRCGSSEVANSSRSRISREKTCTSSSRRKDRSRANNSSNSYYRKSSSRRTPVSGSHSRRTGVRLHCCPASSSSPSSSAA


Sn  AAGTAAAMPAAAVVGVGSRQQSRPSRVSSAVSSSSSSSPSSPTSSSSPPLSRVCWSLPRSLCSS (1377)



# = numbering start is arbitrarily assigned
MS = sequence supported by mass spectrometry at ToxoDB
EST = sequence supported by EST sequences at ToxoDB
```

Origin of Sequences:

Amoebozoa–
Dd: Dictyostelium discoideum, GI:60467473
Dp: Dictyostelium purpureum, jgi|Dicpu1|159362|GID1.0049988
Pp: Physarum polycephalum, (gene locus 131, Phypoly_transcript_04421:Physarum database (*http://genome.fli-leibniz.de/ggl_blast/blast_ggl.pl*)

Apicomplexa–
Tg: Toxoplasma gondii, From ToxoDB 7.3: TGGT1_260650
Hh: Hammondia hammondi, From ToxoDB 7.3: HHA_260650
Nc: Neospora caninum, From ToxoDB 7.3: contig FR823389
Sc: Sarcocystis neurona, From ToxoDB 7.3: Contig sneu_scaffold00020

Chromerida–
Cv: Chromera velia, From CryptoDB, CVEL_18515
Vb: Vitrella brassicaformis, From CryptoDB Vbra_12849