

Advances in Novel Animal Vitamin C Biosynthesis Pathways and the Role of Prokaryote-Based Inferences to Understand Their Origin

Pedro Duque, Cristina P. Vieira, Jorge Vieira

Supplementary Figures

Figure S1 – Complete AO dataset MUSCLE alignment file. The alignment presented was the one used to perform the phylogenetic analyses discussed in the manuscript. The relevant amino acid regions used to extrapolate putative AO functions are tagged at the top of the respective alignment positions, and delimited by the black colour background.

	1	10	20	30	40	50	60	70	80	90
<i>Mycobacterium tuberculosis</i> H37Rv NP_216287.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Streptomyces albiflaviniger</i> MBI0379321.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	MAV
<i>Pseudomonas aeruginosa</i> PAO1 NP_254014.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	MVGNP
<i>Myxococcales bacterium</i> MCO4763851.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Bacteroidetes bacterium</i> MBS1617249.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Flavobacteriales bacterium</i> HIA12326.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Moorena</i> sp. SI04G3 NE081287.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Calothrix parasitica</i> NIES 267 BAY85903.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Asgard group archaeon</i> MCP8715902.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Halobacteriales archaeon</i> PSP48796.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Euglena gracilis</i> BA094255.2	MARNKVMRMLF-----	-----	RAVA-----	-----	-----	-----	-----	PTTAGICLWQVCNQFSS-----	-----	VEVWC
<i>Trypanosoma brucei brucei</i> TREU927 XP_844937.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	M
<i>Leishmania major strain Friedlin</i> XP_001682394.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	M
<i>Mus musculus</i> NP_848862.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Rattus norvegicus</i> EDL85374.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Xenopus laevis</i> OCT81467.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Ciona intestinalis</i> XP_002122023.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	MTS
<i>Falco cherrug</i> XP_005443605.1	-----	MGPAKGGAAD-----	-----	TPGR-----	-----	-----	-----	ALSASLEKFGQGAKVSLW-----	-----	LEAEA
<i>Rosa chinensis</i> XP_024182616.1	MQRALTVKRSL-QSL-----	-----	QRIANSTTK-----	-----	NPLSSGRAFC-----	-----	NAPSPSPSPAPSSASSASASASEVRKYMGYTA-----	-----	-----	LVLFC
<i>Rosa chinensis</i> XP_024165881.1	-----	-----	-----	-----	-----	-----	-----	MYRSWWLLGAIH-----	-----	LLVS
<i>Prunus mume</i> XP_008240731.1	-----	-----	-----	-----	-----	-----	-----	SSASELRKYLGTYA-----	-----	LVLFS
<i>Prunus mume</i> XP_008220057.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Fragaria vesca subsp vesca</i> XP_004291192.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Fragaria vesca subsp vesca</i> XP_004303609.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Arabidopsis thaliana</i> NP_190376.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Arabidopsis thaliana</i> NP_182199.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Arabidopsis thaliana</i> NP_564393.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Arabidopsis thaliana</i> NP_182198.2	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Arabidopsis thaliana</i> NP_182197.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Arabidopsis thaliana</i> NP_200460.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Quercus suber</i> XP_023896433.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Quercus suber</i> XP_023896434.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Quercus suber</i> XP_023900169.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Quercus suber</i> XP_023914113.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Aspergillus chevalieri</i> XP_043131130.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Candida dubliniensis</i> CD36 XP_002422445.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Saccharomyces cerevisiae</i> S288C NP_013624.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Thermothielavioides terrestris</i> XP_003651626.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	M
<i>Colletotrichum lupini</i> XP_049145240.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Homo sapiens</i> DHCR24 NP_055577.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>Bombyx mori</i> DHCR24 XP_004926865.1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

FAD-Binding motif
(X[GR]X[GS][HLK][SG]) region

	91	100	110	120	130	140	150	160	170	180	190	200	210	215
NP_216287.1	-----	MSPI-----	-----	WSNW-----	-----	PGEQVCA-PSAIVRPTSEAE	LADVIAQA	AKRGERVRAV	VGSGHST	FTDI-AC	TD-----	GVMIDMTGLQ	RVLDVDQ	-----
MBI0379321.1	RTVKTVRSSV-----	-----	-----	WRNW-----	-----	AGNVMAR-PT	TRIVAPSTQ	ELAEVRRATAE	GLKVAVGT	-GHSFT	TA-AATD	-----	GLLI	IRPDRMAGVRGLDR-----
NP_254014.1	ALAELLRAPLIP-----	-----	-----	WRNW-----	-----	SGAQSC	L-PLAREAPK	LDLDELVI	IRQA-----	-----	SGRIRPVGS	-GHSF	SAL-VPTD	-----
MC04763851.1	-----	MSQS-----	-----	IANW-----	-----	SRTHHCT-PA	RLERPRSE	EDTIALVKNA	ADNGD	TIRVIGA	-LHSNS	DC-AMSD	-----	GIVVSLDDMQVLEVD
MBS1617249.1	-MAIISQRAYR-----	-----	-----	FRNW-----	-----	AGNQCT-PS	HYQPTEAE	LIEVI-----	-----	KTHSKVRM	VGT-GHS	WSAL-CH	TD-----	EALINLDKYNR
HIA12326.1	-----	MNAPHK-----	-----	FVNW-----	-----	AATESN-PE	KFLKPSNQ	EELVE	INSAINANS	KIRVVG	A-GHS	WSSV-AC	TD-----	HWLITLDNYCNVLS
NEO81287.1	-----	MNQNK-----	-----	FINW-----	-----	SQQYCT-PT	SVHPASTE	EARDIVC	VMQQGT	KIRVFGS	-GHS	PSDI-AM	SN-----	QDLVCS
BAY85903.1	-----	MNQEJ-----	-----	FINW-----	-----	SGEYRCT-PS	KVHHPR	TDVREIV	CQAIKRGS	KIRVFGS	-GHS	PSDI-AM	SD-----	EELVCLDGLNTIKD
MCP8715902.1	MSSGIPESLKPF	LT	TKLV-----	-----	HS	TW-----	AGTF	LCK-PQAIF	QPRNEKE	IKE	LINQAR	INGKT	IMT	VGS-GHSP
PSP48796.1	-----	MLRLTDGV-----	-----	-----	-----	WENW-----	-----	SGSVRAR-PTH	VHPETE	ATLRD	VERA-----	-----	DDTL	LRVAGA-GHSP
BA094255.2	-----	RSLP-----	-----	-----	-----	VS	NW-----	SSTHSVV-AG	KYYQP	TEVEEL	KD	IVTIC	HATG	QKIRPAGA-MLSP
XP_844937.1	-----	QGETMSDGT-----	-----	-----	-----	WTNF-----	-----	ANIGKCF-PR	KHHYP	NTVEE	VSSI	KVINS	AGERC	RVVGG-GKSP
XP_001682394.1	-----	MSAHSAAAR-----	-----	-----	-----	WTNL-----	-----	AGIGSCH-PR	HHPYT	STKEV	QDAVEL	VRSQNG	KCRV	AGA-GKSP
NP_848862.1	-----	MVHGKGVQ-----	-----	-----	-----	FQNW-----	-----	AKTYGCS-PE	MYQPTS	VGVE	REV	LALAR	QNK	KVKVVG-GHSP
EDL85374.1	-----	MVHGKGVQ-----	-----	-----	-----	FQNW-----	-----	AKTYGCS-PE	VYQPTS	VGVE	REV	LALAR	QNK	KVKVVG-GHSP
NP_1821467.1	-----	MVLGRGGYK-----	-----	-----	-----	FENW-----	-----	AQTYGSS-PE	LYFQPT	CVEI	KE	ILDLAR	QSRK	RVKVVVG-GHSP
XP_002122023.1	-----	FDISQKGSGYR-----	-----	-----	-----	FYNW-----	-----	SKTYSAS-PE	LYFEP	SRDNL	ILSRA	KENK	TKV	KVVG-GLS
XP_005443605.1	-----	ITEVHQGGVK-----	-----	-----	-----	FQNW-----	-----	AKTYGSS-PE	LYFQPTS	VEE	IRE	ILDMAR	QDKR	RVKVVVG-GHSP
XP_024182616.1	GAATYYSF	FPF	PENAKH	KAQIFRYAP	LPEEL	LHTVSNW	-----	SGTHEVQ-TR	VFHQP	PETLEE	L	EKKV	KEAN	ENKYRIRPVGS
XP_024165881.1	VTAMLQSP	IQCD-GT	GGCNLFNS	-----	-----	YGVW-----	-----	NDRKNCH-V	PNVTYPT	PTENEL	LRAVAYAN	QNK	LKL	RVVVSQFSHT
XP_008240731.1	AAATYYSF	FPF	PENAKH	KAQIFRYAP	LPEEL	LHTVSNW	-----	SGTHEVQ-TR	VFHQP	PETLEE	L	EKKV	KEAN	ENKYRIRPVGS
XP_008220057.1	IQAMPPEL	PVRCN-K	NGCNLYNS	-----	-----	YGVW-----	-----	GDRKDCH-V	QNVTYPT	TEEBEL	LRAVAYAN	QNK	LKL	RVVVSQFSHT
XP_004291192.1	VAAMLQSP	VQCD-GT	GGCNLFNS	-----	-----	YGIW-----	-----	NDRKDCY-V	LNVTYPT	TENEL	LKAVAYAN	QNK	FKL	KVVSQFSHT
XP_004303609.1	GAATYYSF	FPF	PENAKH	KAQIFRYAP	LPEEL	LHTVSNW	-----	SGTHEVQ-TR	VFHQP	PETLEE	L	EKKV	KEAN	ENKYRIRPVGS
NP_190376.1	GVATYFS	FPF	PENAKH	KAQIFRYAP	LPEL	LHTVSNW	-----	SGTHEVQ-TR	NFNQ	PENLAD	EALVK	ESHEK	L	KLIRPVGS
NP_182199.1	TISTPPED	VPKCVSG	NTNCTV	TNS-----	-----	YGAF-----	-----	PDRSTCR-A	ANVAYPT	TEAE	LVSIV	VAAATK	AGR	KMRVTRYSHS
NP_564393.1	ASSTPPDD	VPKCESG	NMCTV	TNS-----	-----	YGAF-----	-----	PDRSTCE-A	AKVEY	PKTEAE	LVSIV	VAAATK	AGR	KMRVTRYSHS
NP_182198.2	AVSTPPDD	VPKCVSG	NTNCTV	TNS-----	-----	YGAF-----	-----	PDRSTCR-A	ANVAYPT	TEAE	LVSIV	VAAATK	AGR	KMRVTRYSHS
NP_182197.1	VISTPPED	VPKCVSG	NTNCTV	TNS-----	-----	LGAF-----	-----	PDRSTCR-A	ANVAYPT	TEAE	LVSIV	VAAATK	AGR	KMRVTRYSHS
NP_200460.1	ASSP	PPDD	VPNCE	FGTMC	TV	TNS-----	-----	YGAF-----	-----	PDRSTCR-A	AEVAYPT	TEAE	LVSIV	VAAATK
XP_023896433.1	VRCSPPED	HIKCSS	KNTDCT	ITNS-----	-----	YGVF-----	-----	PDRSTCR-A	AEVAYPT	TEAE	LVSIV	VAAATK	AGR	KMRVTRYSHS
XP_023896434.1	VCCIPPED	PIKCSS	DNTDCT	ITNS-----	-----	YGAF-----	-----	PDRSTCR-A	EVAFP	STEQE	LISIV	SKAS	RNKR	KMKVTRYSHS
XP_023900169.1	GVATYYSF	FPF	PENAKH	KAQIFRYAP	L	PDDLHTVSNW	-----	SGTHEVQ-TR	VFHQP	PENLQ	QLENNV	KEAH	KA	TRIRPVGS
XP_023914113.1	IYAMPQ	SPVQCS-TT	GGCTLYNS	-----	-----	YGVW-----	-----	GDRKDCH-A	PNVTYPT	TEEQE	LSRV	VAYAN	QNK	LKLKVVTKFSHT
XP_043131130.1	LLHAV	SSAYR-----	-----	-----	-----	WFNW-----	-----	QFEVTCE-S	DAYI	APED	ESAAS	FLKQ	DYARG	SLIKVGN-GHGF
XP_002422445.1	-MADIPES	LKPFV	TNKL	I-----	-----	HS	TW-----	AGTF	LCK-PQAIF	QPRN	VEEIQ	ELIK	QARL	HKKTIMTVGS
NP_013624.1	MSTIP	FRKNYV-----	-----	-----	-----	FKNW-----	-----	AGIYSAK-P	ERYFQ	PSSIDE	VVELV	KSAR	LAEK	SIVTVGS
XP_003651626.1	ASPIPE	LPPE-----	-----	-----	-----	LASL-----	-----	PAE-----	-----	PESVAE	VEH	AVT	LAR	QRRRLVTVGS
XP_049145240.1	---MPSL	RPLPL	AMLL	ASTAV	TAAP-----	-----	-----	YNTFDG	PGFP	PACND	VAAVHN	ATS	VS	DIQSLVKDAIASGK
NP_055577.1	L	FL	LPLSL	IFDI	YYY-----	-----	-----	VRAW-----	-----	VVE	KLSS-AP	RLHEQ	-VRD	IQKQVREWKQGS
XP_004926865.1	L	ALLP	SA	AWKL	SI-----	-----	-----	IRNY-----	-----	VVE	FMNS-AP	KMHDDK	-VKE	VRQRIKEWLSGDKS

NP_216287.1 -----PTGLVTVEGGAKLRALGPQ--LAQRRLGLENGQVDPSQISGTA-TATATHGTGVRF-G-NLSAIRSVRLRLVTAGG-----EVLSSLSEGGD-----YLAARVSLGALGVISQVTLTQVTFPLL
 MBT0379321.1 -----EAGTVTVAAAGTSRLQLNET--LSAHSLSLNVMDIMEQTAVAGA-TSTGTHGTGRDS-A-SIAAQIKGLELVTADG-----SVLRCSAENRNEPFAAARIGLALGVVSEITFGVPEEFL
 NP_254014.1 -----VTLQAEFGGGTQMSQMGAP-LKEIGQALVNMADIDYQTLGALA-IATSTHGTGVGF-G-SYSAQVRGLQLVTASG-----EVLECDAKRNVFVDFAARVSLGALGVTVRVLQNRRAAYL
 MC04763851.1 -----DGLRVVQGPGLRLARLNEE--LAAGVALPITVSGVSGS IAGV--MSTGTHGSSHLTG-G-NIPSAVVGMRVLVTGKG-----DVLVL--DEHDELLPAARVALGALGVIETVLSVQPAFL
 MBS1617249.1 -----EQKLTIVQAGIKLQWLQND--LDETEGLALINLGSIAQSIAGA-ISTATHGSGIYQF-R-ILASQVQFRTLQADG-----TPLTLDNRADAKETFDLAIISLGAALGVSEITTHICEAPFL
 LIA212326.1 -----EAKQITVEAGRLKELNVI--LKDGLALSMLNLSIEQS IAGA-ISTGTHGTGIKF-S-ILSTQIALTLVNVNGK-----ELVSLDHEKDADIKAAALVSLGALGIIETVTVQICEAPFL
 NE081287.1 -----DSQTVVTQAGLTKDLQNS--LAKHRLALPNLGSISAQTSIGA-MATATHGTGLNY-G-VLSTIIQEITLVNKG-----EVKIKSDENSQFLNAAKCHLGSGLVLTLEYKLVKTFEFDL
 BA859503.1 -----ENQTVIVEAGVTIYELSQN-LARYGLALPNLGSISAQTSIGA-MATGTHGTGLGY-G-TMPTIIIEEITLVNGKG-----EIKVSNSESNQFFNAVKCHLGSGLGILTEYKLVKTESFDL
 MCP8715902.1 DTPPEVKFVLTIVEAGTRVFELNEY-LKKHDLATQNLGSISDSQVAGL-ISTGTHGTSYHG--LVSQQVVSITFLNSAG-----ESITCSSIEKPEYFKAILLSLKGIGIITHVTLRTCPKYTI
 P5P48796.1 -----ETGVRVTVRAGTPLSALDLA-LADHGLMLENMGDIDATQVAGA-LSTGTHGTGTFD-F-VMPTQAAAMRLVTADG-----DVLNL--ETDDDFRFLAQVSLGTLGVSTVTLVDVDPYRL
 BA094255.1 -----KGGQTTIVQAGISQVLEE--LRKHGLTLQNFSSIQDQMGVAG-TQVAAGHTGATFL--TVEEMIVMRKLVTPSL-----GELESEQHEPELFQMAKVLGSLGCLGIVSEMTLRICPDHKL
 XP_844493.1 -----KSMTIVCEGGAISVDVEF--LSAHLMLRCVPSFVQTTVGGV-IATATHGSGIQR--SISDCVVRLLQVLDGRG-----ILHTFDATSPKELSLSACHLGMGLGVVSVTLQAEKKRLR
 XP_001682394.1 -----VAHTITCEAGAVMEVMSR-VSDHGLMVRNCGSVYRVTVGGC-IATATHSSGIQCH--CLSDYVRGLKIVDGC-----QIRTLVAGKDAAELRLAACHLGMVGVITVTLVQVPIRW
 NP_848862.1 -----EKKQTVVEAGIITDLHPQ--LDKHGLALSMLNLSIEQS IAGV--IGSGTHNTGIKH-G-ILATQVVALTLMKADG-----TVLECSSESNADVFQAAARVHLGCLGIVLTVTLQVCPQFHL
 EDL85374.1 -----EKKQTVVEAGIILADLHPQ--LDEHGLASMLNLSIAGVSDVTAGV-IGSGTHNTGIKH-G-ILATQVVALTLMTADG-----EVLECSSESNADVFQAAARVHLGCLGILVITVTLQVCPQFHL
 OCN1467.1 -----EKKQTVVEAGMLTDLPNKE--LDKRGALSMLNLSIEQSEVAAAGV-IGTGTHNTGITH-G-ILSTQVVALTLMTALG-----EILECSSESNPEIFQAAARLHGLSGLVSLVTQICRSAPFL
 XP_002122023.1 -----KRCQTVVEAGVTINELNENLPSHGLSLINLGSVSGQTVGGI-IGTGTHGSEKFG-G-SFATHVLVLMKADG-----TILRCSQQENAFISFSAACHLGLGILSVKLQCEAAFL
 XP_005443605.1 -----EKKQTVVEGGIISLDLNE--LSKHGLALANLGSIEQSEVAAAGV-IGTGTHNTGITH-G-ILSTQVVALTLLTASG-----EILECSSEINAFIQAARLHGLGCLGIVLTVFQCQVQFHL
 XP_024182616.1 -----EKKRVVQAGIRVRQQLVDG-IKDQGLTLQNFASIREQQIGGI-QVGAHGTGARLP-P-IDEQVISMKLVTPAK-----GTIEVSKKDEPLFYIARCLGGLGVVAAEVLQCVDRQEL
 XP_024165881.1 -----ANLAVTVDSGVGLRLKIDT-VEAAGLSLVAAPYWEQVSGIGL-ISTGSHGSSWMWKGGSVHDHIVGLSLIVPAKQSEGYAKIVL--EPKDQFLNAAKVSMLGLGVISKVLKLEKGFKR
 XP_008240731.1 -----EKKRVVQAGIRVRQQLVDG-IEKHEGLTLQNFASIREQQIGGI-QVGAHGTGARLP-P-IDEQVISMKLVTPAK-----GTIEVSKKDEPLFYIARCLGGLGVVAAEVLQCVDRQEL
 XP_008220057.1 -----KNLAVTVDAGIGLRLRIDT-VEAAGLSLVAAPYWEQVSGVGL-VSTGAHGSWMWKGGSVHDHIVGLSLIVPAKQSEGFYAKIVL--EPKDQFLNAAKVSMLGLGVISKVLKLEKGFKR
 XP_004291192.1 -----ANLAVTVDSGVGLRLKIDT-VEAAGLSLVAAPYWEQVSGIGL-ISTGSHGSSWMWKGGSVHDHIVGLSLIVPAKQSEGFYAKIVL--EPKDQFLNAAKVSMLGLGVISKVLKLEKGFKR
 XP_004303609.1 -----EKKRVVQAGIRVRQQLVDG-IKDQGLTLQNFASIREQQIGGI-QVGAHGTGARLP-P-IDEQVISMKLVTPAK-----GTIEVSKKDEPLFYIARCLGGLGVVAAEVLQCVDRQEL
 NP_190376.1 -----EKKRVVQAGIRVRQQLDGA-IKDQGLTLQNFASIREQQIGGI-QVGAHGTGARLP-P-IDEQVISMKLVTPAK-----GTIELSREKDEPLFYIARCLGGLGVVAAEVLQCVARHKL
 NP_182191.1 -----TAMTLTVESGVTLRQLIAE-AAKVGLDALPYAYYWMGLTVGGM-MGTGAHGSLSWKGGSVAVHDVYTERIVSPGSVNDGFAKVRVILSETTTPNEFNAAKVSMLGLGVISQVTLKQPMFKR
 NP_564393.1 -----EAKTLTVESGVTLRQLIGL-AAELELALPHAYYWMGLTVGGL-MGTGAHGSLSWKGGSVAVHDVYTERIVSPGLASDGYVKKVRLSETTDPDEFRAAKVSLGVGLVISQVTLQPMFKR
 NP_182198.1 -----TAMTMTVESGMTLRQLIVE-AAKVGLDALPYAYYWMGLTVGGM-MGTGAHGSLSWKGGSVAVHDVYTERIVSPGSVNDGFAKIRVILSETTTPNEFNAAKVSMLGLGVISQVTLQPMFKR
 NP_182197.1 -----KAMTLTVESGVTLRQLIAE-AAKVGLDALPYAYYWMGLTVGGM-MGTGAHGSLSWKGGSVAVHDVYTERIVSPGSVNDGFAKIRVILSETTTPNEFNAAKVSMLGLGVISQVTLQPMFKR
 NP_0020860.1 -----EAKTLTVESGVTLRQLINE-AAKFDLALPYAYYWMGLTVGGM-MGTGAHGSLSWKGGSVAVHDVYTERIVSPGLASDGYVKKVRLSETTDPDEFRAAKVSLGVGLVISQVTLQPMFKR
 XP_023896433.1 -----EAMTITVESGVTLRQLINE-AAKAEVLALPYAYYWMGLTIGL-LSTGAHSTLRDKGSVAVHDVYTERIVSPGLASDGYAKVREL--KDGQDFNAAKVSMLGLGVISQVTLKQPMFKR
 XP_023896434.1 -----EAMTITVESGVTLRQLINE-AAKAGLALPYAYYWMGLTIGL-LSTGAHSTLWANGSSVHDVLEVRIVTFFGA--GFARVRL--KDGQDFNAAKVSMLGLGVISQVTLKQPMFKR
 XP_023900169.1 -----GKTVRYVEAGIRVQLVDG-IKDSGLTLQNFASIREQQIGGI-QVGAHGTGARLP-P-IDEQVISMKLVTPAK-----GTIEVSKKDEPLFYIARCLGGLGVVAAEVLQCVDRQEL
 XP_023914113.1 -----DNLVTVTDAGVGLRELIDK-VEEAGLSLVAAPYWEQVSGVGL-ISTGAHGSWMWKGGSVAVHDVYTERIVSPGLASDGYVKKVRLSETTDPDEFRAAKVSLGVGLVISQVTLKQPMFKR
 XP_043131130.1 -----EDMTVTVFAGGDVYDLQIE-LKANGLSFNNLGSISQVQNFIQA-VSTGTHGTGSHV-G-NMATQVGLRVLDLDSG-----SLRVIDQTTNABELKAFRVLGALGLIETMTIKVHTYLV
 XP_002422445.1 KTFPEIKFVLTIVEAGIRFIFELNEY-LKRNRLATQNLGSISDSQIAGL-ISTGTHGTSYHG--LVSQQVSVKFLNSG-----ELTICSDIEKPEYFRAAILLSLKGIGIITHVTLRTCPKYTI
 NP_013624.1 -----PELHYADVTDGAMRLYQNEF-LGAKGYSIQNLGSISQS IAGV--ISTGSHGSSPYHG--LISQVNVNLVNVNGK-----ELKFLDEADNEPEVFKAAHLSVGKIGIIVSATVIRVPGFNI
 NP_003651626.1 -----ATGLCTQAGIRLRLASLA--LAAGLALPVLNLSIEQS IAGA-ISTGTHGSSLRHG--LVSECVALKVTLADG-----KTRCSPQOREPELFRALLSVGALGVVVEVTLRAVPAPFL
 XP_049145240.1 -----EGGSVTVIEAGVTFQLQAEY-LHARGASAGYTLVNNNITLAGC-VAMGAHRSSIREDS-MVAAGVLELIDIDNG-----ELRHLVRNNDDTTLAASSTSLGLGVVVRMKFKIYPDFKV
 NP_055577.1 -----KKQIVRVEPLVTMGQVLTAL-LTSGIWTPLVLELDDLTGVLGMLTGEISSHHY-G-LFQHCITAYELVADG-----SVFRCNTPENDSLFYAVPWSCTGLGFVLAARIVIPAKKY
 XP_004926865.1 -----ENNTVIRVEPLVTMGQSLT--LAPGLALPVVLELDDLTGVLGVMGTGVIETTSVHG-G-LFQHCYLEVLELADG-----SVNCSKDENADLIFYAVPWSYTGTLGFLTSIVIKVIPAKKY

	341	350	360	370	380	390	400	410	420	430	440	450	460	465
NP_0126287.1	HRHQD----	RRSLAQTLER--	LD-EFVDGNDHFEFFVF	FPYADKALTR-----				TMHR--	SDEQP--	KP-----				
MBI20379321.1	SAREE----	PMPFDRVMAD--	FE-QLVAENEHFEFFYF	PHTGNCNTC-----				RNNR--	STGPA--	APVG-----				
NP_254014.1	RERQW----	TARTEELLED--	VE-KNTREQNQHWEMLV	THSDYALS-----				ALNE--	TDEPK--	TPPL-----				
MC04763851.1	CETTE----	TIPFDDAVRD--	ME-AIARSAEYVKLWM	LPHTDIAVVF-----				RSER--	TTEPG--	R-----				
MBS1617249.1	HDRTY----	SMDFDKMLEK--	LD-DLVTTTDDHKMWMF	PHVKDILVY-----				QYQR--	TTAPA--	N-----				
HIA12326.1	EEFAY----	PVEFDEGVKR--	IP-ELLKEHEHMKLWMF	PHVKELQVY-----				CQNRTKAEHLN-----						
NEO81287.1	EVREQ----	PETLETVLDK--	LP-ERLQ-ADHYRWFYL	PHADRVWE-----				SATRKPPGKTA-----	P-----				KQNF	
BAY85903.1	EVKEQ----	PQSLATVFEN--	LP-ESLK-SDHYRWFYL	PHVDMAWEW-----				TATRQLPEEST-----	V-----				KKNI	
MC08715902.1	KSKQE----	IINFDTLLSN--	WD-NIWLESEFIRLWMF	PYTNKCVLW-----				--R--	ANKST--	DPLS-----				
PSP48796.1	RERTF----	PADVEAVLAD--	LE-AY-RECRNEFEFWF	PHTDTALVK-----				TLDE--	TDAEG--	APGR-----				IDGL
BAO94255.2	LETTY----	PTDIAGIKKD--	HE-ELLARKYRHYRVM	MLPRTGKVVV-----				--VSNP--	HAGG--	QPLL-----				QKVS
XP_844937.1	RESR-----	PIPRKLTGEBTKL	KRIASESEFYRFWM	PNPTDQCYESTAEFVGE	EAGDQTKRVDSEIKL	AMGK--	KHEAT--	LPMTAGNTIKTLTSS	SKLRNFSSEKCNST	GEDYQMWLNRQRTL				
XP_001682394.1	KLVSQ----	PLLMKDATNAALVA	-EKVSKTEYYRWMWV	PHPTDGCYES-----				YGRV--	ESTTGISALPPL-----				PDALTAQDDAPVQLS	PAALQEESSERTN
NP_484862.1	LETSF----	PSTLKEVLDN--	LD-SHLKSEYFRFLWF	PHSENVSI-----				YQDH--	TNKEP--					
EDL85374.1	QETSF----	PSTLKEVLDN--	LD-SHLKSEYFRFLWF	PHTENVSI-----				YQDH--	TNKAP--					
OCT81467.1	KEVPF----	SSTLQEVLDN--	LD-THLMSSEYFRFY	WFPHTENVSVF-----				YQDP--	TDKPP--					
XP_002122023.1	HEKKA----	SSKETVTLID--	LN-EHVSSAQHFQFV	WFPHTDNVVT-----				ARNR--	TRRLK--					
XP_005443605.1	HEVAF----	PSTLTEVLDD--	LD-DHLKRSQYFRFLWF	PHSENVSI-----				YQDP--	TNKPF--					
XP_024182616.1	VEHTT----	VSNMEEIKKN--	HK-KLLSENKHVKYLI	PIPYTDTVVVV-----				TCNP--	VSKWK--	GPPK-----				FKPKYTT
XP_024165881.1	SVSYD----	FTDDQDIENK--	YM-DHGKKEYFADITWY	PSKHTAAYR-----				YDNRVPLNTSG-----	DGTY-----				DFLGFQANPILVKS	SVRAS
XP_008240731.1	VEHTS----	VSNMEEIKKN--	HK-KLLSENKHVKYLI	PIPYTDTVVVV-----				TCNP--	VSKWK--	GPPK-----				FKPKYGS
XP_008220057.1	SIAYD----	FTDDAQIENI--	YM-DHAKKYEADITWY	PSKHTAAYR-----				YDNRVPLNAGS-----	DGTY-----				DFLGFQNSNILISK	SIRAT
XP_004291192.1	SITYD----	FTDDQDIEDK--	YM-VHGKKEYFADITWY	PSKHTAAYR-----				SDNRVPLNAGS-----	DGTY-----				DFLGFQANPILVKS	SVRAS
XP_004303609.1	VEHTT----	VSNMEEIKKN--	HK-KLLSENKHVKYLI	PIPYTDTVVIV-----				TCNP--	VSKWK--	GPPK-----				FKPKFTT
NP_190376.1	VEHTY----	VSNLQEIKNK--	HK-KLLSANKHKVKYLI	PIPYTDTVVVV-----				TCNP--	VSKWS--	GPPK-----				DKPKYTT
NP_182199.1	SLRYV----	MRNDSDFGDQ--	AV-TFGMKHEFADFILW	PLSQGKVYR-----				MDDRVAVNTSG-----	NGLL-----				DFMPPFRQLSAAIAI	IRSS
NP_564393.1	SLTFV----	MQNDSDFGDQ--	AV-TFGKHEFADFILW	PLSQGKVYR-----				MDDRVPVNTSG-----	NGLF-----				DFFPFRPQLSVALAI	IRSL
NP_182198.1	SLTYT----	MRNDSDFEDQ--	AV-TFGKHEFADFILW	PLSQGKVYR-----				RDDRVAVNTSG-----	NGLF-----				DFLPFRQLSAAIAI	IRTS
NP_182197.1	SLKYV----	MRNDLDFNDE--	AL-TFGKHEFADFVWL	PLSQGKVYR-----				MDDRVAVNTLG-----	NGLY-----				DFFPFRQLSAVLATRRS	
NP_200460.1	SLNYV----	MRNDSDFGDQ--	AV-SFGERHEFADFILW	PLSQGKVYR-----				MDDGRVPLNTSG-----	DGLF-----				EFFPFRQLSVLVAI	DRSL
XP_023896433.1	SITYL----	RKNDSDLGDM--	VA-SFGNEHEFADITWY	PSQRKAVYR-----				IDDRVYNTHGHGMNGLY-----					DFTPFRATASVELGIVRT	
XP_023896434.1	SITYL----	TKDSDSLGDK--	VT-SFGNEHEFADITWY	PSQRKAVYR-----				IDDRVNTNSPG-----	NGLY-----				DFTPFRSTASIALGA	IRTT
XP_023900169.1	VEHTT----	VSNVKEIKKN--	HK-KLLSENKHVKYL	PIPYTDTVVVV-----				TCNP--	VSKWR--	GPPK-----				FTPKEYSS
XP_023914113.1	SVTYN----	FTNDHDFEDV--	FM-INHAKKYEADITWY	PSKNTAVYR-----				YDERVPLNVSG-----	NGVY-----				DFIGFQSNLILISK	SIRAT
XP_043131130.1	KKTTRVLNATTDY	QAQMYKE--	IA-QLYKEHDMTYV	WGPHFD-----				WDSK--	AQDWA--	TEPTY-----				
XP_024224445.1	KSKQE----	IINFETLLNN--	WN-NIWLESEFIRLWMF	PYTNKCVLW-----				--R--	ANKSI--	DPLS-----				
NP_013624.1	KSTQE----	VITFENLLKQ--	WD-TLWTSSEFIRVMMW	PYTRKCVLW-----				RGNK--	TTDAQ--	NG-----				
XP_003651626.1	AWSQT----	IDADARLFTS--	WEGELWQAEFVRVWMW	FPMYRRAVVH-----				HADKVPREDIE--	SGRV-----					
XP_049145240.1	YAQQK----	TLEESEVLDD--	DIYGMIAAPYATANF	WMMYPRKHFHRR-----				YDVVPTNTSD-----						
NP_055577.1	VKLRF--	EPVRLGLEIAI	CAKTFHE--	SQRQENHFEVGLLY	SLDGAIVM-----			TGVM--	TDE-A--	EPSK-----				LNSI
XP_0042926865.1	VRHHY--	YPFNNLSDPAKRSES	OKSLPNPHOFVEALL	YTRDSGVLW-----				LGDM--	VDEVG--	RDGK-----				LNPI

Conserved
Cysteine (C) region

	466	470	480	490	500	510	520	530	540	550	560	570	580	590
NP_216287.1	--RVSGW--	VDDELLSNG				GLSLICQTG	RRF	--PSVAPRLNRLM--	TNM	-----MSSSTVQ--	DRAYKVF	--TCEKV--		RFTEME
MBI0379321.1	--RVSGW--	VDDELLSNG				IFQVACSVG	RAV	--PATIPGIAKIS--	SRA	-----LSARTYT--	DIPYKVFT--	SPERV		RFVEME
NP_254014.1		DPAKEGEGNEFVGLIEGLD					KYL	-----SDFPETRRTLNSLR--	HFARFDERV	--DDSYAVYA--	NVENV			RFNEME
MCC04763851.1		ISPVVRW--	IDEHIVNQL			LFATVVLWG	RWF	-----NGLIAPLNRLVARTYL--	KPRRVV	--GRSDKVMs--	LAMPF			RHREME
MSB1617249.1		DSRLRQW--	LMDEVLSVS			VYRFLVWIG	NFL	-----PSLRPFPPKILVLSFR--	SPLDRI	--EKSYRVFN--	VADPP			IHREAE
HIA12326.1		TGLFIW--	VDEKILAKY			ITFLLNLG	LWF	-----PTLTAPINRIKILKF--	KKEHRI	--DYSYRVFN--	TEPPP			KHHEAE
NEO81287.1		QRFQFSW--	YREKLIGY			TFELLLYA	TYN	-----QRLIPLVNRWVYQQMF--	SKSKQSR	--GNSVSQFN--	FDCLF			KQQVNE
BAY85903.1		LQRLRDW--	YQKLIIGFY			TFQLLLYA	TYN	-----QNLIPGINRWYAQQM--	SQPKQSR	--GDSLQSQFN--	FDCLF			KQQVNE
MCP8715902.1		--DRPWSW--	YGTKLGRF			FYESLLWVSVHLY	--PSSTPYVERFVFGQQYGNVETLKGKDIAV--	Q-NSVEGLN--	MDCLF					SQFVNE
PSP48796.1		EERLENL--	VWEACCRAG			-----TRA--	RPLAPALNRAVAATFS	-----DSErv--	GPARDVFP--	TTAV				RFNETE
BAO94255.2		EDPTLPMRQLLSKKVGVP				AGESEGLSFAQLRDALLDLN	PLD	-----TAHVDEVNRAEAEFWQ--	RNSAQLR	--S-DSTEVLG--	FEQGG			SQRVLE
XP_844937.1		RTRICKI--	LKGSWLRHG			-----VVEAALAAA--	VIQ	-----PGIQPYINRTYRRLFY--	NAPEVQY	--G-TSLECF--	FDCLF			KQWACE
XP_001682394.1		ESAEERAGDRSISLIVKSA				LKYIATDFVRHQVVECSLYAA	CLC	-----PAIQPVYNKAYRRAFF--	SAPQVQR	--G-SALECF--	FDCLF			KQWANE
NP_848862.1		--SSASNW--	FDWYAIIFY			-----LLEFLLWTS--	TYL	-----PRLVGWINRFFFWLLF--	NCKKESS	--NLSHKIFS--	YBCRF			KQHVDQ
EDL85374.1		--SSASNW--	FDWYAIIFY			-----LLEFLLWTS--	TYL	-----PCLVGWINRFFFWMLF--	NCKKESS	--NLSHKIFT--	YBCRF			KQHVDQ
OCT81467.1		--SSKANW--	FRDTPFLGNY			-----LLEFLLWIS--	TFI	-----TGMVPWINRFFFWLLF--	ASKSEQV	--NISHKVFN--	FDCLF			KQHVDQ
XP_002122023.1		--LVKNWN--	FKDIIIGHH			-----LLEFCLWIA--	TFP	-----SSIVPLITSVFFKFCY--	EGSSECI	--DRSDKIFN--	IDCLF			KQVYTE
XP_005443605.1		--SSASW--	LWDYAVGY			-----LLEFLLWIS--	TFV	-----PSLVWINRFFFWLLF--	SSRVENI	--SISYKIFN--	YBCRF			KQHVDQ
XP_024182616.1		DEAIQHVRDLYRECLKKYRVV				P-DNSVDINELSFTELRDKLLALN	PLN	-----KDHIKVVNQAEAEFWR--	KSEGYRV	--G-WSDEILG--	FDGGQQ			WVSETCFPAGT
XP_024165881.1		EKVMDNNGHNLGGKCLIAS				-----SFLGYKKLVANGL--		KNGLIFTGYPVIGNQG	-----KMQTSGSCL--	Y-STGKDS--	CAWDP			RINGLFFYETT
XP_008240731.1		DEAIQHVRDLYRECLKQCSVV				PVKNDVDINELSFTELRDKLLALD	PLD	-----KDHVVKVNQAEAEFWR--	KSEGYRV	--G-WSDEILG--	FDGGQQ			WVSETCFPAGT
XP_008220057.1		EKTMNDNGRSLGGKCTMAA				-----SFLGYKKLVANGL--		KNGLIFTGYPVIGNQG	-----KMQTSGSCL--	Y-STAADNS--	CAWDP			RINGLFFYETT
XP_004291192.1		EKVMNDNGHSLGGKCLLAS				-----SFLGYKKLVANGL--		KNGLIFTGYPVIGNQG	-----KMQTSGSCL--	Y-STSQASS--	CAWDP			RINGLFFYETT
XP_004303609.1		DEAIQHVRDLYRECLKKYRVV				P-DNSVDINELSFTELRDKLLIALN	PLN	-----KDHIKVVNQAEAEFWR--	KSEGYRV	--G-WSDEILG--	FDGGQQ			WVSETCFPAGT
NP_190376.1		DEAVQHVRDLYRESIKYRVQDSGKKSPPDSSEPDIIQELSFTELRDKLLALD	PLN			DVHVAKVNQAEAEFWK	-----KSEGYRV--	G-WSDEILG--	FDGGQQ					WVSECFPAGT
NP_182199.1		EETQERFRDANGKCAAGT				-----LITSLTFATSYGLT--		NNGMIFTGYPVIGSQN	-----RMMSSGSCL--	D-SLHDGLI--	TACPWDS			RKSGEFFHQT
NP_564393.1		EESSESSGDANDKCARAE				-----QITSPFLFSISYGV--		NNGMEFTGYPVIGSQN	-----HMMSSGTC--	D-SHQDGLI--	TSCPWDP			RKGGQFFHQT
NP_182198.2		EETQERFRDANGKCVGAT				-----IISSTLFPAPSYGLT--		NNGIFTGYPVIGSQN	-----RMMSSGSCL--	D-SLQDGLI--	TACAWDS			RKSGEFFHQT
NP_182197.1		EETQETLRDANGKCVTAT				-----TISSTLFSTSYGLT--		NNGITFTGYPVIGSQN	-----RMMSSGSCL--	D-GLDKLI--	SACAWDS			RVKGVFYHQT
NP_200460.1		EESSESLDANMCKVRAK				-----LVSSSMFLMSYGV--		NNGLIFTGYPVIGMQN	-----HMMSSGSCL--	D-SRQDGLI--	TACPWDP			RKGGQFFHQT
XP_023896433.1		EEVQESLSDAAGKCLGAE				-----TVILALETATGYLS--		NNGILFTGYPVIGYHN	-----HLQASGTC--	D-SLNDALI--	TACAWDP			RVKGEFFHQT
XP_023896434.1		EDFQESLSDAIGKCLSAE				-----TVILTLKTATGYLS--		NNGILFTGYPVIGYHN	-----RFQASGTC--	D-SLDEALI--	TACAWDP			RVKGEFFHQT
XP_023900169.1		DEAIQHVRDLYKDSFKYRKITTT				ETPDSNEPDINELSFTELRDKLLAMD	PLN	-----KDVIKINQAEAEFWR--	KSEGYRV	--G-WSDEILG--	FDGGQQ			WVSETCFPAGT
XP_023914113.1		EKTFTNTRDVGKCTLAT				-----TTVAACKLVGNGL--		KNGLIFTGYPVIGQQG	-----KMQTSGSCL--	Y-SSSSFST--	CAWDP			RINGLFFYETT
XP_043131130.1		--FASWWEPTNYTGVRNC				-----TLNYCANGCG--		DCKKNYICYDEVSDAV	-----SCGPEGVC--		THQFYA			HFFPAE
XP_002422445.1		--DRPWSW--	YGTKLGRF			-----FYESLLWISVKEF--		PKLTPTFIEKPIFNQYQGVETLKGKDIAV	Q-NSVEGLN--	MDCLF				SQFVNE
NP_013624.1		--PAKSW--	WGTKLGRF			-----FYETLLWISTKIY--		APLTPEVEKVFVFNRGYQGLKESSTGDVNV	T-DSISGFN--	MDCLF				SQFVDE
XP_003651626.1		--RHRRTSFLDSDRLGYH				-----VYHNLALA--	RWF	-----PRITPVIEWFVGLNNGYFRNGERTRVSGV	Q-PSYAEFL--	LNGLY				SQFVNE
XP_049145240.1		---QQGFQNTFSVTDLEG				-----TTARTLLDSG--	RYL	-----PTSNMLAEIIFFLGWSKPNFREKTTNQPIE	W-PVYGNW--	YDVLIGGLYDPQRPLWEVGVQGYTLE				
NP_055577.1		GNYYKPW--	FFKHVENYL			-----KTNREGLEYIPLRHHYHRTSRISFWELQDII	IPFGNNPIFRYLEFGWMVPPKISLLKL	--T-QGETLRKLYE--						HHVVDQ
XP_004926865.1		GVWHAEW--	FFKQVEKHL			-----KRRKRTAIEYIPLRDYHHRTSRISFWELQDII	ISFGNNPIFRYLEFGWLMPEVSLLLKL	--T-QPEAVTKLYN--						AHVVDQ

Glu-Arg pair (EXR)
region

	591	600	610	620	630	640	650	660	670	680	690	700	710	715
NP_216287.1	YAI	PREN	---GREALQ	RVIDLVR		-----RSLPIMFPIEVR				SAPD	DSFLSTA			YGRDTCYIAV
MBI0379321.1	YAV	PREA	---AVTVR	ELKATVER		-----SGLRVSFPVEVR				APADDI	PLSTA			SGRETAYIAV
NP_254014.1	YSV	PAEH	---GPAC	LEILALIRD		-----KDLRTWFPIEVR				VKADDI	PLSMF			EGRDSCSISV
MCC04763851.1	YSV	PMTK	---AGAL	LQRTALIGS		-----HNLRVNFIIEVR				VRGDDA	WLSPA			TARDSQCLGA
MSB1617249.1	WAF	DLRD	---AREV	LISAYKMI	DE	-----SDHKINFIQIE				VKGGDY	ALSPC			YRRDSIWI
HIA12326.1	YAI	PVEK	---AGEAM	QIRKMI	DE	-----RLRHVNFIIEVR				VKADDI	CLISPA			FGRTDTCYVGA
NEO81287.1	WAI	PIEH	---TAEAI	LIQIRMIKE		-----KDYQVHLPIEVR				VKGGDI	WLSPC			QCRDSCYIGV
BAY85903.1	WAI	PIEY	---TVQAE	LEIRDLIQ		-----KDYKAHLPIEVR				VKGGDI	WLSPC			QCRDSCYIGI
MCP8715902.1	WSA	PLVS	---GPEV	LKELKQIITN	--AA	-----NNGDFYVHAPIEVR	CSNVVTSY	SNKPF	IENDGKPS	LYPSQEWL	SHRSKTS	SAGPIPGNNLR	APFLD	NSPNLPYTTKE
PSP48796.1	YGV	PRDD	---AVAA	FRDLRSVSES		-----HDVMFPIEVR				VAGDDI	PLSPA			TGRDSTFIAV
BAO94255.2	VAF	IKTPLADIE	FMKVLAVIEA			-----RGIPAPAPIEVR				TARSAQ	PMSPASPT			PGDFPSWVGI
XP_844937.1	WAI	DISN	---VMPA	FHYLRLGISS		-----ENLSVHFPIEVR				TGADTA	ALSPA			HGRQTCWIGI
XP_001682394.1	WAI	DASR	---AVEAF	NRLRDMIDR		-----EGMLLHFPPIEVR				TAPDVS	DMSPA			VRGPTCWI
NP_848862.1	WAI	PREK	---TKEAL	LEKAMLEA		-----HPKVVAHYPIEVR				TRGDDI	LSPC			FQRDSCYMINI
EDL85374.1	WAI	PREK	---TKEAL	LEKAMLEA		-----HPKVVAHYPIEVR				TRGDDI	LSPC			FQRDSCYMINI
OCT81467.1	WAI	PIEK	---TKDAL	MQLKGVLEK		-----NPHVVAHYPIEVR				ARGDDI	LSPC			FQRDSCYMINI
XP_002122023.1	WAI	PRRH	---TAIV	LREMKNWIEN		-----NPDCKIHFPIEVR				VQKDN	IMMAPS			CEQDVTYIGI
XP_005443605.1	WAI	PIEK	---TKEAL	LELKGALEN		-----NPKLVVAHYPIEVR				TRGDEI	WLSPC			FQRDSCYMINI
XP_024182616.1	LAK	FSMK	---DLEY	IADLKQIEK		-----EEIPAPAPIEVR				TASTKS	PMSPASSR			EDDIFSWVGI
XP_024165881.1	AIF	PAPK	---FRDF	ILDVKKLRDLVKPE		-----NFCGVDIYNGFLIE				IKASEA	YLG			PQESVSVVDF
XP_008240731.1	LSK	PSLK	---DLEY	IEDVKQLIEK		-----NEVPAPAPIEVR				TACTKS	PMSPASSTK			DDIFSWVGI
XP_008220057.1	AIF	PASK	---FGDF	IRDVKKLRDL	KPD	-----NFCGVDIYNGFLIE				IRATGA	YLG			QFEDSVVDF
XP_004291192.1	AIF	PAPK	---FRDF	ILDVKKLRDLVKTE		-----NFCGVDIYNGFLIE				IKASEA	YLG			PQESVSVVDF
XP_004303609.1	IAK	FSMK	---DLEY	IEDLKQIEK		-----EEIPAPAPIEVR				TASTKS	PMSPASSL			EDDIFSWVGI
NP_190376.1	LAN	FSMK	---DLEY	IEELKKLIEK		-----EAIPAPAPIEVR				TARSKS	PISPAFSTS			EDDIFSWVGI
NP_182199.1	FSI	PLTQ	---VKSF	INDIKSLVKI	ESK	-----SLCVLELYDGLIMRY				VTSSPA	YLG			KETEALDFDL
NP_564393.1	FSI	PLTR	---VKGF	INDIKALVKI	EPK	-----SLCALERSNGILIRY				VTSSPA	FLG			KEEKALDFDL
NP_182198.2	LSV	PLTQ	---VKSF	ISIDIKSLVKI	EQK	-----SLCGLLEHYGILIMRY				VTSSPA	YLG			KETEALDFDI
NP_182197.1	FSI	PLTQ	---VKSF	INDIKSLVKI	DSK	-----SLCGLLEHYGILIMRY				VTSSPA	YLG			KETEALDFDI
NP_200460.1	FSV	SLTN	---VKSF	INDIKALVKI	EPK	-----SLCVLEGSNGILIRY				VTSSLA	FLG			KEEKALDFDL
XP_023896433.1	FSV	GLSV	---AKSF	IQDVQKLIEL	EPK	-----AMCGVELYNGILIMRY				VKASSA	YLG			KQEDAI
XP_023896434.1	FSV	GLSV	---AKSF	IQDVQKLIEL	VPK	-----AMCGVELYNGILIMRY				VKASSA	YLG			KQEDAI
XP_023900169.1	LAK	FSMK	---DLEF	IEDQKLIEK		-----QDIPAPAPIEVR				TARSKS	PMSPASSG			EDDIFSWVGI
XP_023914113.1	AIF	PAPK	---PGDF	IRDVKKLRDL	NPE	-----NFCGVDIYNGFLIE				IKASGA	YLG			PQESVSVVDF
XP_043131130.1	YL	LEAA	---NYTL	FQSQS	TPRMK	-----APYNEQMVPIEVR				LKGGDA	YLS	SPVNTYN		LGPESSGVFA
XP_002422445.1	WSS	PLNS	---GPD	LILNLKKI	IID	--AA	-----EKNDFYVHAPIEVR	CSNVVTSY	SNKPF	IENDGKPS	LYPSQEWL	SHRSKTS	SAGPIPGNNLR	APFLD
NP_013624.1	WGC	PMDN	---GLEV	LRSLDLSIAQ	--AA	-----INKEFYVHVMPIEVR	CSNTTLPSE	PLDTSKRTN	--TSPG	PVYGNVCR	PFPLDNT	PSHCRFAPLENVT		NSQLTLYINA
XP_003651626.1	WAI	PLHK	---GPEAL	QRLSAMLNRLRPGDPGYVEHRI	PFS	EAAGLWYHS	PVEV	YSDTTVNAS	-----ARGNR	PWLVDTP				ADGALYILNA
XP_049145240.1	LAF	PIITQ	---ANAML	KRVRAFLDD		-----EAKKLIVMTSTVRSKINIK				GRPY	FDLLGQV	TYGTADGAD		WSKGVIMLNA
NP_055577.1	MLV	PMKC	---LQQA	LHTFQNDIHV		-----YPWLCPPIE				LPSQP	GLVHPK			GNEAEYIDI
XP_004926865.1	MLI	PIEL	---LEKAI	IAFFHDEF	EV	-----YPWLCPPIE				NNPQ	GLKIKP			GEESQMFVDI

HWXK motif

	region															
	716	720	730	740	750	760	770	780	790	800	810	820	830	840		
NP_216287.1	HQY	-----	AG-MEFESYFRA	---VEEIMD-DYAG	---	RPHWGKRHY	-----	QTAAT	-----							
MBI0379321.1	HLV	-----	RG-TPHRAAYTA	---AEQIMI-AHGG	---	RPHWGKLHS	-----	RDAAY	-----							
NP_254014.1	HQH	-----	YS-MDHHNFFAA	---VEPIFW-KYAG	---	RPHWGKLLHG	-----	LNAHQ	-----							
MCO4763851.1	YMA	-----	ES-EGIERYPGA	---FQGAMK-ALGG	---	RPHWGKEFQ	-----	ANVAE	-----							
MBS1617249.1	YLI	-----	GD-RGWDALLHD	---FEALAV-KYNG	---	RPHWGKXYH	-----	IDRAY	-----							
HIA12326.1	YKA	-----	GD-RGWKEYLHG	---FEEIMA-AYDG	---	RPHWGKEFS	-----	IDKNA	-----							
NEO81287.1	ITY	---MPY	---GKY-VDCQSYFDD	---YEKIMA-KLDG	---	RPHWAKKFG	-----	PDADK	-----							
BAY85903.1	INY	---MPY	---GKR-VDCRDYFND	---YEKIMA-KFDG	---	RPHWAKRFG	-----	PDADK	-----							
MCP8715902.1	TMY	---RPF	---GTN-VETHQWFQL	---FENVMS-KAGG	---	KPHWAKNFIG	-----	LTQDEKYNEKEDLKTQLEFGGKPFYSMIGFKPVMKDFWGEDLVKFNNVRKETDPHG VF	-----							
PSP48796.1	HAY	-----	HR-REWETLVRE	---AEA VFD-RYDG	---	RPHWGKHHT	-----	KTAAE	-----							
BAO94255.2	IMY	NTPGQDEQGLEKVAKAFKSYCAALEGVHQ	---EYGA	---	VPHWAKVEP	-----	CDDPA	-----								
XP_844937.1	VMY	---RPY	---LRHARDTLRYDA	---FSKAMT-VMGG	---	RPHWAKYYT	-----	WGPEQ	-----							
XP_001682394.1	VMY	---RPY	---GQEARDTRRCYDG	---FCRVME-EMGG	---	RPHWAKYYD	-----	WGHRE	-----							
NP_848862.1	IMY	---RPY	---GKD-VPRLDYWLA	---YETIMK-KFGG	---	RPHWAKAHN	-----	CTRKD	-----							
EDL85374.1	IMY	---RPY	---GKD-VPRLDYWLA	---YETIMK-KFGG	---	RPHWAKAHN	-----	CTRKD	-----							
OCT81467.1	IMY	---RPY	---GKD-VAHQEYWVE	---YENIMK-KVGG	---	RPHWAKAHT	-----	CTRKD	-----							
XP_002122023.1	ISY	---RPY	---GKL-VPHDKWFTF	---YENLAL-KYGG	---	RPHWAKDHK	-----	VNSQT	-----							
XP_005443605.1	IMY	---RPY	---GKN-VPRLNMYLT	---YEGIMK-KYGG	---	RPHWAKAHS	-----	CTRKD	-----							
XP_024182616.1	IMY	---LPTTDAQRKRDITEEFFHYRHLTQTRLWDITYG	---	---	---	RPHWAKIEV	-----	PKDKE	-----							
XP_024165881.1	NPY	---RAD-DASTP-RFDQDVWEE	---VEQLAFFKHSS	---	---	YBHWAKNRN	-----	MAFLN	-----							
XP_008240731.1	IMY	---LPTTDAQRKRDITEEFFHYRHLTQVLWDKYGA	---	---	---	KPHWAKIEV	-----	PKDKE	-----							
XP_008220057.1	NYF	---RAD-DALTP-RFNQDVWEE	---VEQLAFFKYSA	---	---	YBHWAKNRN	-----	LAFLD	-----							
XP_004291192.1	NPY	---RAD-DALTP-RFDQDVWEE	---VEQLAFFKHGA	---	---	RPHWAKNRN	-----	IAFLN	-----							
XP_004303609.1	IMY	---LPTTDAQRKRDITEEFFHYRHLTQTRLWDITYG	---	---	---	KPHWAKIEV	-----	PKDKE	-----							
NP_190376.1	NPY	---RAD-DASTP-RFDQDVWEE	---VEQLAFFKHSS	---	---	YBHWAKNRN	-----	MAFLN	-----							
NP_182199.1	TYT	---RAK-DPLSP-RLYEDFIEE	---IEQIALFKYSA	---	---	YBHWGKNRN	-----	LAFDG	-----							
NP_564393.1	TYT	---RSKDDPLAP-RLYEDFIEE	---IEQMAIFKYNA	---	---	LPHWGKNRN	-----	LAFDG	-----							
NP_182198.2	TYT	---RAK-DPLTP-RLYEDFIEE	---IEQIALFKYNA	---	---	LPHWGKNRN	-----	LAFDG	-----							
NP_182197.1	TYT	---RAN-DPLTP-RLYEDFIEE	---IEQIALFKYNA	---	---	LPHWGKNRN	-----	LAFDG	-----							
NP_200460.1	TYT	---RSKNDPLAP-RLYEDYIEE	---IEQMAIFKYNA	---	---	LPHWGKNRN	-----	LAFDG	-----							
XP_023896433.1	TYT	---RSK-DPLTP-RLFEDILEE	---VEQLALFKYNA	---	---	LPHWGKNRN	-----	IAFDG	-----							
XP_023896434.1	TYT	---RSK-DPLTP-RLYEDILEE	---VEQLALFKYNA	---	---	LPHWGKNRN	-----	IAFDG	-----							
XP_023900169.1	IMY	---LPTMDAQRKRDITEEFFHYRHLTQSLWDYSA	---	---	---	YBHWAKIEV	-----	PKDKE	-----							
XP_023914113.1	NPY	---RAS-EASTP-RLNQDVWEE	---VEQMAFFKYGA	---	---	KPHWAKNRN	-----	LAFLD	-----							
XP_043131130.1	VLE	---IDW	---MQEYNNFTTLWQNQELAEFLP-RFGEAYNVRS	HNWKMSP	---	---	---	ANSTY	-----							
XP_002422445.1	TMY	---RPF	---GTN-VETHKWQFL	---FENVMS-KAGG	---	KPHWAKNFIG	-----	LTQDEKYDKLSDLKTQLEFGGKPFYTM LGFKPVMKNWFGKDLIDFNNVRKQTDPEGVF	-----							
NP_013624.1	TIY	---RPF	---GCN-TP IHKWFTL	---FENTMM-VAGG	---	KPHWAKNFIG	-----	LAAGVPVKKDTDYDD	-----							
XP_003651626.1	TMY	---RPYFRDPTG-EATERYYAA	---FEWLMR-DLGG	---	---	RPHWAKTFT	-----	VSPDE	-----							
XP_049145240.1	PTF	---RPT	---VGDHSRFNEPFYHN	---LAKVLIDEFPC	---	RPHWTKNTR	-----	EVFAQ	-----							
NP_055577.1	GAYGEPR	---VGH	---FEARSCMRQ	---LEKFVRSVHGF	---	QMLYADCYM	-----	NREE	-----							
XP_004926865.1	GVYGVPK	---AKG	---FETIASTRH	---VESFVIQNGQF	---	QMLYADTYT	-----	TREE	-----							

	841	850	860	870	880	890	900	910	920	930		
NP_216287.1	-----	NDYTRRVLPZ	-----									
MBI0379321.1	-----	NDYLRRVLGDZ	-----									
NP_254014.1	-----	NAHLSLIGVTZ	-----									
MCO4763851.1	-----	NRFVARLFESTEEIHZ	-----									
MBS1617249.1	-----	NMTIDAICGTIZ	-----									
HIA12326.1	-----	NRFVREYIMZ	-----									
NEO81287.1	-----	NSYSDRVLGKPKQKTALSSYZ	-----									
BAY85903.1	-----	NSYSDRVLGKGRGZ	-----									
MCP8715902.1	-----	LSGKVWAERNGILLDZ	-----									
PSP48796.1	-----	NEHLRVDVFGAZ	-----									
BAO94255.2	-----	NDLIERVIGHZ	-----									
XP_844937.1	-----	NGWFNSLSGNSPVLNSTISHLZ	-----									
XP_001682394.1	-----	NCWFNRNLSFDRV-NSTACTPZ	-----									
NP_848862.1	-----	NSYLEKVFY	-----									
EDL85374.1	-----	NSYLEKVFY	-----									
OCT81467.1	-----	NAYLEKVFY	-----									
XP_002122023.1	-----	NEYWKRLK	-----									
XP_005443605.1	-----	NSYLEKVFY	-----									
XP_024182616.1	-----	NIKLEKLFLSDTIZ	-----									
XP_024165881.1	-----	SEWSDEIL	---VFGKEGAKSDGCGLEGMCICSEDRHCS	PGKGYFCKPGVYEEARVCRYSPSSNVESDSPKTDLZ	-----							
XP_008240731.1	-----	NNKLEKLFLSDTIZ	-----									
XP_008220057.1	-----	SEWSDEV	---VFGKEGVKGDGCALEGQCICSEDRHCS	PGKGYFCKPGLVYQEARVCRYSSSPSPFVSGSGDKNDLZ	-----							
XP_004291192.1	-----	SEWSDEV	---VYGKEGGKSDGCGLEGMCICSEDRHCS	PGKGYFCKPGLVYQEARVCRYSTSSNVASDNPKPTDZ	-----							
XP_004303609.1	-----	NVLEKLFLPSSDTIZ	-----									
NP_190376.1	-----	NNMVEKLFPVSTTAZ	-----									
NP_182199.1	-----	SEWTDQILGKGNVTI	IKDGALEGLCVCSDDAHCAPTKGYFCRPGKVYKEARVCTRADD	---ISVIQSLSYZ	-----							
NP_564393.1	-----	TEWNTQILGLKGNVTI	IKDGALEGLCVCSDDAHCAPTKGYFCRPGKVYKEARVCTRADD	---ISVIQSLSYZ	-----							
NP_182198.2	-----	SEWTDQILGKGNVTI	IKDGALEGLCVCSDDAHCAPTKGYFCRPGKVYKEARVCTRADD	---ISVIQSLSYZ	-----							
NP_182197.1	-----	SEWTDQILGKGNVTI	IKDGALEGLCVCSDDAHCAPTKGYFCRPGKVYKEARVCTRADD	---ISVIQSLSYZ	-----							
NP_200460.1	-----	TEWNTQILGKGNVTI	IKDGALEGLCVCSDDAHCAPTKGYFCRPGKVYKEARVCTRADD	---ISVIQSLSYZ	-----							
XP_023896433.1	-----	SEWTDQVLGLKNSVTI	IKDGALEGLCVCSDDAHCAPTKGYFCRPGKVYKEARVCTRADD	---ISVIQSLSYZ	-----							
XP_023896434.1	-----	SEWTDQVLGLKNSVTI	IKDGALEGLCVCSDDAHCAPTKGYFCRPGKVYKEARVCTRADD	---ISVIQSLSYZ	-----							
XP_023900169.1	-----	NNKLEKLFLPSSDTIZ	-----									
XP_023914113.1	-----	SEWSDEIL	---FGKEAANKGDGCALEGQCMCSEDRHCS	PENGYFCKQGLVYKEARVCRYSQSLSGNEIFYGIRYZ	-----							
XP_043131130.1	-----	NDFLVQQLGITRCSGYLSMZ	-----									
XP_002422445.1	-----	LSGKVWAERNGILLDZ	-----									
NP_013624.1	-----	INGIIDPSELSDZ	-----									
XP_003651626.1	-----	GPWHRRYLLGERSAGEGGEMERLPLEEVGFTTREPAGGGZ	-----									
XP_049145240.1	-----	KSVVGEAIGVYZ	-----									
NP_055577.1	-----	PEVYDKICKAARHZ	-----									
XP_004926865.1	-----	PEIYDKVNNVRKZ	-----									

Figure S1