

Supporting Information S4

The benchmark dataset $S_{\xi=7}(T)$ used to train and test the model for predicting the possibility of carbonylation at Thr site. It contains 121 positive samples and 732 negative samples. None of the sequences included has $\geq 30\%$ pairwise sequence identity with any other in a same subset.

(1) List of the 121 peptide samples in the positive subset $S_{\xi=7}^+(T)$

HLELRNRTPAAVREL
 PRKRPPVTQAAGALP
 LLAQKHKTQRSPVRI
 NEPEAEKTHLFAKQE
 RKSLVMHTPPVLKKI
 GSKRQLQTPKEKAQA
 KDGSVTGKRLRCMP
 DTEEADRTSSKKTKT
 MPRRLTQTGKRKHGA
 GTARRTGTPSDPRRR
 NSKHLNGTITAKYTY
 GKEMLEQTLQKVEL
 NELLKQKTIELTRAC
 LDTMSKKTALDQLS
 ASKEQGTPVEKSKK
 DLIQHRRTQDHKIAK
 IEEELKTTKRKMNLK
 VQKANPGTLAAEIRA
 WIWVQLRTGLARDGR
 VSAAQGTTGTAPGAR
 FQASLPLTRIDEEEA
 LNDKHLKTLKTVFS
 RPLKSQITVTRHDKR
 DSKNSKKTNTEFLHT
 ERVEQFLTIARRRGR
 PDDRTARTLTIAKV
 RDLRFVRTAEILAGK
 KDNIPPLTMKRIRER
 KGITFIFDSEIKDE
 MHFRSLRTKLLMSR
 RQPVTKNFRQYRVL
 SEVLVQQTQLQIRQS
 FVQRRGHTVLKIHL
 KAKKPAATRKSNNP
 MAVLEALTGVLRSCL
 HSGKLYKTKSNKELH
 VQSIKLITEGASKRI
 YSAEWKSTVSRLGRW
 GAFKAVMTSIKQLSS
 RDKLKFTRPKVRVP
 VKRGDIFLLEELPK
 QLIPRHKTREKVD
 NEIVNIRTSLLNLVD
 KQELEVLTANIQLK
 PKDEDERTVYVELLP
 EKIRLAKTQQASKHI
 EDLGRKRTTMLKIAT
 EVAEKQATDVKPKAA

EKFKALYTLPRSVDD
ARIKAKLTQIRRYGE
SKGELLSTLLTKEKA
LTKLKAKTDNVVQAK
QWIRATATALERCSE
PLGAEKITGELEEMR
ISNVDVETQSGKTVI
LPDMSVKTTPKISMPD
RSIRKSSTSSDNFKA
TPESKKATSCFPRPM
SVFLRVRTNVGVRVL
SVSVEKSTSSNRKNQ
TSPLQSPTKAKPKVE
SQGKLLRTLFGVKR
STAVGLVTKDPGKKP
LGKKGSTVGLVAKD
NYDKILATKKNLDHV
RAALDKATVLLSMK
LLRAGYLTYGIEAL
ETKETTNTNKQTSTD
TPKKPAPTPKEPAP
RTARAITTRSGQTLS
KNGKVYFTSDAGIAG
IALTTYETLRLCLDE
SSVGWDATEDLRISR
TEELIRLTQRLRFHK
NHFLQVETRRAGERL
QTSLIVATLKKMLPI
DVKSTAFTFRSSKEK
NKEKEMVTSLFCKLG
LRIHAHFTGLRYLLY
TSSPPRKTRRLSPSA
RLLRIAATPSCHLLH
PPKRAPSTTLTLRSK
LRKGIPRTKSVGEDE
EPARWDTTLPTSPK
WKKLLEATELKGIKL
VEKPPKFTEKGNLEV
LRLLNKHTALRGEMS
ATRVRNITKKSSHLP
SYKLKSATGSVVSTR
ITKKMDKTIPEGRIR
VARQTLETIRSVGY
DLKTIYQTGCQTSTV
ITKSSTSTIKDKDEL
VRVTGIPTPVVKFYR
IMKETRKTVPKIV
KKLSDISTVVGKEVQ
SSEPKECTYTIPKLL
VDVKFKDTVILKAGE
KILSNNNTSENTLKR
CLQTLTDTKVFHFID
KPLKKKPTVLLPQS
GISSPLTGKRQDSG
VLCIIQDTTNSKTVN
QLSKKLKTVLDQARQ

LEEIKERTATGLTHI
 KKTILMVTNRRVLCI
 RRIKVSNTLESRLDL
 KGFTVKRTLVIHQRT
 LLAPLNDTRVVHAAK
 NETRESETKLEKELPG
 DGTWSPRTPSCGDIC
 RDSARQSTLDKELXX
 AVSEHEATKCQSFRD
 KEGYYGYTGAFRCLV
 SSLEACTFRRPXXX
 LEGNKRITCRNGQWS
 QKLEPLGTELHKNAK
 CLLPKIETMREKVLV
 KVGTRCCTKPESERM
 TPVSEKVTKCCTESL
 EKQIKKQTALVELLK

(2) List of the 732 peptide samples in the negative subset $S_{\xi=7}^-(T)$

HGSIFSGTAAPRCEI
 SRNKSGRTALMLACE
 LPAIFEYTVVGEPA
 SIAFQMPTIAKDGNV
 YDINQIITAVMTHT
 GAAGLEGTAATAAAG
 AREPPGSTAGLPQEP
 KMMHIAATLIQRRFR
 CPPGLYQTVPPEEHY
 QNIPPSFTKKLTKMD
 SASAPSSTPTGTTVV
 VSPRGMVTRSSPGAG
 DEDDNMSTVMRLRTK
 ARNGISPTNPTKLQI
 RELQARLTLVGKEGP
 NLKFQPPTLGPEPAA
 SHRVGAATAGRLPAS
 PAAASLSTATDGLAA
 PKESMLKTTLTAFGF
 YPEPRVLTLLVRITPV
 DSLPSPSTIVSGDIP
 RVNKVPVTMTRYRST
 HLRKAAITIQSSYRR
 RLLMASPTRKPEPQV
 QYHKMYRTVKAITGR
 KSKKRNSTQLKSRVK
 KAKEDDKTLSKEPSK
 GGGRNDITGRFTRHL
 ENHPQWDTAIEGDED
 DRWNSAFTRKDEIIT
 SEPSDAITCRDDVEA
 PRCAEHGTCRDGKCE
 NGDGGCGTGGRNCSA

LCKAFVPTCIEQIHV
DFEVLTKTAHGHLVP
IVYHRKVTDDISKIK
RSSRKICTIGIAPWG
VQGDSQGTPTSSQGS
YIVQKRETSRLAWTN
DRKILDDTETVVSQ
CRDELVRTTAQYDQA
QEKSPVNTKSLFKRL
IEELLSQTTNPDRFS
ATQPAAPTPTATITS
SKKDDFFTSFKISCQ
EWRRANHTPESCPET
PSAFYQQLPNSHLT
MREEAERTRDELERR
GHMYMILTPSDIQIQ
SPQREVLTVPEANSE
KLESLVLTHLSRCD
FSCGADGTMKMRILP
KPESIKVTTGDTCTL
VVKYRAGTSVKLRAG
LKDLLLNTMSQEEKA
SYSSLLETIEVKGAG
PHGGAMDTHFANMRS
KFTSITDTPEQVLAK
AQEERLKTQKPESV
NPYSLLDTSEPEPPV
ICESLEKTKQKISHE
ECGLGHRTVPLLLAE
SFIDGLATFQISGAR
SSAMESLTKHLFAIL
RKLDLSDTKSIRAF
PEAKCACTDSMGVPR
PGPPVNVTVKEISKD
VSPENLHTEEILVCG
VNIQLKGTNEYVPRF
PIAAVLATFVTHAYA
LESHMILTETLFRKI
RLTYTERTKSTITLD
LSRKLRETDSQLLKT
GDKRRVFTFPCLSAF
VLDEDEETKEPLVQV
SSQEKVATLTSQLSA
ESVSSLSTNDFSIPW
CFIKLCITLNEGKSI
SIC SARVTLREPPSF
KELNALETSSSAMDM
SSTCEPSTVAAVLSR
ILKAALITEENQQLS
NKYHWEHTGLTLREV
MRTQEDLTQGLLLID
DKEGRELTLEKPELK
HQYKLETIQKLSNV
QAEAVLKTQLQELKKL
LDCVLDLTKLKHNS
AKGDRAATLEEGNPT

LKGTQEITGDDRFEL
LVVFNHLTPPPDEH
VSPLLSRTEFCTAPL
FGQRLDETVAYEQKF
KRLQEEKTQEKIQEE
PTGPELDTSYKGYMK
VTPEVKKTSFHVTNL
NSSNFMNTTNFQSLR
SASKHSPTETEESAK
VEGAKKVTVTELLAG
ENMNRSETEACFFIC
EVCQKSLTGYLEKKR
GLSPGSLTLSAPSTH
GEQLGVVTNWPPSLE
LIGPCCATANLEAKW
AELERLRTQLLSHE
GQTSMPGTLVLCLPQ
TLQRTGQTILPSLNE
VNSSVKRTQIKVTHL
MASFLYSTALPNHAI
TVAKELQTLHNLRLK
IARILKDTAIKSADN
GDGDLMMTSFERMLS
RIFADGKTWSYTYLE
DLRKVYGTVLSRHHH
ALGRDILTKTKQARE
AMMSVDETLMCSFQI
XMSLADLTKTNIDEH
AAGGPKLTKRGLAAP
SPAARISTSPIRSVR
FKGRPPPTVTWRKDE
AALECYNTFIGERTV
MAEILSGTESVPLTQ
LLAPVALTCGSDGSL
YEGWRIDTYLDIPLV
QQMCCVSTSIVSFVR
YEGGAVCTHARSLWR
KFHGVLTLCLEVVL
RMVYCSLTDQKAVY
SQRVNGLTSSKNSQP
CSLRPNQTEEGTTPP
HESQISFTIEGPLTS
AEAKPKSTCELYSSQ
SESPVYPTKFDSEKE
VEKLLDCTVIVDSVF
TVLKSSATFQSTVAG
KQQKEEKTQEAAYAQ
GVCSEGTILHRRHS
LERQNLETFKDRMTE
IFESVHGTAPDIAGK
LGPADQKTGTPTPTS
KLCYDAFTENMAGEN
KKGASFQTVASLHKE
KDAYKVNTNLDYKKQ
VGKPGIPTGPIKFDE
QKSDDKVTLEERLDK

RDFVGQRTVTFSSLV
LEISEDCTYADV KAL
KNCREAFTADGDQVF
EPLPAKVTERHMQIT
ALMRSEETADLLAEK
IDHKEKSTEINHEIP
STERLKKTNEILKGI
CEESLSQTPPRVTGT
GPSSERATPAFHPVC
GPKQASFTDVRDPSL
KIQQLVNTLKRPKRP
ICELSFETE HILLQH
KKALLPPTVSLSATS
LHNFSIYTLLGKQVT
KDLVNRYTQNGSLDF
KGEDPLATDTRVSVE
LKANQVATGIRYNER
LQSHFIPTIGRLRKR
AFMDISATDLVLRKV
AFLRHKMTLISP IIL
YLRNLPITAHPEVFG
LWEGKPRTYITEERA
IRILCCLTFLVKVKS
GYGSNKKTKHMLPSG
EPALVPGTPKAEDRM
GYHVINNTFQSLG C
XXXXMWPTRRLVTIK
RFRHLK KTSKEAVGV
KLAQLIATCPPSKSS
LKDSDEITEDDNIRM
GGEVIEDTYGAGGGE
PKAAKEKTVKKETKV
PPEVLDVTKSSVSL S
HPSVVMMTAGRCHTL
AVPAAISTSEAAPYA
VDRHREKTLRLLWKI
SLKSLILTALQRETE
RVDALERTLEQKNKE
AGGSLKV TILQSSDS
TLTAYIVT SLLGYRK
GKPDVGLTTLRNGTL
TKQSFSLTMSIEMPY
LVSKELSTWKERPAR
GEAGGGSTAAEEASE
IACDNIGTPLAVFSG
RPCDEDKTDSETGKL
RTSAPSPTALKLATV
RGQPRTRTRASVRGF
VENMILLTIQYLVRL
VMADDEFTQDLFRFL
NVSQDLDTIRSNINN
SRSKADYTSHLRSLV
KSRHSDGTEKNKLPS
SVPKVEGTLKGPEVD
KEEVPEPTPEPPKKQ
ADGLQGETQLLVSRK

NKVQLMVTDSSELSNQ
ILYGPPGTGKTLLAK
CIEVVFVTKFLYSIL
EQSIWNVTVLPNSKW
GLNMCAPTDQDLITL
SLIKQIATKVHPPGT
HGREAKVTETARVPA
TDGSEIKTDEHYTVE
SFIWELITPTKDGQA
AAMRAFKTVTTKCSK
SDKFYKQTVNLQLQP
EIVFCAVTSNEQVKG
PPHLKPATEKLIVVN
EVKGAGMTEHYVTQL
CLKEIDSTLYKNLFV
SPNPDVWTSEQNPPY
EWPFLIITDLFLKSP
YEAQKRITQVFELEI
KAVVNEDTQGNVSQL
GFTFGTATLASGGTG
PSEAIEITAPEGSFA
AESGIRYTMPHVME
DALCLVLTLMNGGDL
KSLQAVITNLTQGEE
LQPEPINTPTHTKQQ
GTLVSHVTLRLLKPE
VNVRRDSTGMPLWLV
MFDERIFTGNKFTKD
KVLKWNFTTPRDEYI
DITENAATVSWTLPK
SSIQEQUITKANEEFQ
GSMLVSWTPPLDNGG
FLPLFKATINPDHR
PRAEKEDTAELGVHL
CLSVLSSTSLRLHSN
QSLQLSTCDVESKR
HSADVFHTFRQLDRL
HQQALPGTHIPEEAQ
SAPYLLATSCSDEKV
GLVVKAGTTVRFP
ALYSRIGTAEVEKPA
IALSSSETTKHATNT
NRLVDYITKTSCHLA
AEAQPEATAKNLLHS
FFLYSKLTVDKSRWQ
CVSSIFWTQEVSQL
ALQPSVLTMCNGGAG
YKDSRGLTPLYHSAL
EPKECTYTIPKLEGG
EALIKKLTQDNLAL
AIQIASATMPALSLN
QKEVEQQTGLSVFLP
LEEKAAMTDAMVPRS
NVIRSIQTDKREKYY
AGGHPLDTPHLPQEL
KVQIHHNTWLHFFPGH

KAEPLAFTFSHDYKG
KLYIEAPTFDLQGSY
DCNKAAVTIQSKYRA
TILEPLFTESESKIF
RHLKESQTESTNAIL
GDGFYLPATAGAPGSL
LSLSVELTEAKLHHG
PSQFPSETAFLEEFQ
RVASNPYTWFTMEAL
PSPFDCSTDQEEKIE
TSESIHTITPSVVN
AELENIATLCFKALE
EQVAWALTKFPCLQL
KNYSLKYTKYTKKDT
QHPAASPTHPSAIRG
QNAQVCQTNPEPPAT
IPVNEKDTLTYFIYS
XXXMAAATVGRDTLP
IILGNPATELSVATH
KDNSRDETFHFQCDC
YLDGNEITAIQKHGI
LFGDVGDTGERLLYN
KKAEAVATVVAVDQ
NLELRNLTVELEQKI
VEDNKRFTRIPLVQL
EALHELRTAEFKTE
QDLRVLPTIDLSTMP
LADDLKKTVTIRAGA
AVNNYHKTNPQTQE
INKFAEYTHEKWAFD
TIRGDLETSNLQDMQ
GEGECEQTPEELEV
LREAGSVTALVQCVL
VARREQNTKRDIPDK
ELHICFETTKSNEAM
LGHDEDNTRISSAGC
VVRDAVITSNGTLPD
QMLDDSTQLNSDLK
VKS VN SGTWIASNSE
RGMINVSTSSDMDHN
VRRQEDGTQPQVNGQ
LHPRAVSTSGCRVN
NIAVILLTDLIIDHS
DYLRWVQTLSEQVQE
CHALSGSTAELREDL
ESWLVPATLPGPGES
HFHIRCFTCQVCGCG
QIQFSQCTDRIKDLE
NKPVSFLTAQQLQQL
WGWRIERTREGDSMA
PPPAPRPTSRYNSSS
WSWSEQNTNEQQPDV
NLGKKPGTITTVGLL
LEVIYGDTSIMINT
ARLMSALTQLKERY
TINIICTVDYLLRL

QKREAEETAQKMAQL
PEKAKSPTKEEAKSP
DYESYEKTRTSDVG
SGNIKNWTSLMAAVA
PPARLAKTPSSSSSQ
VDSSQKPTPSAREDP
MLCBEFVTSYKALKS
LVLEIEGTVNSTTYI
SQLKSDLTSRDDLIS
QTKVVPQTVMATVPV
EDSGGGGTGVKKKRK
DQIIPRMTTQSPLNQ
KEIDNDFNETKISP
AENGKEGTAKGAPG
PRPLPRSTFARPAFN
YLLVLATDLGSPQ
XXXMFKNTFQSGFLS
EERCSFATVNIQLKG
TPLGPGSTSSLERRK
PSSLKGSTSLLVHPV
GYPNRDSTKYAEIYG
KVDQHEWTKCNTTPT
SNQPASVTDYQNVSF
KFPGPSHTESYSWPP
PYLSEVLTFCLEVAR
CQDGSNLTAQRAYA
KNLRSISTSNQPRKK
KDVNERNTVKGFRKF
NCQDKLETSLHVLNQ
RSAELGCTVDEVESL
HMHQNVPTTEEQRNKG
KKQSKKGTEEKKKPS
MPDPVDPTTVTKTFK
IASNSELTPSESLAT
RLQMFNVTPHHHQHS
RPDIKKATQAAKQAS
SKKLLKATGASKKSV
APKKPTSTKKPKTMP
ELKHHHVTLASQKE
GWLLEEQRKKQKPL
EVIYDAKTVNAVVEF
SECGKGFTMKRYLIA
IRLPRKATKTKKNYK
TPVPRVRTDYVVRPS
PVGKLTQTSKTTQT
QARPGEGTAPLPPPG
TPLQPQATTLPAQTL
KHVSRQDTLQQCQAW
QNEFLGCTDRLVITP
SLTQAAFTPIPVDT
KDPRQQSTPQAANSP
KREKEQVTEPAKVP
EKNQALNTDNYGHDL
PESIPLGTEDRISAS
QHKAVLLTGEQGTAK
DWVDGEPTEAEKDFW

IWIFAHVTQGQDPWI
YLCNAQITMLQMLAL
KTDMLMITRRGVLFV
EELGSAGTSARCPPQ
LEARLWETEKICQLE
TDSRTSLTIENANRN
GPSGSGKTTVITILM
IREIQKVTKLDRDPA
DSLEEKYTEPVILNL
LWAALLVTFLAGCQA
FYYYWKKTPEAASSR
PYRDSKMTRILQDSL
VGEISVNTSVAFLPY
VVVGEIVTEEHLRA
EFRIDSVTGAITVAK
RGASFGSTSVSGSEP
TTNKQTSTDGKEKTT
VEHLEDQTVTEFDDA
EGSDIESTEPQKQCS
QAPPPQPTQPLQPKA
SKSSQIGTGDLKILT
NSQLARHTSPSVISD
DRSQGGETLDECGLK
YYGLSGMTKNTFEVF
ENAEDDKTENQIPQR
XXMEDSSTDTEKEEE
SEAINELTVLVLVNT
MSLYASLTSEKVQSL
KAFCRMVTRKLETQK
QGEVEEQTFKEKELD
LDDNKLLTLPNGERL
EGYYEAITAVELKSR
SLHQEFSTKQKLLQN
VRDPVRDTPGTCHVT
KVEHVKVTVLDVPGP
KESDAMSTQDQHVLV
DSSNCSITREHICVH
NLVNKDSTRRDSGAY
TVKVTEATITGLIQG
TIVPHGYTEVANIPV
KTADHFWTDPSADEL
ALGEPITFYRQALL
NSSNKSYTLNGTKTW
DVSGILATQVYTWVD
NLDLDNETFHYYQPQ
IKRKQQDTPGSPDHR
RHRINPETDLPRKLS
GKDQEERTKAIEAFR
ELYEHWKTLMNRRDQ
VKGAAGRDLRLRR
LVDIQGLTFQVICSY
FNYLKDATEVKVIE
DVSRDGMILTWPPE
PNLSCKRTLSFQLLL
AVKGAPETLHSMFSQ
PSGADSKTVKLSKV

RILRDVHTALSTPGS
PPALQAPTGVTPAPS
AIDGGKNTSGLPAEP
LDPECKCTDDRLWEA
SHREKPGTFYQQELP
IKRQVVETMGLFHD
HPPLQPLTGSAGQPS
MHFESFITEKKQNIQ
PHIKPPPTTPIPQLP
ELAWHVKTITITEME
GNPRVLDTSRSSISI
DNEGISSTVITFMTS
KVEQAVETEPEPELR
RSQKSLWTIKSMS
KKLKHRRRTSETSIS
LSPVLPFTPAAMTQA
SSYKEDTVLHLC
QEI VQKRTFTKWINS
TPPQPSLTASPSSRP
LEDEGEYTCMASNDY
FVIDGKKTCKLKM
SALAEGSTYVPYRDS
DGSRCFFTKSKTDYN
AAETARRTREFRSP
KSLQWGITSPLLRC
PPFSPHPTLLSSTPK
TFDIEAVTLGTT
AEEFNDETNRGR
FHAVAWGTPGFLTVM
VYLGRDKTMQLLEEL
ENLNKLMTNLKSTAP
PFQMDPETGLITVRE
QREIERNTCKIKLFC
TCIHGVWTQLPQCVA
YLD FSVSTGMLAGV
IPEPSNITGNSITLT
NGVSVEDTINLLASL
TSSAQKITKSVKAPT
GSQFFITTKPAPHL
PTDEVPSTPLALEPS
SEADSSSETS VSEKEN
ERICELTYQAEEDK
IYGKPIPTIQWIKGD
YDGGSPVTGYWLEM
SYSENIATLQFSRVE
HSKKVLNTGHPLVTS
VRPARRRTPSPDYDF
INIEHALTFLRNRSI
AVSSFKKTSEMEASS
SGSQMGSTLSLAERP
DGERDPFTCPGFASQ
HGNNQDLTRNSSTPG
YYALGQATGLDEEKR
NIVHLSKTHFPIVQN
LRRIANVTLIDLRC
EYQCSSSTPSLNLTN

VAGQIRATQQQHDF
SSVHAQSTLESVISQ
ESVSSPPTLHMKTGL
ESCGSNSTNRSTAES
NEKVEEFTTFVKALA
VKWFKNGTEILKSKK
LACKVTGTPPIKITW
YWVMAFITKTIKLVK
KKISAAQTRLSELHD
ILRQLRQTVEATNSM
LIRSASVTPKINQLC
PSVICEHTKQFQTIS
NKQQNFNTGIKDFDF
PYPSEASTTFPESHL
DMAPLDHTSDKSLLD
QKIERVSTKVIRTLK
GHPAPPVTSARFPSW
ILNSLGLTISVSPSD
GAEIEIRTYDLKANA
RCAAASATVSVQVTD
VLGQLGITKVFSNGA
VSMHKFLTIPRLEEL
LELYLVKTSDSGTYT
KKDGFNITTSEKCN
VPPSSHETGFIQYLD
ASDKYRMTFIDSVAV
VDPSKFMPYIAHSQ
GSAEGGPTLTPATS
VQGNPGKTKKGELSI
EIEERARTEIISTDN
QLRDVQETVGLPVVF
VPHILRSTLPATALP
DLDIQVTVDPSPST
VDASFMQTLWNMNP
FPLPDEKTKKRIFQI
ENALQLGTPVLIEN
LLAILLKTTNRASDV
INNVRRTSHPTLPY
EKNFSFLTGANQHN
KQSQEEETHRLEIEE
PAAIAAATTLAQEE
LEGKMKGTKVKTPEM
ARYSIENTDSSLLT
GWGSPQATAMVLNNL
FGGLDVLTIISLGGIP
EEGLRAWTEGEKQQV
KAEKSSSTDQKDSKP
YSLWEHSTKNHLQLE
THEASKLTNHVELFK
SLFVRDATRDHRGVY
XXXXXXMTTQAPTFT
DAYLVLHTAKTSRGF
YQVGKEFTSKMDLEK
KKEQCSSTQSKSNKT
QINKRYKTKDDIKST
REKVASDTEEADRTS

LGHLAKFTCEIQSAP
HV VFGSETGEGENND
TAITVSDTVAVKETS
ITLPVASTASTSLVV
ELNRKIATQEVLRH
YQLSKEYTLDVYRLS
SLGGNAKTAIIANVH
MIKVRQYTVKEIREG
EFVSHPSTTCREQMY
DSGEKRVTITEKSA
SFLSSNSTFCNKQNV
DCKKLIDTFGIRWPE
PNRTGARTLADIKAK
VQQVHEETANTIQRK
KVAERKETEFFSSLP
WAVDRCYTGSGRVAA
KKQSIFTSVPDTPD
EARN DAGTASCSIEL
EMKKS VHTQVNTNTT
QELYLFDTTGKHLYT
DQLNEQVTKLQQQLK
ARLREVQTGWEDLRA
TIKAIKDTEAQKVST
NFKKENETLKSDLNN
DVNVIEGTKAVLECK
SCPSEPNTAALQAAL
VLPRVRFTRHGKELK
AASRPDSTHMVEARQ
KSDVPPLTTKGKKAT
AAQLKPMTPFKRTRI
KINREKATFKHPRTT
HCISLKDTNKRVESS
SPSADIGTTTEGDSI
PSRKS VLTDPAKLKK
FSVSICQTLMQDQQT
AEELSYKTAHEDISQ
LFVIDPNTGVITTQG
LDLGLSLNTSSHRTVP
WVNEKMKTATDEAYK
ISVSPLATSALNPTF
LSVSVVQPTTFQET
CSELSGNTKSRESIQ
VIFYRDKTEELMELT
SCGKKEYTIVVKVLD
IDQFLIATFMQLRLI
QQQAFPVTEIDCAQD
VKG GVDVTLPRVEGK
NSSPDEDTEVKTLLK
MATFECETSEPFVKV
FYLNKMKTFEPPFE
KIPSRFSTPEFTILN
LYHSEAF TVNFGDTE
AGCPVGQTRTNTFEQ
VYWDYAMTIRLEEIV
PLKRLN TW TNRYPD
SCQRDPNTGRLNGES

LILEGTFTGTPPISV
LQPTAGGTSKNEGTI
EAISETVKKVEYD
RRGRKPKTVLNGILS
RNPEGSRTLSPSRGK
VHERDVNTDFLLILL
KQRKDKDTLKAEKK
ELPLIFITPLSDVKV
PKPAVAHTAEREQVI
SSGNCLHTDGNETNV
PPDRFKFTSVTDSLE
GTTVVKVTPDLKPTE
GIGPPVETIQRTTAR
RMYLSCLTTSRSMTD
RARHRCQTPHLLAAX
IEYLMGGTYRCTYTR
QPRPPPSTQAPSVNG
MIKLRGKGLVRLF
AGKASCTTHLYIKAP
SMALYNRTRRISQTS
DIKWVDDTHALGVFS
VLSNVEVTLNVLADS
RVFKIKSTVVVKDLV
RELEIARTNQEHAAL
EDISPVQTDALDSVR
LCLEPPNTHRVSFL
XXXXMVLTLGESWPV
SLGVILYTLVSGSLP
VVVISGDTGCGKTTR
RVCSKRGTTESKEVL
DSEDESEIPTDKTV
YGADHPYTARELEAL
TSLHCALTEGLEILS
LGGNSYVTWNYAWSK
PRVRKPKTTPRKM
QYHRGEMTALREELK
NMVQRPSTVDFIDWV
NIEECSLTQQNKSTV
ILEILESTIEDYAQY
SEDDNFGTAQSNKAI
HQNVVAETLSALENK
ESDSGDYTCIATNMA
WKYYCLPTGWANFGV
EYDPSKATFVKVVPT
IYRGEELTYDYKFPI
PGEKRKMTDSHVLEE
ARAYPELTLAIFSEI
KCIDPGKTLPEALDY
KQLLEYQTMLKGRCP
RLHQLHITGKLPAGS
GDLPEAGTTRTFSVV
LNPAVLSTASLGSSQ
MTAFADRTVVVIAHR
DDQELQITDRIKIQK
ILGAKEGTQEFVHTE
DGEWVYNTFCIYKRC

AGAPPCTAVTSDHL
ELPRHPPTFDNLYEY
QNYHLECTETQAWMR
PAIFGVLTKEQLLF
RGISKHSTSGTDEGE
LLHQGLSTSHFAWAF
KVDMKDVTKINREKA
AVQLNIFTAVLSALK
DTTKVSFTYDETAGM
KAVLQRKTACQDTDE
TVQSKILTTGENYKT
RRNCCYKTNSHGETV
PDCNFRVTDVIEGTE
RRGRAGATSGELPVV
GKQPFLGTLQDNLIE
KDAASFLTRMKEKHG
CGNFDLKTINEMRTP
KPDDKAKTLSYEPLS
SYWSIFMTLLHFVAS
GDGPRSETKEFTTPE
SRESKMQTRILKQDL
CYIQHVDTGLWLTQY
YIIEKRETSRLAWAL
VSEPELATLSGDMAQ
AYDPEDETILEEAKV
SSKRRPRTRAQKVEV
SPDSPSSTPTSPVVS
LNESTFDTQITKKMG
EVRKSALTLVMGPLD
VYYRSKTTTHIATN
YVIVPLKTGLQEVEV
VGIDLGSTGDTTEAL
ALAVVDVTKRHVDLK
ERKQERETFSSAEGT
APKAPEVTTVTKDSM
HYLHLRATVVSQRR
TKDKKIITNHQKKNL
SDLDSNQTHSSGTVT
PFKAAPHTIHPPCVV
EGNDIRKTNVAQIRM
SPFSSPETSFRGHHI
IQRAAESTEEVSSLR
ANGEIAYTIAEGNTG
YRKHYEDTKANVHIP
GMMSDYGTARRTGTP
MAHMFVSTNLGESFM
QKICGKKTAFDPSVD
GQDYEHVTMLRDKFR
RMEIKDQTKCAASVG
EISSASRTCTFLPPF
RDGQVISTSTLPGVQ
KRFDYDVTMMHRIKL
DLKSLPDTELMKDTA
EQNMIGMTPTVIAVH
PKDPVIETIRNYFPE
QSLPASQTHMRVTAS

AVALVFLTGCQAWEF
DTDENVGTCPVSAF
HSNMQFSTFDRDADQ
GFSWIEVTILKPHLQ
KLKDPNTLCDEFKA
HHLGKEFTPSAQAAF
YILIGSSTSYCEIQG
LGANGRHTSSWLVT
GRIGRLVTRAAFNSG
DSRFGSLTSNFKEFG
PENLPPLTADFAEDK
LKSNSMVTLGCLVKG
FPESWLWTIEELKEP
EPKNLVKTNCELYEK
KAQEAGYTNATTTDQ
GREEEFFTYRSSVTY
SMILDICTRYLGDVD
VKDSCVGTLVVKGDP
YRLVGEATISCWYSQ
YVVLARLTAQPAPSP
DEANQDFTNRINKLK
QDIIWDDTKQTVRVL
CQNDIMLTATCKSDG
LAKKYEATLEKCCAE
NLLDNWDTLGSTVGR
SLKYNFVTPLTHMVV
DPPACYGTVLAEFQP