Supporting Information S2

The benchmark dataset $S_{\xi=7}(P)$ used to train and test the model for predicting the possibility of carbonylation at Pro site. It contains 126 positive samples and 792 negative samples. None of the sequences included has $\geq$ 30% pairwise sequence identity with any other in a same subset.

(1) List of the 126 peptide samples in the positive subset $S_{\xi=7}^{+}(P)$

GKWELYIPPKQNKSV
PLPRKRPPVTQAAGA
QRPARRGPPPLARAV
ASKLQVLPQKASERL
GYKVLIQPSNRRAIH
KVSAAPGPADQKTET
PASKDTRPKEKRLVN
GRKSTEFPRKIREQE
KQRAKKKPNPVEEEP
YYYWKKTPEAASSRA
QRNRETAPRTIFQRV
KSLQAKFPSNLKVSI
NSRTMLLPLLRRFPE
EPQKQVQPQVQPQAH
IAKQSLRPFCTVCNR
ERKQNLRPMPKKYHS
SPVQKANPGTLAAEI
PTLPKRVPRAGEARN
VVRLSCLPAFKDLIA
KERKMAVPMPSKRRS
LKKAGRIPEQILGKV
ASLKRLKPHVPLGRN
STKEKAGPKGSKVSE
VGIPDKIPFKRPCTY
YQDLDGLPCIVILTG
KTVMGAAPELKVRLE
AGRLFLIPATIQDKL
AKARVTKPKTAKPKK
PNSRFSAPSCGSSEK
VFKLLECPHLNVRKA
CSGKKVEPSALACRS
KEKQPTMPILKNEIK
RLKNEKEPMVLKLKD
STIPLSHPKIPRCQE
VAKAVECPPPRTTAR
GSETPDGPLSPGKME
KQATDVKPKAAKEKT
KKEEKEKPKKEVAKK
EVKKETPPKEVKKEV
KKPAALKPKVPKKEE
KESLKVGPLDSVTYL
LFNRIGDPSVDEDAQ
LGGKLKGPDVKLPDM
KIPKMKMPKFSMPSL
VDLHLKGPKVKGDAD
NMPKISMPEIDLNLK
KKSKIKMPKFNFSKP

PKGRPDRPRDLELTD
DSKKEEAPKKEAPKP
KEEKAKKPEEKPKTE
ADGRGMLPKRAKAPG
VVRPEEIPPVPENRF
RKRVSLEPHQGPGTP
KKATSCFPRPMTPRD
FPENKKQPYYYPPFD
STLERVLPRVRFTRH
VVERRRPRDSPTLS
YKPKRGRPKSKEMPQ
IPREKKPPRPPKKKY
KIHRNRYPVCTKPYA
ERSPGGRPIHKRKRQ
KGSDEKKPFKEKGKT
RALGVQTPSIRKSLV
ELIEELHPIIKEALE
NPKEEKKPKEKKKKG
MDPSAPQPRAETSGK
ILDLCAAPGGKTTHI
KHTFSVLPGDCQQRL
RVPRRASPLRTSRSR
KKKGRPDPYAYIPLN
LLKTHFLPLMEKLKK
KIKKAASPSPQSVRR
KKAKSPTPSPSPPRN
KKLLVLNPIKRGSLE
RRRAPPPPKRAPSTT
LPGPEKLPGSLRKGI
PDMLKNAPMLKSMKN
LMEKEVEPEGSKRTD
KLTRCLKPFENSRLR
SYRTVEKPPKFTEKG
EEIRSHRPTLDALRE
KVSRTVSPKSFAKAF
KSTLAIDPATSKEIP
KEPFKAAPHTIHPPC
KVPIKQVPGGVKQLE
PAKVPEAPKKIVPEK
PEEKVPVPIPKKLKP
PEVKPKVPVPAPVPE
RPIKPPGPPINPKLK
PFVVPDAPKAPEVTT
TQEGRQQPVPAPQQK
ARKALCDPLEEVREA
LRLLKPQPNIIPTVK
KKPLKKKPTPVLLPQ
VIRSLKRPPPADAPT
PRKRTHAPSALSPPR
KQDKLKSPFKFSDSA
GNGIISSPLTGKRQD
IREATAAPHGKRKRK
LLVNEQEPCKFLLDA
KTLGIKLPFLVMIIK
ESKKNELPDGLNKKR
LKLYRLLPMRAAKRP

ASQRCRDPPSNPVAA
RRAHTANPLHRCRCG
AVETEPEPELRQQTE
VYDAYLAPNNLKPVV
GLLYCDLPEPRKPLE
CLDNTRKPVDEYKDC
EFQLFSSPHGKDLLF
FFSEGCAPGSKKDSS
QNTGGKNPDPWAKNL
KTPKYFKPGMPFDLM
RNGQWSEPPKCLHPC
EFLRACTPGASLRKG
SHFRHGIPFFVKVRL
ENVGTPCPVSAFPSS
LSHKDDSPDLPKLKP
DKGACLLPKIETMRE
ARLSQKFPKAEFVEV
KVPQVSTPTLVEVSR
GTRCCTKPESERMPC
CVLHEKTPVSEKVTK
ESLVNRRPCFSALTP
VELLKHKPKATEEQL
KRVIISAPSADAPMF

(2) List of the 792 peptide samples in the negative subset $S_{\xi=7}^{-}(P)$

LCGMLSLPWIYSHSD
SGGSYVYPPVQEFPL
YIVPDYMPVVYDKLP
TTAGATQPAAPTPTA
QMANNFTPPSATPQG
KGSETYSPNTAYGVD
STNMGILPSVTMYNF
GKSTDFAPIKEDFGQ
HSWQEQAPNQPPGPT
REKLKAKPLLLQLPI
SSFIKRQPLVIARLA
VLGEYYTPALKCEVN
LRRVAGTPCGVFGQR
ALEHEYEPYSRLKSD
VIINQLQPFAEICND
KSDDPVVPSLFKQKT
SDFDFSAPKINLEIE
VGGTIKSPKNNEKNF
TGRQIFQPLHALRNA
YLSTEDVPLARMLVP
PEKEAKSPVKEEAKS
SHGEKELPQWLREDE
SPTASQAPNQPQAPT
LEKQFVKPDGENRAR
RFMEIESPHINENYI
RSLWRLEPLRISWSG
EEYIALYPYSSVEPG

DALKSREPQAKPQLD
AIIALFTPTTDPEAE
SSALGEKPITFYRQA
PIPKKLKPPPPKVPE
SLSSAGRPGPSEGGD
RLCNDSLPPGTVKLX
AELEIANPPELQKHL
NIVGIGKPSKVSECY
RDIPAMLPAARLPTT
VMTLKWTPNQLMNGS
KPSKSTEPILIKDPI
AELLANMPDPTQDEV
IKLNYALPPPLHQTE
VQGLGVMPPKAGQTI
ISQNESRPHFLIELF
PLVDHTSPSSDSLLF
SSTVNGVPSRSPRLV
TPEDQDLPPCPEDIA
RCISSALPRRRPPVS
SSISDHTPWLMPPTI
SLSHSACPTPNPLSR
MVVCWGHPDSDGGSE
SESSFGKPXXXXXXX
RQFDNLKPSGGYITG
TVLDANDPPIFTLNI
SWSRISNPSAFSIVP
SSIALGDPHIPTSPE
FNQSLKSPSRLGCPG
PPLLASMPAQLPPRD
IAQDPKQPPDPPVDV
ECHISGYPSPTVTWY
KASVVTLPVYLNFTR
NTSSDCRPSEESELL
WVKENKVPCLECNYK
SWLKDGLPLKESEFV
EEVDCMDPTCSGRGV
AESPEVLPHIEKELS
SCRSEGTPAWYMHGE
VAFQLHLPLSQRACS
DRFCLSSPTEALKMG
QVRVVAQPSLPAVPQ
WIRDALRPPLQNINS
HKESFLAPVFTKDEQ
AAVQKAIPMYKIATK
WGMESKIPVEDNKRF
AQNPVDAPGRPEVTD
SEDSVQVPRNLVGKV
YPDIGWNPCGGERYL
IGSKAGVPSKSSGSA
SLEPHQGPGTPESKK
LWGCHSAPHGPGLIY
EIGDGFLPVCSLGPG
GAETLTFPCDRWLAT
LKASVCKPFLFLLKK
ESGKIQEPFSAMSKS
GTQVSQRPGAQALKV

KEQQKDSPVFCRFFH
QSRPTGVPTPTSVLS
EMFGVWEPLLEPLEI
YTEVANIPVARPGRR
SRLDLDLPGCQAEPP
YLNKKGNPKKFAGLA
AIESLVKPQAKKKAG
ENIDGILPLHDAVAN
GKLYIEAPTFDLQGS
VQIKKMEPDEQLRND
KRKKKKRPEDTAASA
SERLQRAPLKSVGPD
DPRQLGVPVIARDIE
ELLSTTHPANKASLT
SVTVTTIPASQAMNP
LSAHTILPGSKTREP
ALKISVAPGLADQKT
DRGEIIDPFVEVEII
HKDLIKKPTISTAVG
DRETILDPNLQATLX
DGAVPAAPASADAAR
VLAERKSPEISERIV
SPHPPLQPLTGSAGQ
FLPASTAPMQGKRKS
LNNSSSSPQRSVDQR
KNALENYPNFRSVVD
ERLALAGPQLRPEMS
LKSQTQHPWEKLLNL
NANQNASPNVPGKRG
SLEMGILPREIRKLV
EATPTAAPPTLPPTT
EKHTHHAPLSPRTFH
QYRKVLMPKLKAKPI
ALSELDVPFKVKAGQ
DAEVYNVPLDSQSDD
DKLTLKIPWKNLYGE
LSPTLNTPAPVAMPA
TRSLGEEPVGGLGSL
EEQVNSLPGSITKAG
KVVIGFVPLAEIMGY
TYIELMRPVSELIRS
PDVEAHGPEWNLKMP
YFPNRKFPSSSVAFK
EGSRLSLPRLIDMSA
PRLTPVRPAAASPIV
PQPGLVVPVPTVRPL
LLVTAMGPPGGGRND
KIEKELKPYGSSAIN
ENLELTNPQEFGSSW
AVDKRFLPDDHGKNA
PSAFSIVPRRAAKSS
RLLSLSKPILPQQDN
IKRQCVWPFIVMMDD
ANLHPAKPKDFSAFI
TQITTESPEKTLFSS
SDCCGQKPTGPGGPL

ACVTACEPPKTCQDG
TMATRTSPRLAAQKL
ALTALTTPIQTAALQ
KLQSILKPMMLRRLK
RKKMNLKPIMRMNGN
EKAFEGSPARELDVP
GSSGAGVPGGAAAAS
SKQKKVAPRPSIPVK
LPGSVLSPPPPPPLP
GDSGGLIPGKSLVFA
NKEEGARPGTLLGTF
RRTEGQYPSICPRLE
HEAIILKPEAIEMYI
HPFASFHPGLNPLER
PDGLKLYPTLVIRGT
PQRPASEPHVVPKAV
EIPKKKVPEERKPVP
RLFNDSSPVVLEESW
GDRSGIFPSNYVKPK
QAAIIQIPKPASNRT
EKRDYREPFFRKKDR
FTNTSLHPQRMKALA
PQFSSLQPPCFPPVQ
ACRCSIFPDLSFVTF
LMDYADLPYQIEDIF
FIDDINMPVINEWGD
GTLVLCLPQIKIISA
ISTSEAAPYAPPSGL
TPSWLGLPNNAERVL
DLEDESTPIVKLGDA
EDVVASLPLCHAALR
RTHTGERPFQCHSCG
GNLHLLSPGNSARLT
EAIIDVRPASTRFLP
EMSDSNSPPIPYSQK
DSVGVMSPPLSRSSV
PRSVSRSPIRMSPAR
NMVNGMGPMGTEGLF
TNTSVLGPMGGGLTL
TQGSEKCPQKTTRRD
EESQHQQPDDSNNIA
FAGTEIDPENEELML
KVTIKDKPAVAPATK
NHDESLLPESLESMM
QHLPSVSPSVSDAFL
LNRDLSTPGLEKDSG
CEELRAAPRQGPAPA
PSKEEPSPVKAEVAE
NSSRQERPPVKPFIL
GTEEKKKPSDFKKKV
GGESSASPGEPQRTL
QADRKILPFTSMRHQ
RARTDEVPAGGSRSE
RFEMEGLPVDESDSC
LLSQLGQPSIFDTQK
FDAEDGEPQTCKSMK

AASLAELPLPTPPPA
ISDTTLLPDLSQISP
NRFTMEVPKKLTEWF
TPESSSLPVALQTPT
ITESQSSPPPVIDLI
PLLPSSEPPPAPPLD
EMVSGFGPIYNYKDT
VQVLKYEPFLDCALS
LGRASVVPLPYERLL
KKTKGDNPNSHPEPA
ARNLFSVPGAPDKPT
AKVMELLPTHAFSTL
HAVAWGTPGFLTVML
TTSQDEAPREELAVQ
SYQEAQLPALPESVP
IVEREMAPDFELDAE
ATLEEGNPTDEVPST
LLKPEADPRPCERAP
PKKVAKSPAKAKAVK
SRRRPRAPKEKAQPL
IPMLYVVPRPGKAAF
PMNPIPAPLPPDIPS
GNDSDEEPPAAAPTP
DKVVIQDPYRLPGPP
FAKSDNQPSTEKAME
IAKEKRTPAPEPEPC
SAKITSLPPPPPTLF
LHLQRINPTTVKMKS
LDIKAKAPKVKMPDV
ESGDAPRPPPAATPP
VDLHRFHPEPYGLED
RRMQQAAPTDPDLPP
KAVSSADPRAPGESP
TIELEWEPPAFNGGG
XXMNTSIPYQQNPYN
GALMTLGPSVVRYHL
VTFKANRPFLVFIRE
VKVAQRSPVDSGTIL
PMKPLLLPQPEVLSP
NLVLKRVPSPVLIKK
MDDLISSPGKNKSGR
PRESACLPEKLKEKE
RAWGCAGPCGRAVFL
NWRQPELPEVIAMLG
MDRDCRMPMGLSTGI
TVENASKPDFTKNSQ
RTDLYFMPLAGSKLA
DRGLQLVPVWDIILS
NALNSCDPWVQAKCY
TNMGIIIPDFARSGV
SKTIYTAPLDMLQVT
RTPPRRMPPPPRHRR
XXXXXMHPPKTPSGA
YYRQSEVPDSVYQHL
EKDDSVRPNMTLKAM
ECLGAAVPARLRKVA

RSLQKDFPIQILSID
VDSDVAKPNQACIST
SLKKAKKPPLPPSRT
EKFRIELPDLGRFYK
AARISTSPIRSVRSP
AFPDGHLPEEALKVS
ETYTEEDPEGAMSVV
ALDDAEEPESPPPPP
ECTTIEHPFCMYDAD
GAAGSMSPSRVEANV
EDVQVLIPFKGRPPP
EKVSEVLPIQKNDEE
LKFSSMAPDLDRLNE
HPQRPPSPHPPPHPS
NRLCFLEPTSNAQNV
TDEILLNPNNKWFKP
GKLSNWEPKDDAMSE
GSGTDMEPSLYRCSN
FKSDGIIPFAHAIWH
DMLKRVEPLRNELQK
FGYIKIIPFYILNAT
GESKQHIPEKKNEVH
ATLSVLEPATIVEKP
LQGGNKSPVTLTAYI
DIIQVTVPDPSPTSE
ALGNQSTPAPPTGEV
FTNFMRSPACDIFNP
PDVSMVDPEALAIEQ
TSLFSNKPFLKLGAV
LLCGSDLPLHKMAIQ
CTVLTLEPNSQVQQR
YNGKLTGPAAVELKR
GDASIAAPFTSKLSS
IRDWLVKPIRDQHVK
SQVFIDHPVHLKFNV
AEDVSGMPALCSPIS
NHARLRTPPPPLSHA
LSHTLGDPIKIRAWQ
FQELSQTPGHTEELA
LGRPPSAPADKDGSK
GFLSYLGPFNQIFRN
ESQGEVQPNFSTSSE
LMLTAGLPELTSVKD
RPGPHSVPSQAPRLE
ILDEFYGPEKSLQPI
KIINFKTPEDARWIL
DREDQEAPDSDAEVD
KQNGIPHPRGSTAIF
AGQVAVLPEVQVTQN
LAELLRGPCGSFDVR
KAKKRRLPRGFPPSA
LQRAFPAPAACQCHC
VTRQLYEPLVMQLIH
EWFLRTFPDPTSWYS
NPKAITAPQMFGRLD
KLNQHDSPRIKALEK

MMLCSECPTVFVDAE
ILKKYGIPFSRITQE
EKTAAPSPSLLYKST
AECWRSRPEERPTFE
TLTYTAKPVGGIGGG
VTSSQVQPVTIQQQV
CQTSTVFPLKKKVSI
RTVKDSTPSSLDSDP
QVTDANDPPAFHPQS
LADTDLSPMDGISTC
EADASGLPHKLGEED
KKIIKEQPQPSGKQE
DCFSEEAPDATKHKL
SGQRSASPSVPGPTK
KTVKLKSPVLSNTTT
QKSLSLSPGKXXXXX
FVASGMGPSASSHGS
LGCTAATPPSKDPRQ
QWIRNLVPEFGVSSS
MRKFLSDPQVHTVLV
DRSLESNPEQLQAMR
VASSEAKPAATIRIV
DLQVRACPFLPVKRK
ESVHGTAPDIAGKDM
XXXXXMVPGEENQLV
FLRMSKVPTDLAVEE
DFLKRRLPKLSKSTA
QECKTKHPHHLSLKE
VLKFSCKPGFTIVGP
RLEVELYPSEVVEIQ
TTKSSAVPPGLPVYL
QLYDDSFPMEIRQYL
KKDTKKYPESTDTES
SHGDVIRPLRKQVEL
ESGLMGEPQPQGPPS
GHETEEFPDEEQLRR
GDSRGPSPSYLGGPR
ESRYQTLPGRGLSGS
DNANSRLPEDTTSVL
KAAYLSDPRAPPCEY
IGFLDNWPLLEQWFS
LFQKKIDPVTMDPEK
SVAQTGGPPEADGLV
EMEVCKLPRLSLNGV
SVQSSHFPSGSYSVR
ITEVWGIPSPIDTVF
SAQDMHVPVPKQLAH
SKDIPLSPPAQKNDP
YRGVKVMPFSTACNT
YQEKLYRPPPVLDAL
PAGPTESPASKGVTA
WVRCNKMPVKDTTYR
EKECAPTPAPVTRAK
RHAVKGDPQSPGRHW
TASLLAIPHTPESSS
VEAFFYEPYQGAIQG

```
FMAVKNWPWMRLFFK
ENYVQETPLVLSRCS
MTISWHEPLSDGGSP
SAEKNQRPRKQTSAP
FAASSTVPHCEQSCR
VATPQVSPNTVKRAG
EEDLLDDPSLEGMCG
ASYESESPENQTEKE
WQMWTLLPGFCTRPT
RQLDRLLPKLARGSL
GKKVRKMPVSYLGSK
ICHEISEPRLQNYDE
IVDVLDVPGPVGTPF
KKSRFWEPNKVSSNS
QRFCIGHPTQEMTFG
RARSLKFPGLISGCT
LIKEQLEPPEIDMKN
HSDSGYEPSFGKSSR
GLAWARAPQPKAWVS
ASSVETKPGASKVAT
LIRICSKPVVLPKGP
AKKEDKTPIKKEEKP
GKASIKDPHFLNFNK
MTMMSHIPSVMKAHG
ISELTGTPFDVESDS
DYALLTLPAEFSSQL
GATAPPAPAEPTPPP
EIYGKWSPLILNPSY
EGKSLRFPLALEEKQ
NRFGVSKPLESAPII
PQLSVCNPHSGKEYF
SNRKGETPLKVANSP
YGIWKLIPNGQYEFR
GGAEEERPGTPELAP
APLNVVPPEVPSEEL
ELVQAMFPKLNNQER
GTELSLNPTCAVFIT
KVFEEAKPNSELCCK
VTSAISLPNIRKPDG
CDEDHQTPRDGETSH
VNYCECNPCFNGGSC
AFYQQTLPNSHLTEE
VYIESRRPNTPYFIC
YLAQQIHPVVARICE
VEDKSDPPEGSEEAA
XXXMTAEPMSESKLN
KSVHLQKPKPDKKKF
ADSVSLLPAAPEGSR
GCEGSSKPFFYNRQD
RRREQRDPTVHDDVL
LDGEEEEPEEETTLG
SGIRAGYPLSERQQV
MKRESQNPDQKDETN
NKVGASDPSDSSDPQ
MKVVTLKPTVLEHTV
ITPKSDVPIQAPHFK
```

ENLKIDRPEDAGEKE
ISDHSETPNMELSCR
SCSENQEPGYCTVSN
SFIKQNNPKFSAVQD
PEDDGGSPITNYVIE
TLVSGSLPFDGQNLK
LARYSYNPFEGPNEN
PASTRFLPQGTRIAA
LHEIYFLPDHPELKK
LNDRMTFPEELDMST
GRIFVFEPPPGVKAN
QSKLYHLPPPTVGPH
IKNSNGIPKLKYIPL
FLEVDEYPEHIKNLV
QETAEGIPPGSQDSS
XXXXMSEPGGGGGED
VYCLALVPANTLPKT
RCFLEADPYIDIDQN
ASLEVEVPAKIHLPK
KKVPAPVPKKEKVPP
QVFKPFLPSIIALCM
YSVYGTEPYIRLGVL
MSPARMSPARMSPGR
PVPEEKEPASSPWAS
KKTFVGEPQRLGSET
KDMVKMFPLLVEKLR
LGAMSAAPSQPNSQI
TDRLVITPLTDRCYI
SLRQNETPQPAAQRS
LSKDLLKPIQDVNSL
TASPSSRPVASPGAI
KSQVRDYPKHNGQIS
GKDVLRLPPSSITTD
GLETYLGPLQVAYRE
REYQEYNPYEVLNLD
RVELLHNPAFCSLAT
SVVSITGPLIRILGD
PEQNQDMPPHLPTAS
RLTPIGDPTMVVEWL
TCRDEYAPPKAELDA
GDEDIKIPENPLEPL
GEIRSVRPLDREKVS
LSPPPTEPTDGEQAG
KKELKVGPYANTTRY
GKDELSKPSSDAESR
ATLVDQSPESLKRKS
KLKVLDKPGPPASVK
CTDSMGVPRALGETW
ANGDSDGPSYMSSLS
LGRVFIHPESPSTGH
SLKTAHSPNVFLQQG
DKIQQMNPPEFEMIE
ALLKDQQPGTFLLRF
AVGTSAAPVPSDNHX
EVKISHFPAADLGFS
CVDGCYCPNGLIFED

```
EGVLLTLPRQSGGSG
ESMENSCPQVTEVTA
SSSGAVDPSVQRSFT
XXMMAQFPTAMNGGP
EEVRVRSPTRSPSVK
CLEENERPLEDQLEN
SSSSGISPDNRDFYQ
ATSTSSEPLSSNQPA
TTCNCKKPDDDTRKG
MLTGESVPQMKEPIE
VPGGSVKPKIMSPEK
ILTTLQEPHDKHLLD
TNVEDVCPNIPSLEK
RLEYLEGPTVTSSYR
QSPTKAKPKVEDEAP
PPPSTQAPSVNGVCT
HGDACFLPEECPCTW
EVSIGSAPLAKQQSY
NILVELDPDGCPWLV
DDTERRSPTPERTRP
GSFGYKKPPPATGTA
XXXXXMAPVHGDDSL
APQQEELPLSSDMVE
KKLRSDEPLIDPKPV
AFTRTVSPGSVSPIH
QLSELADPLKDELNL
TPNPLERPIKMGWLK
FKELFQTPGHTEELV
EASKEYSPCVAPCGR
KSFATGRPKKTKKRS
HPYVTHIPSPTLPGS
DAWLQTDPEIQPSPF
SEIKISNPTEFQNHE
HNIENASPATVSRMG
SLESSSFPLPKQYLD
TGQLGVKPQTGNSIP
CSPLMADPLLAGDAF
FVYKDIKPGRKRKKH
TSANEVSPVSSSGVT
FQRSRSEPSPDAPES
ADSDGWTPLHCAASC
VEENNASPHFEPDLH
NQIGCLTPELSELKK
ETQAEDHPPRLYGCS
FKISFGTPAPGFSSM
GVCTDGEPSQPLLGQ
LSPDSHYPLEEEKTD
SPISQRLPPKVESLE
AAEFISKPQNLEILE
NKCLIKNPAERADLK
ELVAAVWPYRRLALL
RQTVAQMPPQLVELP
NPALTLWPMFLQGKD
QDIAKDFPRGEESLE
VTLKWAKPEYTGGFK
QHVFSAGPVSDEVLP
```

LYPPGRSPLHHAQQL
HLSSKWLPSSPSGVA
QLAPDPGPAGHTLFQ
GLCAALEPILSAKTK
PADVTGLPMPKIEWS
QKFRFNHPAEAAVLR
SAMEMNEPQAPAPSL
FHKVKALPLVSWSRR
AFKRANDPVNYLESK
VRATRASPDPRAYVP
KVIKMEEPLPAKVTE
QQVTAEIPVDLNTRE
GKVIGEFPGVVHCLD
MIKKQDQPTFDNSGN
SASDGYCPREHMLPC
LKHSPEDPEKYSCFA
LIKKHLLPLWNDGCI
SSSLGKKPSLTSESS
GILKRRYPNSLPALI
SKRQLQTPKEKAQAL
VYVDYPGPASLTQIY
AEEDLTDPVRCNSLE
NSAGRSAPRESRPVI
FQEFLKRPDHKQDFV
QAESQVEPSYSEQAD
LTAKQKLPSYILENN
RSTEGIVPQYDGSSS
LSLSLAQPLRSERSG
EMKEIERPFETYKEN
NATSAFGPNLRYIVK
VDQKIEHPLQPQPGL
HRRTSAIPRAFTRER
GDFKVTTPDTSTTQH
VISRTGVPQPTQAQS
RGKKAKNPKEEKKPK
YTLKKDVPDGVKELS
GNNDVKQPEGTGMID
VESEIKVPDVELKSA
KLGVQVRPENWCSQM
SWLKDDEPLYIGNRM
PAYDGGSPIIGYLVE
LAKILQNPITHSLQV
KLSSQALPSFGYIKI
PPARQLGPRSPRVGR
SKVWYNCPXXXXXXX
GKDFERKPLLEMNDP
GQAVTLLPFFTSLTG
KLAHGQEPWTTDAKI
PGGKEEAPEPPDGGD
QMHSTSDPSHRSSSP
LKNKGIQPLLDAVTM
SNIDIRIPTGQLTMI
CDEHFETPSGSYWDH
IDKNTKIPCKSPPPE
NAWERRAPLAPKHIK
TDWEKTKPVTGNLRP

CYHFGLSPDLPICKE
EALVIREPITVPERP
GDMGDTTPAEPPTPE
FVRPDLGPKMYNAYG
DQKYRQHPSNFQFKK
QMSFSQSPFLSIAKS
FSALRALPDMEVVGL
TSPEGVAPTSGHDRR
GSHVNQGPLHLGGIY
GRITDVIPSEAINEL
VAEYEVTPDDKRKAC
LVRDPLSPAVRQKET
GNICVYCPGGPDSDF
RLLERYIPKHQKCLT
PLTTALNPPVTATEE
FLTYHMWPLTFLSPS
GHGIMEEPELTLIST
KLDSEAWPGVLDSER
LRISLTHPTIPFNLM
ENSIEGTPSKCYCRL
NESKKTAPRQEAIPD
INPSTSLPLIGSPPV
QHRTFIKPKELEGLP
RVRVVGKPDPECEWY
LQALDLLPLLIQTVE
WAQQIHSPLTCEQLT
ITVITLGPPSKPKGP
AWDSGQGPRLPSSVA
STIGLNQPSTPTHAA
GEYDVTMPKVESEIK
FKTSIFSPMKKEVKT
DEREFSGPSTPTGTL
GDEGADEPRGAGRKA
STSKNTSPMRNSFAH
LRPHVFMPEVTPDMD
VDLSLGSPKLKGDIK
LLVGAKSPNLPEHIL
ILDGLYGPIAFKDFI
ARHRCQTPHLLAAXX
AAAARTEPYCAICTL
LFHILVTPLFLLSTV
AQILILEPPYFVKQL
PPPGQLDPEDQDSEE
CLVKDYFPEPVTVSW
TPELSAEPTPKALEN
TAALAVAPGPRFLVT
GRPAEPQPEPEEPAE
DLRTDAIPIRAAKAA
AGIAAIGPAKELEAE
EEITDNYPQAIVYPF
APPGLQPPAEDEARA
LVIPPQPPTTGPPRK
SRGTDGQPLLLPYKP
KVDALHCPPAAVVTV
TMKMRILPDQFSPLN
QVTTVKVPMMKRLGM

KKACMFEPLPIKLSK
RQQNTRDPDTMVPLT
EEALKVSPVSIPAEQ
IWWANKDPLPPLDPQ
ELKLSESPEPGRSPE
MMGLDLQPDLCSLLI
TRLFVKEPAAFLKRL
FDMHKRSPLLRAPRF
ALGRTYHPKCFVCSL
QESARAVPEKSTAPL
VVLCNLKPQKMRGVE
DASSSWGPQMSASVH
AGVSSGAPPGRNSFY
PSASERLPDEKVELF
TVSQFTLPKSVSDGI
TNSKATTPKPQKPTK
KESKVEEPRKRETVS
ALSSVGSPKAKEALN
EGQTQRQPGSPMYRR
KKPKTVKPKKVAKSP
AAVRGSEPISVTWMK
QFPFADTPDTYKSEA
LAQADKLPMTDKRVA
XXXXXMDPPRPALLA
PEVRSTEPVAPPEQL
EDEAPPRPPLPELYS
EKPVDPPPFMFKPVK
IQTDVVLPSCKKKAP
ALSLFTVPEDCPIGQ
SQPRDGWPMMLRIPE
EGGGRSLPRSSLEHG
TLSISVQPGEGNKAA
TKSNDTLPSSHREIP
ASAEGREPSPQCGGS
SGGEWKGPQVSSALN
YTIIQQMPRGNHFTI
PGQHLLAPEDVISCC
SLCVNEVPSFYVPMV
KGVTSILPVLREGVL
EQQAQFTPLADPPDI
IPLLLPTPEEKKPPP
PPTLSRTPEVQSRVP
TDQKDSKPPEKATED
EAMGLGTPSEAIEIT
IKTLSDVPAPRRLAE
TLLIELLPWALLINE
RIEEGERPDRAKGYD
WARVTKDPIHPYTKF
KLLVNFDPKILEVVR
ESKQATTPASRKSSK
LVKMKFLPVDDNFSL
NIETSNFPKLLLALF
HLDFLKQPLATQKDL
LNKLSMKPEPLFRSV
KDKIHTTPDTPEIRQ
DAAHGYSPRAIDMSN

KTKEKAQPLEDLAGL
ECQITGTPKIRVSWY
FGTLEVMPQNESETQ
ESEKSDFPPTPVAEV
FVNKADFPKVRAKEQ
SGQLVVIPSDHLPHF
ATGRPWLPLHTLSVS
ADLGVSGPKVDIDVP
WKQVHNYPMFNLLMD
PPCLQQNPGTMQGVY
GSAGPVHPLVDPLTA
AHLFSWIPLSASITP
LKWNFTTPRDEYIEQ
VKKTDSQPTSVMLLD
EGCANRLPNGHVNFE
RGGKKFLPVLKEILD
STCSSTFPEELSPPS
RSKLEDDPLYTSYSS
SVVSQRFPQNSIGAV
SLLALGSPMYSIITP
ASKGEMSPMVVETLK
HCLSFTIPRLPSSEE
IDKTCLSPTPTLEQH
TIFQERDPANIKWGD
NTDGLFFPVMSFSAG
KADIGCTPGSGKNYA
MTSTSLSPQGISTLE
KTVGVWSPSPPTCER
PRERSMWPLRKCPWK
VDYASKIPAQGGTNI
NRLCFLEPTSEAKKI
ETERCLQPLQMMALH
AVRMKIRPLVGQXXX
TLVVKGDPRDNRQPA
LLRRIEHPTAGNIEA
HVLLTLKPYYAKPYE
DEEVLTDPKIQALLL
LISDTKAPKRQEMES
SSVTFEYPSNAVEDV
TSSWLVTPKSLGNVN
KVYDRYIPPDLCVCN
CWYSQWTPAAPQCKA
SNGYKPAPLDLSDVK
LGVAEGSPQYKKWYF
IALTKLQPLLNKDSP
DTWVEHWPEAEECQD
GYLAATYPAVGQTSP
GGEFEMMPLGVNKSP
LSAEAACPMFNFSIR
YLCSAVLPLLTRCAP
MDVSLHLPSRSSPTV
TIEELKEPEKNGIST
AIDRFNAPGAQQFCF
GGFQVLLPLVDQYFT
ILLTDGEPTVGETNP
KMLSGFIPLKPTVKK

EELTYSVPYGREKDV
SLKFKEKPTDLETRS
DLGQGLMPVLEAWKA
SRLANRAPEPPPQQV
RTKKEGIPDARQATR
GCKEICCPVPDPKSV
FSHIDVSPGSAQVKA
IILLILCPEIIQDIS
KAHIQFKPTLSQQRK
MGNCGHKPESPTEVE
EKASKEAPVKGLLQT
LKNLLFNPSKPFSRG
RGGKGISPPSFSSYL
SRTQWEKPQVKESKR
NSALDWWPKIDAVYC
GVQREDVPAADLSDQ
YRDRSFSPGSYERET
KVNSGAFPAPIEKSI
AIMRYGMPPQDAFTT
KEVKSTAPEATVECA
SEGNIDSPVSRFMDR
TLGEKAKPALDDLGQ
GAAFSQKPQSNALIG
SNQAELLPSKSVSLI
LRNDKDKPLPPLLAR
EKSLSHSPGKXXXXX
GYYCLKVPLVVFKRE
FYQEVANPLLSSVTF
ALLESSLPDTDPVVS
EKCLAGQPKDTMRLD
GTSKAGNPIFYYVAR
YSRIKYDPDLFRMAL
TQKTVDGPSGKLWRD
EAERSGIPIVTSPYQ
TNLKIKHPTYCYEMN
YAAFNNRPPSTWLTA
ICLAQNSPSTFHYVL