Supporting Information S3

The benchmark dataset $S_{\xi=7}(R)$ used to train and test the model for predicting the possibility of carbonylation at Arg site. It contains 136 positive samples and 847 negative samples. None of the sequences included has ≥ 30% pairwise sequence identity with any other in a same subset.

(1) List of the 136 peptide samples in the positive subset $S^+_{\xi=7}(R)$

ASKLSKIRVVRKSIA
IIIQSYYRMHVQQKK
HCKAFKIRKHYLHLR
IHLELRNRTPAAVRE
LEELDRERCFLLNEI
YEMHKMTRDPLNLAL
RPPVSGLRRRKPRAT
KTQRSPVRIPFMQRP
KQFFKDWRDKDQSDG
KSEYKSYRLRAQLYK
KSRDEGPRLGDAKLK
SLRKLPPRSEISSEK
RKSTEFPRKIREQEP
GSVTGTKRLRCMPAP
ELQDMAARQQQQIEN
RKLANAPRPLKKRSS
ASSRAHRRHRRQAVF
RALLTAPRSLRRSSC
TMLLPLLRRFPEQVR
SIDSLKQRLNVAVKE
EADEEKERILAQLRE
VADLEKQRDCSQDLL
EEEEEILRLHQQKEQ
IELLRKVREQVLKCP
LLAEVERRIRRGHAR
PDGENRARFLPGKDL
WERKQNLRPMPKKYH
RVLSGRWRLPLRALP
VSWGKLKRRVKGWAP
TLPKRVPRAGEARNL
EGGHFKLRVTISSNN
TLVEKEMRVTLAKLL
GDSRLKMRAELGEYI
REQEPTERPLKGIQS
ALAKHKERKMAVPMP
KDLGLHPRAVSTSFG
CLEVVLCRVEGMTEL
IKGLTYLREKHKIMH
MQKLNQKRAELKLVV
RIPQLWKRVSWDSST
GYKRKVMRMLKRQVS
ALVPPWTRVYRVQRD
KCSEETFRFELGGGV
GAGSGLGRLLALQFA
SQDKMEIRSCLPKLL
VDKSGVVRVRVEGDN
KERLQEVRDKLGLQY

YVGKAATRLSILSLL
RSVFTVMRRQLLEQA
QRGLTEKREELLSVP
RKQLIPIRHTKTREK
LLEAEKERNNKLSLQ
STEPLPGRKQVRDTL
RQLAIETRKTLAAQT
LSHPKIPRCQERLLR
RDRPGDPRRKRSSDG
FKDTDMKRLSMEIEK
LKRSLEMRERVLGPD
DRGHEELRKLESTLD
SGILRQLRQTVEATN
TAEDLRTRKSKLQEL
KAAAADGRGMLPKRA
PPSAASSRYSMRNRI
SRSYTRSRSLASSHS
HIAQKLFRDGHFNPE
ETSLNKAREHSLLRS
HVTKTRRRCKTVRVD
RLSSVFLRVRTNVGV
GQDGSGRRGPQKRGV
PRVFPPRRPHTPAER
APYVTLRRGLNAESS
TSQGKLLRTILFGVK
IYKPKRGRPKSKEMP
ENELTFLRSQIAAIV
IANVFQNRDYQCDTV
LMEFITSRKRGPLWN
VVRPKDRRVGTPSQG
RRSVRMKRPSSVKSL
PKTMPRVRKPKTTPT
HFHTVIARRAGMAAS
RIKNPKARVTEVMKA
LFTQELARRLKGSGV
WDATEDLRISRTDSF
ELIRLTQRLRFHKVK
LPLVSWSRRLEFQGE
FLQVETRRAGERLGW
LSYRSLRRRVRRLRR
VGLGLPNRLRFFRQS
GLAARLQRQFVVRAW
SLRKGIPRTKSVGED
ETRQRLLRTVKKEVK
AIVRQRERESHAAAL
ASPWKSARLMVHTVA
LGKRADQRKAKLGDS
IKSMAASRRAKLNES
QRMKALARALPLQPR
SASKASVRSISVEDK
EEELLNVRRELMSVF
AQLNIKERLIPPSFT
SDPIKACRPIKPPGP
HHVVSGLRENSEYFF
VVESSVLRAKEVTWY
LKRTKKPRPPSLKKK

KESLLQDRRQEEDQR
QEERLEQRLKREEVE
PEKENAVRDNKVYCK
EKLHLITRNLQEVLG
HGKQPSLRAAKEHAM
NCILDDKRPHVKKAT
NRRTLLVRPISKQDP
QADNPLGRSVLRKDI
GPPTGAGRGPSQLQA
KVLKNNIRFMNHMKH
SRLSRHRRAVHGPPE
LAPLNDTRVVHAAKA
YCDLPEPRKPLEKAV
CALSHHERLKCDEWS
VKVPMMKRLGMFNIQ
YQLEGNKRITCRNGQ
HKSEVAHRFKDLGEE
LPKLDELRDEGKASS
SLQKFGERAFKAWAV
EKTPVSDRVTKCCTE
CTLSEKERQIKKQTA
PLPKEPPRKDPPPKD
XXXXXXERIDGGITG
RGGEGNARTHGTPDL
VLGNDGHREVVKEVV
RGADYSLRAVRMKIR
QQARLHERGAAEMVL
LPKIETMREKVLTSS
CTKPESERMPCTEDY
YLSLILNRLCVLHEK
TESLVNRRPCFSALT
VGVNGFGRIGRLVTR
AITIFQERDPANIKW

(2) List of the 847 peptide samples in the negative subset $S_{\xi=7}^{-}(\boldsymbol{R})$

VTPHTQCRLKLLKLE
VLSFSETRAHQVVQQ
ELSKANSRFATTFYQ
EEAAGSDRASDLDSV
GSNGVDHRSYRIIRG
HSGASQHRIARPSRQ
DSLLGKNRKALAKGL
IVDEVHERDVNTDFL
SDVEAHFRAAHENTK
ELFVKGVREVYDYYC
HASCYGIRPELVNEG
ASPATVSRMGMVYIS
RVQRLGERVVDSGRS
RSLSRFERSARFDIF
PNTYDIHRLEKILKA
EKPGGKERGSSASHP
VTEASPWRATNPFLN

KKKKKKGRMKKEDNI
GEPKDVIRKDVRAIL
PHQLLLQRLQEEKDR
LSLSHFCRSPFLCFG
TTEAPVKRASLLGDM
QEECGYLRRHHQEEV
VSNTLESRLDLIAQQ
KKKMTKARLERFKLR
VYTTKSPRERAILGL
PASPISQRLPPKVES
GAQMEAARPLVQENP
TKDNEHKRSLTKTPA
LHVTDAGRKHIAIAW
INYVLESRLIGTEKF
SSASLIWRSEAEASV
KFTTEEKRFVESRDV
EGEQYLFRIRAQNEK
PPIYTAAREQTPFRH
IRGNEAGRFRLDITL
EWDKDVARKLGSTTV
TDDSGTYRAVCTNYK
ESKGPMTRRLLLHEV
LDLKETLRLRILSED
DELLRHHRDEQKPAT
ALNMLTWRAEQEGGM
PSSGLKKRPISRLQT
CVIQSYWRMRQDRVR
DVWIDVQRRWVYLEG
WLLAHCGRPQTECRH
AACVMPSRLKALGTL
HKQTKKRRILIPSDI
SIQDWAKRMKALVEQ
QENCILLRLDNELGG
SQFNLSGRNPQKQAR
KSMSIVLRENHQKPE
LPEEIALRASDGPQL
EIEKLEFRVRELEQA
EKQVELYRNGQRLLE
AKKPAATRKSSKNPK
EGQEYSFRVRAVNKA
SISGCKMRLSYLSSR
RFRTLMMRRRFLSLK
TYDSYYSRSRSRSRS
EKETLLQRLTEVEQE
HWGQQQIRPIKSVIP
AVPPAKGRTVLEEKV
DLSTIQERMEELKGQ
SECGFCLRQFPRSLL
TPSEAHSRIFESVTL
RIQVFLARYSYNPFE
KREKEVARKLEFDGL
PLKTDIPRHPMPFAA
CYIFPGGRGDSALFA
RDMEKVERQAVPQAN
LDLDSELRKGIVVRA
SNQSVLHRWERKQNL

ECSKTSFRVANLEEG
MSTIEPHRQVAWKRA
YKDKRELRSGKKYKI
LEAQVLPRFQPRVLQ
SPQPPLSRAEAIKQV
ELEAHLCRMMKHSMD
EKKLEDARARCEGQE
TYSSAEERMQSEQIR
GLELSIDRFRLRKKT
SEIRITHRSHHFIPR
CSKDELERALHYLKN
HFIPTIGRLRKRAGK
KQLPQVPRPLQLFSA
CEPLSNKRNSNSVTN
HVHEASSRSHAIFTI
KNPASSKRRLKTSLG
GPKLEALRAEVSKLE
WICTKCVRCKSCGST
NCVSGQERGRLGVLA
KLNETMERQRTEIAR
NYSESLYRQAMEEAK
LAMLAKERLQEVRDK
EEKQEEERSKVDDLR
SEPGANMRHIRKPVI
HVEAVYSRRCVSFIL
QNQVGIGRPAELKEA
IHYLMLPRVREELID
YGLITHQRIHTGEKP
TGLETPDRKGKGTTE
KVVPPSMRRIRAQKG
RRELLELRRKGREEK
PLGLILNRFSADTNI
GNLSILSRQESSFFT
IQGKDVLRLPPSSIT
LNTSIDLRTNVLNDA
EKTINQQRIHARIGQ
KAKEQETRISEEITK
VLIPFKGRPPPTVTW
MALEETLRQYQAAKS
PGFKGEGREVDVNLP
RDLLLQTRLVNHVSS
DNKKKKQRTSSKKKA
WDDIITNRCFFLSKI
RKTRSARRRTTQIIN
YIRISPTRAYNRPTL
LEEELPKRVRSRLSL
QLKSFPARLRQYASY
STGLLPGRGPGTSAP
VKAGDSSRLECKIAG
ATEISTERDLGQCDR
CREVGLQRRSVQLFC
EFKQKLCRAEQAFQA
KVNGCLGRNCKLPIT
HQAAAATREASSESD
GAQMLAMRGEQLGVV
TEGMDDFRYACQSPE

```
IKKANMKRENKAYSF
FSCQRRTRFISLGEA
NSIEASLRCSSVVPR
ALTDEESRKNWEEFG
EEKPEEERSAEESKP
SQELDVKRREAIYND
EKKSKKIRDKTSKKK
DVFLDSGRKTRSARR
KIDHELHRLQALLKH
LCPEGVHRFQWIRNL
QDQKVNIRVAVLPCS
GATTSTIRSSCLVGC
APMTPAARPEDYEAA
KNLVQKERELEEQEK
QFGGIDQRKIFTFAE
QERDRKFRLEEQKVR
VVLPLDERAFEKTLT
ALAVEGKRIGKVLDH
EFLELKYRLLSLLVL
AKWFKDGRELSADSK
RVRPLSKRETKEGGR
IDVNVIARPSAPKEL
EAYITATRQYVAQAN
AQSEAAGRTEPTGPK
CYFHSGPRGTHDLWD
VNLVSVGRVLWQWTK
ALHTEITRNKDSTLE
LQVNDHVRKVTDQIS
VTLQSAYRGMKVRKR
IVTEEHLRRAIGNIE
CPGMMLWRYPEPRVL
EDVHIKGRIIDTVAQ
ELVPPSFRGGKQIDS
KLQTAIHRVTLAQTA
VHSPDIYRVVVEGER
DIANVMQRLQDEQEI
KFVGDATRLTDKLLE
GLTNPSLRSPEESTE
KGNEIVLRQDIMPTT
RSRSGKGRGLPKKGG
SGAGCSSRDPGPPEP
LEKIKAKRSDSGKYC
KFSSKYLRTEHGSEM
SKSIEKKRGRPPTFP
RKVFSPIRSEPRSPS
WYNVENWRTAVTSPD
HRLIAKSRELYEKTE
GLLFYSLRTLTTMAP
ESDSFCLRNIMEAPL
VKEVAAWRERYEDSQ
KLGKVNDRWQHLLDL
LGYWLEKREVNSTHW
ERDYLLERRDLAVDF
IIDEGLKRFIIEITF
AEAAHNQRAGAPGIQ
TSVPTTPRATPILQP
```

PLDQTLPRQGPGQSL
IKALADEREAVQKKT
LINGGSERKSCFWKL
VADLGNIRIRFIRKN
AKISDLHRESVDHLT
EDVTFNKRKMKELED
KKQQQQCRPSISISS
PKVLQDYRKLKNTAE
IKKLEPSRIVKQDEF
GKPKVLARTKGSMLV
AWIPVPARREAEKVP
SLDWLISRLERLEAS
RLKAQEHRAEMEEKG
DEAKTSARDAALAYH
PTTRSSARRSQAGVS
LYYYNFTRTVISSGG
DTQTIWARTGYLGFL
NELAYFKRENQELMI
ETQIDETRMGYKPVA
ALFPKGSRSVVPRDT
EVFTYSVRGLYENHK
PRARPRERDPGRRPH
NAALDNLRKVVPCYS
ESTSQIERPLSQEPA
YDGPKNVRSDISDQE
PFTPVPPRGEPVTVY
MKKFEESRAELEKVL
GTIDGEVRLTGELDR
SETKDPLRVQPHLKK
LMVHVDKRITLAAFK
FFAKLNCRLYRKANK
QKVAEQERTAQQLRA
SEILSNIRSCRLQCF
PLPTVNERDTENHTS
EQEQLYLRSGVVSSA
AVVDVSDRAVPPSFT
LEYCEVQRDPYSNYG
EDISKLNRAAKVVQE
QGETNKNRTKGGWQQ
EFITSRKRGPLWNHE
DMSMLEERIKRSAKR
AKQFRAVRTTEGEEE
VKERESERNQCNFKP
DKINKTYRSQLSSEE
EGVKYQFRAMAINAA
GPPASNLRKQKSLTN
DVLLGASRTLSLALD
EEVMHTPRTPSCSAD
PGLPVEVRGSNGAFY
PGSTVLYRIFTVNHK
AQGALPERDASRGGL
PVWAAKQRVLCALNH
LEKRQMDRAEHKGEL
ITESDAQRTMYPGSC
ETPSYAKRRRLAGPS
EDEVQFLRTDDEVVL

AVVLTSYRSTAERKL
RSEGEKVRVGDDLIL
LHNSLLQRKSKLQSL
GWWVRNTRILRKCIE
LLANAPERVVERRRR
RNGPAQTRVSSSSSH
SVQLTEKRMDKVGKY
EDTWKLGRNWGQSVE
SWIWVQLRTGLARDG
HCFHMCDRMVIYFFI
STPSLTSRKIHGLSH
QSVAKMFRAVEEGLT
LSVKDAVRVDSGNYI
KIYNIEDRTRVPNKF
DADLEALRRKLANAP
NFLASKHRQPPEYNP
CEPNCYSRVINIDGQ
PPEPVDKRAKAPKAR
NKPIPALRVVEEKKK
SKLLKNWRDLDARIT
IRNGVTKRGETLSWL
YTMESTDRNQTFSKE
CNGQNTLRNIVHLSK
PLPIDLYRYDDVSGK
LRGLLRSRGAWEQQI
EDELQLPRLPELFET
FKHIKDTRYMSSYFK
QRAAAMARTKVEIAN
QNSSVLDRLPQPAES
TSKSSDNRETPRNHS
KDTLMISRTPEVTCV
QLKNLKIRTNRIQRF
HQWIQETRTYLLDGS
QALLSKMRAIDLQIK
KKTWAKVRTAKFFIL
MSFKDAERGDDTSCE
VEEIMYLRQIYKQQL
TTAATVQRPGPGQTG
VGLWSMCRTNRGTVG
XXXXMSARTPLPTVN
PHPSPEARAPLASAS
DTPPSALRGSQSVSL
HCSLDALREFFSTIV
STSYMHQRSPGGPTK
LLIQKYYRAYSIGRE
ESESDNHRSSSDFFE
ARFDIFSRYESMKAA
TNYILEKRETMSKRW
KAPSESARGPPGPTD
RKKKSKDRPSKLEKK
KVELSQLRVAKVTGG
VTSQFTTRDDILRDW
GHDEDNTRISSAGCL
SGSFVSSRARREKKS
HLFYKIHRNRYPVCT
GLCIFFVRCRNDVAI

SLEKNLDRMKLDWVN
MTCLSIDRFIAVVHP
AETVPEERARGDPDS
NEHCPSKRDPSKLYK
LQQQEQLRALQGRQA
RLPTLPKRVPRAGEA
PDSMFGDRGEIIDPF
EGQEYYEREEGYDEG
EVNAAWQRLKGLALQ
TFQTLEPRASDAKNQ
FLALILQRTFLQASY
EHQMYEFRVKAVNEI
QVWETWPRAQELCPE
IVNLLKMRQGEEVFQ
QIHMIAQRRDAMAHR
SLSEEIQRRLEATGM
VYTTAEQRPNAYIPE
RRVRETQRHISTVKY
AFYEKMGRMLGSAFP
PGAKVTIRASLGPLQ
TKLLKALRVRKKTGG
ITLMDVTRNSVSLSW
LVAVLLSRTWHVRRQ
YLTVTLQRPTKELHG
AAMFQSERKNPAPQC
QDHFQVARQNHFPQR
SPRRRKSRSPSPRRR
SSEEKTQRMSLMRHH
KHKGTEKRESPSPAP
FLDVFLPRVTELALT
HFFFPKSRIVRSLMP
AVQKTTFRVTRLHEG
GYTEVQFRPYRTDDF
DPVEHGQRHTATKRE
HEYEPYSRLKSDIKD
TKFQKLHRDMEEATL
PSAMPQSRSTESAHA
VWEPLIERVEGKRQW
PNVDYQFRVCARGDG
SWIADCARHHCSSTP
VLESTTGRENSDQHY
ELSLNKQRKLNKSES
AWYMHGERAHYSHTM
SATLKNFRYHISLKT
YIKSELDRELQDRYV
LGLDEHNRVKVYRFX
LDTYNVKREAAEQWL
KAWKKHCRGEGKISK
VFKHGYERYNAMRAD
QLSEEQGRNLELQVL
IKVLEKKRAMFECEV
GTPVFADRLPEKMKT
DPSAPQPRAETSGKD
KSLLDWLRQQADYSL
KAAQQHWRHELASAQ
LNKHTALRGEMSGRL

TCMKFYARVCRSGSG
PKKVEMQRSLPGSLL
AKCQVTPRRNVLQKR
FVELGVTRVGHRMNI
IMEEAVTRKFVHEDS
EPIQFTKRIQNIVVS
LVIQKYWRRVLAQRK
LEAKQKARKKEERQS
LLEEALLRLLPAELP
YQTMLKGRCPSVSAP
LTKEKVSRLDRIVAE
KELADLARLHPTSCA
FNELEALRAESVATK
YLVLQQLRCNGVLEG
QMVTPPPRLVVGTYD
VPYFEQCRRDACRCG
EQEIVQKRTFTKWIN
KNQGLFHRVLREEIS
AACCGSCRNVSCLFT
SLAKFAGRKLKDCGE
CSSLSPQRLCSKHMP
DAAAAGDRRPAPDTW
LDINVVIRQVYHLME
DDQEIQKRLAEQQDL
RYIRSALRQEKALSA
LEKAAAQRKAKLDEN
SGEAPGIRKEMKDVT
REEAEKQRVASENLR
VLQKCSHRLQELEKN
MEFAVNGRPLEPSQA
AREHSLLRSDRPAYV
SNYQSTTRVKPFICT
AVASAASRPDSTHMV
LKYQSLLRMFLSYCI
NSQVTFPRIQVTSLS
DALTDPSRKTQKCLQ
EVLFLSSRSKVFERA
DTIAHLERTRENMEQ
AEQKSLLRSVASRGE
TSAEQDNRFSNKQKK
PANAGNMRYSHSTGK
MVAVLVSRTVGPTQR
NHHMEVLRELCEELP
LVNKDSTRRDSGAYT
AMVFMDVRMPKFKIY
LVSHVTLRLLKPECV
EKFGLEKRQGALELI
EGCEYEYRVYAENAA
YACGVEDRKSRARLF
QPDVTGIRLLSLGAG
EELHKEIRLIGNESK
SFLKLTTRAQLGAKS
LRVIEMEREGKGRIS
QIYGQKGRTDEILLN
IYNTKSSRERAALSL
VELLFNTRYAKAIGI

YGGSYGGRGRGRRTG
EDTQYQFRVYAVNKI
NSQPVAVRGGGGKQV
TPSPPPRRRSPSPRR
LEFDGGSRGKGEHFP
YLMGGTYRCTYTRKR
MAVGMGERPEVLHLT
EGGKGEKRKRETDDE
ITMWLSKRLPQFLQV
NIVWKYQRYHFIMAY
LTASPSSRPVASPGA
XXXXMAVRQALGRGL
QAMPYNCRNNLAFPA
LFGKMIARAGRAGNL
LIVCFIKRSRGGKYP
AHEKRKERLQMLQTN
ASLQPKGRRDPPLLP
LQRQHYLRQRHSAVI
ELKALAQRGIGYHHS
QDEWEIPRQSLRLVR
AALESLNRAVLAAMD
QFLEEEGRTLEDVAR
KDDIKSTRKLKESLI
YVRVKTERKNFLAVQ
VNSTHWSRVNKSLLN
VDTKYQIRIQIQEKL
AGLSTDNRNVCLWDT
PRRASPLRTSRSRPH
MPSKFSCRQLREAGQ
GENQLLERRRLYHCA
TTIDDGSRCFFTKSK
AAKVFGLRSRKLKLF
ERLKEPTRQALQQRL
VIFYDFLRGLEASWI
DEEASYLRFRNSIWK
PKNDGGSRIKGYIVE
LQVITTLRTAAKEME
SGEGDLSRHLGLLGL
NDQLEEQRQEQALQR
QEGSLSARWPVAGQI
RQILRLLRLRCTEND
DPVDSTLRDFCGRCI
PVYKTSARRGTLSTT
KTIGSPKRIQSPLNN
TNSVTAVRIQPLEDI
LIPVIGPRKNIKKQQ
ILVNTNLRALINKHT
WGTATVGRPRGPPKA
WLEAQEERLKTLQKP
GETPEACRQARSYLE
LRQCLAPRLPHSVTC
ADEIYELRVTGRTQD
NAEKYYGRKSPVGRD
LELESQKRLYEKNLT
KSEAKLSRKQVDSEH
SSLSSAGRPGPSEGG

KVFAKVLREEDSKDA
FIAKAFKRIKDSEFE
QHHRLESRYSSSSGG
LAVGKFTRTSGETTH
VKMCVGSRRMVDVMD
QRPSGEDRWQEKDQD
GSKKVTIRSPVQIRN
TSENTLKRVSSLAGF
QPGTFLLRFSESSRE
TDFWKTLRYLSLLYP
AEPGEGTRPATVGDS
PLSPPQHRYSEGERT
GCFLFISRSKPLKTL
VFLHQKPRLPKLFKQ
HSYSVLERFVEECFQ
TNDLTAKRLLHVKGR
KSLKAKFRKSDTNEW
LNELEILRNSAVSQE
EWKFKLFRVRSFEKT
KASAAMVRLRLYDIL
EYNAVKEREFHNQYR
MVTMGFARDEINDAL
INIDQFVRKYRAALG
DISALTLRISELESQ
SCFKDGVRQIPIESI
PWAKYVVREGDNVNY
LSFLDAYRNYAQHKR
SESEDEQRPRTRSAK
IGSDSQGRATAANNK
EEPGSSSRESVSKAG
GPIFTTARQLVHALA
GQSPFQQRKKKIKRE
DALEEQQRCISELKA
YKTLHDTRTHFLKTK
HSSEHAGRNGRNAAD
AAIGATPRAKGKAKH
SKRKVLGRRDSDDDH
INKEAQKRWTRREQA
SFMSLQLRELVIKSL
EHKTMASRAKVMADW
VKQKELSRIEEALQK
QYSFNSQRFAELLST
SLHYALARKGGAGGT
AHTKPLNRRSVLEKN
SLDAANARLMSALTQ
SGLDIFERINTSAFE
AVDSPAGRQQLLQRG
LVFHSITRSHAENLE
AEYTVVARNKYGEDS
NAQILSQRIEKAKCL
EDKGEPKRRGYLQVN
YGQIESVRFRSLIPA
ENQQCLLREETWLAS
HQVESFIRKKLESLL
MKFKQKPREEQLEAD
ASAADTSRLSLWGIQ

LPKVTLSRDGVPLKA
ETLGVLARAVGEPMR
EPRRLSSRRSVLTSP
SVVWLKDRGEVIRES
ILTLAWDRVDIAKNH
VYNENSNRQKLEHVK
FKIAGDMRATCPAFN
PSHGLADRVINCREV
AQQEELERRKRLEQQ
MEDKIYDRQVTKQSL
EEQPTSQRDRLSQVL
TNYVIELRDKTSIRW
QKKSAFTRTVSPGSV
EVLNEVNRRSLKKLA
QTAGVIDRWELLQAQ
RSKVDDLRGTPMSVG
REGITTKREQVQITQ
EASTSALRESCQAEH
PANGFLVRYLRRKLV
NNTDEMARLIRSVMQ
ASERKLNRRGSMSYL
YSASTPVRKPRPGQQ
EMDISADRVKVEGEL
WDLSKKLRIPAISRL
SLALKNIRCRRGIHK
ALQAALARVVPSYMQ
GMYSEPLRQFRDSSV
MDGAAWDRRNGKLME
GSYLQFLRILFRMLI
LKHFIGQREEFETAR
LVSKDLIRKAGVGSV
GFKSHLIRLIGNLCY
AHEDISQRVAAENQD
EIEEDKARRILELSG
SMLQHLLRRLVFDVP
ISPPSPDRPPHSQTS
RSRGKEQRKLARQRS
EEAEAAARALARFAQ
TLTIQKARVTEKAVT
EFTQDLFRFLQLLCE
GAHKATLRIGQDGIS
ANLKMELRDALCAII
ENSLETLRFSISNLS
MFKWLVARINRALDA
HKFLTIPRLEELYNT
RYNERKGRSELIVVE
SVDVKSVRMGSIQRK
KGGSKLLRMKLSDFN
TGTPSDPRRRLRSYE
PEVFKYNRFLNPDGS
ESVKASERPLPPGKI
AYPICEWRYDACASP
PETIDTARLHYRNSW
RSALFALRYNILTLM
EPLKELFRQQEAVRG
GDEEQKLRFAFSIYD

YLMLDNKRKEVVHKI
RLDPVSGRLSTISSL
KPQDYIPRAPTFLMN
NSHHECIRKLLQSKC
TSSLSPQRSKLRIMS
QGLKIHQRFHFASAL
SSKPIKIRGPSHCAG
HANNVTIRESMQNDV
PGHFVGPRGPHPSQF
GLDPTQFRVHHYHKD
PSTNTESRKDVITIS
IDAIGTKRYDSNSGG
AESNSNMRCTCRIIE
VEVAATERTLLGFFL
YDYVHSVRGKVAPTT
SRALSHDRQLISQDA
GLPNNAERVLLTTQG
FDFPDLNRFLTRTLQ
GGKPVQNRELQGYES
LSLQVVNREIREENK
KAFVTLERFDQLYGS
EAVKYSERSLTKCIG
FGKKAANRSWQNVYC
TLRMRSHRVPCGQNT
QVSETLKRFAGKVTT
EAVNVCGRATAVVEV
EKALRSMRKAQVSPQ
EKVLSPLRSPPLIGS
LDLVSHPRATQTHVE
LISQINKRYKTKDDI
SSEKVAERKETEFFS
TAKPINPRPSLPPNS
EAKAELERKLSEKDE
ITKVGEERKGPLPKS
LVEETEERLVERISF
QIYSTGLRKGNLHNV
KHFPNIDREKAMSRP
SMSKGGKRKDSVWGS
WADPTGTRKLTWTYA
AQIADGFRIRVDLAS
LRDWNLIRDAATLIL
AEAADNQREEAADNQ
VYQIKVHRKYTGEDF
ARLETLVRKAGEQQE
DASPGDKRLAAYLML
LSSYATARLMMNNEY
KLPVDFERSFQEEKP
GPLPYPERQKRARSM
DSAKELQRSVEFDRE
TGTLVSKREVELEKN
GLFNKNPRHSSSENN
FNVPDEERCSFATVN
ERERERREAERAAK
QQLNKDTRSLGEEPV
FSPLISARFGMGSMP
KTKTDYTRIKSLSIN

```
SSPSSGQRSASPSVP
QHTTMATRSPALPPE
VQQVTFSRGTDGQPL
KKKGGLFRRGSLLGS
GNFTDAERRKCEEEI
FLNPDPLRADGISDL
WFETQEKRLKQQHRI
QGYIIEKRRHDKPDF
SLWSNTVRCYLIYTD
FDFQAMGRIFVGLCQ
TPGGTISRASRDLLK
DTDMDSLREARREDM
KKLRLKFRDVKCIPD
EPSFEATRSRNRYSV
SSGKSVARTSLQAED
VMLADSPREESNKRQ
LSCLTTSRSMTDKLA
KNYVIDKRESTRKAY
VWGGIWPRHRFSLQV
GRREAGSRAHPLLRL
YIQLEVPREATISYI
DSVSEATRDQDGVLR
NQLAEKVRLRLRYEE
ILKPMMLRRLKDDVE
VSPFPKKRGRPKRQM
ATDQEKSRKLHELTV
MVVLSLPRIQSQRYQ
XXXXXXMRGAMELEP
FREEEQLRQEREEQQ
QNLPVQDRNLMLPDG
DEFTTLARDLYAFYP
DMAAHEERVNEVNQF
TAEDIQERRQQVLDR
KMAEESERRAKEADQ
ALHQKTLRVDVCTTD
LAAERHARLNDSYRL
RLDDEIGRKVYMRDR
IKIQAMWRRYRAKKY
ESLHQNSRATLLYGK
QSKALTERLKLNSKR
PSLEGDLRGPDVKLE
TFYVTKSRDALTETA
AVSSADPRAPGESPC
PISDGGSRVIGYHLE
PHQIGGIRFLYDNLV
GNMELEKRRQALMEQ
EEQQESARAVPEKST
PDLEKVVRYLAGCGL
EMEPLVPRVKEVKSA
YNINTYARFCKCCRE
EKIDLEVRSLETCMY
LAKMFLNRINYWHLE
TILDMVERECGTIIE
KLAYYEIRYQFYKRD
VPTVTWLRKGQVLST
DMNKVKLRAIERSPG
```

LLDSPYARVHIAEDQ
GKMPLEERIQKIKEI
KLKPVELRELLNPVV
LQVVACTRGGMSRLF
YLQHLCYRNDKVKTD
DEVEVYSRANEQEPC
EDFSQLLRNFGQWLQ
LLFRGANRDVRNYNS
TKSEKLVRLHQEYQR
GPLASHMRWFIWLMA
QSVQETVRVRKVDVS
SLPCLSLRKLPPRSE
KEVEVAKRNLAQQKI
LSPGNSARLTPLRYD
TVGDSSARPARRVLL
QEEITEYRRQLQART
MPWKTCWRYLTSGGF
IVVEEKVRVPEEPRV
PSFSRQLRDVQETVG
KKFGGPGRMKQSCVL
TSLQKLFREVRIMKI
PPVETIQRTTARDPI
SQLLTSVRRMVLTLT
CKLAALVRHRVSLFG
TLFSNVQRVQGVSER
EVQYEIFRSLMYWMT
IELPDLGRFYKIRVW
AAEELLQRCISSALP
ELSHDNIRVKWFKND
VPDALNDRHPDVRKC
NNAENLFRAFLGELK
VIAMLGFRLDAVKSN
KNMLIHGRAMLLVGR
AQAGAKSRPKKREGV
DFEYKLDREFLKGCK
LDALVMDRVAFVKLL
VSIEAPKREPQPIKE
PIGYQCSRVYWSTTD
SANKENFRILVESVK
HQLTEEDRERLKEEL
EAAQLSERTAQEQRT
ECIQMKARLTQELQQ
AGPGLLPRKPPGLAA
NFPVHLSRIRRLRAE
KVDQFVTRFLLRETV
HYPFENKREFEDAFP
NNLHSGIRKLEDYLT
RNLWQSSRPLICLLD
IQSAPNVRFQWFKAG
ISSPPACRSPSPVAN
GYPGGKARIIHKESD
DQRLHTTRSVSMQDE
APPVTSARFPSWVTF
DKCNDIIRNKDDTAA
IEVPISGRPKPTITW
TEINAKLRNIIVLDS

VRQEQHKRLEDLRRR
ITDASALRLVDKAQV
SILLDFTRGSRGRKN
FHHCEQCRQYMDFTS
EYEMPLYRKYWPNHI
IRPRPPSRSRAAQSP
RAARTLYRIELLRKV
VGGSGLPRGRRPPRW
MHHDPFLRSMAYWIL
LCTVATLRETYGEMA
CAKQEPERNECFLQH
DDNPNLPRLVRPEVD
KYLYEIARRHPYFYA
ELLFFAKRYKAAFTE
KASSAKQRLKCASLQ
FKAWAVARLSQRFPK
LLECADDRADLAKYI
MFLYEYARRHPDYSV
YSVVLLLRLAKTYET
FQNALLVRYTKKVPQ
PTLVEVSRNLGKVGS
CKHPEAKRMPCAEDY
CTESLVNRRPCFSAL
PIPQCPPRLDVQTIQ
WNPIEGERYLSFLRF
QFLEQAQRELAELLD
LNFTMNDRHTGPAWN
CDCELQHRVEAIVAF
FCQEMKFRNKMVEYL
QEESQAARIIRNTTA
LVIEEQLRRAAYLNM
DDIKCLTRDLDQASM
DMKADVTRLPATIAR
YLHLLTLRGSVLRVE
MKCSSMQRFGERAFK
GKNIYVERQRPALGE
SKPEGTPRGPQVYTM
KLTGMAFRVPTPNVS
SLGKVGTRCCTKPES
GPLSLASRNRSQDSG
LKNVNNMRIFGEAAE
CQKGFVLRGSGVIHC
VVVKLLIRRPECFGP
CAKQEPERNECFLQH
TCKKFDLRVTIKPAP
KATLKNLRKQYRNVK
LVEVMMARRDDLSFC
TENAIGSRKFLSLGL
YHKDLQTRATFMEVL
ANIHKEQRKFCLAAE
CSGSLVERRPCFSAL
YDDKAIERLLDRNQD
EHMPNDSRLKCLLDP
VLVFEQVRLLMKSYS
TEILKWLREILICRN
SFEVDPTRLEPSESL

NMGQWKIRAFYEHAP
FSKKCSGRWKTMSAK
EIMEACMRDIPTCKW
VNGILVSRGVKDMEL
TLGSTVGRLQEQLGP
GLAESGARYTEMPHV
PLPPLLARVGGNIEV
TYPGLSKRRLKPEAI
STLQTIYRLSKGRSL
YPVGTVLRYICRPGY
GGITGNMRQEAIDRF
EKETDWLRNEMNKDL
ANQDFTNRINKLKNS
KYLPGLLRSGFYDLL
GVVWVPFRGADYSLR
RIPPVAVRLQMSERN
LFLHGYQRFWIETEA
LAFDVGLREHATGYL
VCALGNSRVASALCS
ENAQTGGRKRGMSRR
VSLHLPSRSSPTVFR
LHKTIVKRRMSHVSG
YELLLKVRPEQLVKH
PMKYSCQRRARLITQ
VLVNSLHRIITNSAL
DTTVKKTRRGQKAE
EKADMEDREKEDKIK
SCSRSVSREINLKDY
WLISKLDRLESSSGI
PAVNGTLRTSSYEEH
WVQAVVSRFDEQLPI
VPLVVFKREKEIARK
DLSWSMLRGCKEICC
KGDPKKPRGKMSSYT
NPDGSPARRVPVVTQ
IPITSSTRNIRLYPD
MLQQLLRRLVFDVPQ
IQQQLEQRISASGAE
TSFSNYYRTWNVVVH
EAGGAGTRARSLWRV
GQPKDTMRLDETMLV
DDIAILARYMLMLSF
FQNAILVRYTQKAPQ
AHTVVTSRVVNRADT
GDLYGAERIAELLGL
QTLWLWARPRAGGAW