**Supporting Information S4.** The benchmark dataset 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 pairwise sequence identity with any other in a same subset. See the main text for further explanation.

**(1)** List of the 121 peptide samples in the positive subset

HLELRNR**T**PAAVREL

PRKRPPV**T**QAAGALP

LLAKQHK**T**QRSPVRI

NEPEAEK**T**HLFAKQE

RKSLVMH**T**PPVLKKI

GSKRQLQ**T**PKEKAQA

KDGSVTG**T**KRLRCMP

DTEEADR**T**SSKKTKT

MPRRLTQ**T**GKRKHGA

GTARRTG**T**PSDPRRR

NSKHLNG**T**ITAKYTY

GKEMLEQ**T**LQKVTEL

NELLKQK**T**IELTRAC

LDTMSKK**T**TALDQLS

ASKEQGP**T**PVEKKKK

DLIQHRR**T**QDHKIAK

IEEELKT**T**KRKMNLK

VQKANPG**T**LAAEIRA

WIWVQLR**T**GLARDGR

VSAAQGT**T**GTAPGAR

FQASLPL**T**RIDEEEA

LNDKHLK**T**LSKTVFS

RPLKSQI**T**VTRHDKR

DSKNSKK**T**NTEFLHT

ERVEQFL**T**IARRRGR

PDDRTAR**T**LTLIAKV

RDLRFVR**T**AEILAGK

KDNIPPL**T**MKRIRER

KGITFIF**T**DSEIKDE

MHFRSLR**T**KLLLMSR

RQPVTKN**T**FRQYRVL

SEVLVQQ**T**LQRIRQS

FVQRRGH**T**VLKIHLP

KAKKPAA**T**RKSSKNP

MAVLEAL**T**GVLRSCG

HSGKLYK**T**KSNKELH

VQSIKLI**T**EGASKRI

YSAEWKS**T**VSRLGRW

GAFKAVM**T**SIKQLSS

RDKLKFP**T**RPKVRVP

VKRGDIF**T**LEEELPK

QLIPIRH**T**KTREKVD

NEIVNIR**T**SLLNLVD

KQELEVL**T**ANIQDLK

PKDEDER**T**VYVELLP

EKIRLAK**T**QQASKHI

EDLGRKR**T**TMLKIAT

EVAEKQA**T**DVKPKAA

EKFKALY**T**LPRSVDD

ARIKAKL**T**QIRRYGE

SKGELLS**T**LLTKEKA

LTKLKAK**T**DNVVQAK

QWIRATA**T**ALERCSE

PLGAEKI**T**GELEEMR

ISNVDVE**T**QSGKTVI

LPDMSVK**T**PKISMPD

RSIRKSS**T**SSDNFKA

TPESKKA**T**SCFPRPM

SVFLRVR**T**NVGVRVL

SVSVEKS**T**SSNRKNQ

TSPLQSP**T**KAKPKVE

SQGKLLR**T**ILFGVKR

STAVGLV**T**KDPGKKP

LGKKLGS**T**VGLVAKD

NYDKILA**T**KKNLDHV

RAALDKA**T**VLLSMSK

LLRAGYL**T**LYGIEAL

ETKETTT**T**NKQTSTD

TPKKPAP**T**TPKEPAP

RTARAIT**T**RSGQTLS

KNGKVYF**T**SDAGIAG

IALTTYE**T**LRLCLDE

SSVGWDA**T**EDLRISR

TEELIRL**T**QRLRFHK

NHFLQVE**T**RRAGERL

QTSLIVA**T**LKKMLPI

DVKSTAF**T**FRSSKEK

NKEKEMV**T**SLFCKLG

LRIHAHF**T**GLRYLLY

TSSPPRK**T**RRLSPSA

RLLRIAA**T**PSCHLLH

PPKRAPS**T**TLTLRSK

LRKGIPR**T**KSVGEDE

EPARWDT**T**LPTPSPK

WKKLLEA**T**ELKGIKL

VEKPPKF**T**EKGNLEV

LRLLNKH**T**ALRGEMS

ATRVRNI**T**KKSSHLP

SYKLKSA**T**GSVVSTR

ITKKMDK**T**IPEGRIR

VARQTLE**T**IRSVGYY

DLKTIYQ**T**GCQTSTV

ITKSSTS**T**IKDKDEL

VRVTGIP**T**PVVKFYR

IMKETRK**T**VVPKVIV

KKLSDIS**T**VVGKEVQ

SSEPKEC**T**YTIPKLL

VDVKFKD**T**VILKAGE

KILSNNN**T**SENTLKR

CLQTLLD**T**KFVHFID

KPLKKKP**T**PVLLPQS

GIISSPL**T**GKRQDSG

VLCIIQD**T**TNSKTVN

QLSKKLK**T**VLDQARQ

LEEIKER**T**ATGLTHI

KKTILMV**T**NRRVLCI

RKIKVSN**T**LESRLDL

KGFTVKR**T**LVIHQRT

LLAPLND**T**RVVHAAK

NETRESE**T**KLKELPG

DGTWSPR**T**PSCGDIC

RDSARQS**T**LDKELXX

AVSEHEA**T**KCQSFRD

KEGYYGY**T**GAFRCLV

SSLLEAC**T**FRRPXXX

LEGNKRI**T**CRNGQWS

QKLEPLG**T**ELHKNAK

CLLPKIE**T**MREKVLT

KVGTRCC**T**KPESERM

TPVSEKV**T**KCCTESL

EKQIKKQ**T**ALVELLK

**(2)** List of the 732 peptide samples in the negative subset

HGSIFSG**T**AAPRCEI

SRNKSGR**T**ALMLACE

LPAIFEY**T**VVGEPAP

SIAFQMP**T**IAKDGNV

YDINQII**T**TAVMTHT

GAAGLEG**T**AATAAAG

AREPPGS**T**AGLPQEP

KMMHIAA**T**LIQRRFR

CPPGLYQ**T**VPPEEHY

QNIPPSF**T**KKLTKMD

SASAPSS**T**PTGTTVV

VSPRGMV**T**RSSPGAG

DEDDNMS**T**VMRLRTK

ARNGISP**T**NPTKLQI

RELQARL**T**LVGKEGP

NLKFQPP**T**LGPEPAA

SHRVGAA**T**AGRLPAS

PAAASLS**T**ATDGLAA

PKESMLK**T**TLTAFGF

YPEPRVL**T**LVRITPV

DSLPSPS**T**IVSGDIP

RVNKVPV**T**MTRYRST

HLRKAAI**T**IQSSYRR

RLLMASP**T**RKPEPQV

QYHKMYR**T**VKAITGR

KSKKRNS**T**QLKSRVK

KAKEDDK**T**LSKEPSK

GGGRNDI**T**GRFTRHL

ENHPQWD**T**AIEGDED

DRWNSAF**T**RKDEIIT

SEPSDAI**T**CRDDVEA

PRCAEHG**T**CRDGKCE

NGDGGCG**T**GGRNCSA

LCKAFVP**T**CIEQIHV

DFEVLMK**T**AHGHLVP

IVYHRKV**T**DDISKIK

RSSRKIC**T**IGIAPWG

VQGDSQG**T**PTSSQGS

YIVQKRE**T**SRLAWTN

DRKILDD**T**EDTVVSQ

CRDELVR**T**TAQYDQA

QEKSPVN**T**KSLFKRL

IEELLSQ**T**TNPDRFS

ATQPAAP**T**PTATITS

SKKDDFF**T**SFKISCQ

EWRRANH**T**PESCPET

PSAFYQQ**T**LPNSHLT

MREEAER**T**RDELERR

GHMYMIL**T**PSDIQIQ

SPQREVL**T**VPEANSE

KLESLVL**T**HLSRCDS

FSCGADG**T**MKMRILP

KPESIKV**T**TGDTCTL

VVKYRAG**T**SVKLRAG

LKDLLLN**T**MSQEEKA

SYSSLLE**T**IEVKGAG

PHGGAMD**T**HFANMRS

KFTSITD**T**PEQVLAK

AQEERLK**T**LQKPESV

NPYSLLD**T**SEPEPPV

ICESLEK**T**KQKISHE

ECGLGHR**T**VPLLLAE

SFIDGLA**T**FQISGAR

SSAMESL**T**KHLFAIL

RKLDLSD**T**KSIRAFA

PEAKCAC**T**DSMGVPR

PGPPVNV**T**VKEISKD

VSPENLH**T**EEILVCG

VNIQLKG**T**NEYVPRF

PIAAVLA**T**FVTHAYA

LESHMIL**T**ETLFRKI

RLTYTER**T**KSTITLD

LSRKLRE**T**DSQLLKT

GDKRRVF**T**FPCLSAF

VLDEDEE**T**KEPLVQV

SSQEKVA**T**LTSQLSA

ESVSSLS**T**NDFSIPW

CFIKLCI**T**LNEGKSI

SICSARV**T**LREPPSF

KELNALE**T**SSSAMDM

SSTCEPS**T**VAAVLSR

ILKAALI**T**EENQQLS

NKYHWEH**T**GLTLREV

MRTQEDL**T**QGLLLID

DKEGREL**T**LEKPELK

HQYKLKE**T**IQKLSNV

QAEAVLK**T**LQELKKL

LDCVLDL**T**MLKIHNS

AKGDRAA**T**LEEGNPT

LKGTQEI**T**GDDRFEL

LVVFNHL**T**PPPPDEH

VSPLLSR**T**EFCTAPL

FGQRLDE**T**VAYEQKF

KRLQEEK**T**QEKIQEE

PTGPELD**T**SYKGYMK

VTPEVKK**T**SFHVTNL

NSSNFMN**T**TNFQSLR

SASKHSP**T**EDEESAK

VEGAKKV**T**VTELLAG

ENMNRSE**T**EACFFIC

EVCQKSL**T**GYLEKKR

GLSPGSL**T**LSAPSTH

GEQLGVV**T**NWPPSLE

LIGPCCA**T**ANLEAKW

AELERLR**T**QLLFSHE

GQTSMPG**T**LVLCLPQ

TLQRTGQ**T**ILPSLNE

VNSSVKR**T**QIKVTHL

MASFLYS**T**ALPNHAI

TVAKELQ**T**LHNLRKL

IARILKD**T**AIKSADN

GDGDLMM**T**SFERMLS

RIFADGK**T**WSYTYLE

DLRKVYG**T**VLSRHHH

ALGRDIL**T**KTKQARE

AMMSVDE**T**LMCSFQI

XMSLADL**T**KTNIDEH

AAGGPKL**T**KRGLAAP

SPAARIS**T**SPIRSVR

FKGRPPP**T**VTWRKDE

AALECYN**T**FIGERTV

MAEILSG**T**ESVPLTQ

LLAPVAL**T**CGSDGSL

YEGWRID**T**YLDIPLV

QQMCCVS**T**SIVSFVR

YEGGAVC**T**HARSLWR

KFHGVLK**T**LCLEVVL

RMVYCSL**T**DFQKAVY

SQRVNGL**T**SSKNSQP

CSLRPNQ**T**EEGTTPP

HESQISF**T**IEGPLTS

AEAKPKS**T**CELYSSQ

SESPVYP**T**KFDSEKE

VEKLLDC**T**VIVDSVF

TVLKSSA**T**FQSTVAG

KQQKEEK**T**KQEAYAQ

GCVCSEG**T**ILHRRHS

LERQNLE**T**FKDRMTE

IFESVHG**T**APDIAGK

LGPADQK**T**GTPTPTS

KLCYDAF**T**ENMAGEN

KKGASFQ**T**VASLHKE

KDAYKVN**T**NLDYKKQ

VGKPGIP**T**GPIKFDE

QKSDDKV**T**LEERLDK

RDFVGQR**T**VTFSSLV

LEISEDC**T**YADVKAL

KNCREAF**T**ADGDQVF

EPLPAKV**T**ERHMQIT

ALMRSEE**T**ADLLAEK

IDHKEKS**T**EINHEIP

STERLKK**T**NEILKGI

CEESLSQ**T**PPRVTGT

GPSSERA**T**PAFHPVC

GPKQASF**T**DVRDPSL

KIQQLVN**T**LKRPKRP

ICELSFE**T**EHILLQH

KKALLPP**T**VSLSATS

LHNFSIY**T**LLGKQVT

KDLVNRY**T**QNGSLDF

KGEDPLA**T**DTRVSVE

LKANQVA**T**GIRYNER

LQSHFIP**T**IGRLRKR

AFMDISA**T**DLVLRKV

AFLRHKM**T**LISPIIL

YLRNLPI**T**AHPEVFG

LWEGKPR**T**YITEERA

IRILCCL**T**FLVKVKS

GYGSNKK**T**KHMLPSG

EPALVPG**T**PKAEDRM

GYHVINN**T**FQSLLGC

XXXXMWP**T**RRLVTIK

RFRHLKK**T**SKEAVGV

KLAQLIA**T**CPPSKSS

LKDSDEI**T**EDDNIRM

GGEVIED**T**YGAGGGE

PKAAKEK**T**VKKETKV

PPEVLDV**T**KSSVSLS

HPSVVMM**T**AGRCHTL

AVPAAIS**T**SEAAPYA

VDRHREK**T**LRLLWKI

SLKSLIL**T**ALQRETE

RVDALER**T**LEQKNKE

AGGSLKV**T**ILQSSDS

TLTAYIV**T**SLLGYRK

GKPVDGL**T**TLRNGTL

TKQSFSL**T**MSIEMPY

LVSKELS**T**WKERPAR

GEAGGGS**T**AAEEASE

IACDNIG**T**PLAVFSG

RPCDEDK**T**DSETGKL

RTSAPSP**T**ALKLATV

RGQPRTR**T**RASVRGF

VENMILL**T**IQYLVRL

VMADDEF**T**QDLFRFL

NVSQDLD**T**IRSNINN

SRSKADY**T**SHLRSLV

KSRHSDG**T**EKNKLPS

SVPKVEG**T**LKGPEVD

KEEVPEP**T**PEPPKKQ

ADGLQGE**T**QLLVSRK

NKVQLMV**T**DSELSNQ

ILYGPPG**T**GKTLLAK

CIEVVFV**T**KFLYSIL

EQSIWNV**T**VLPNSKW

GLNMCAP**T**DQDLITL

SLIKQIA**T**KVHPGGT

HGREAKV**T**ETARVPA

TDGSEIK**T**DEHYTVE

SFIWELI**T**PTKDGQA

AAMRAFK**T**VTTKCSK

SDKFYKQ**T**VNLQLQP

EIVFCAV**T**SNEQVKG

PPHLKPA**T**EKLIVVN

EVKGAGM**T**EHYVTQL

CLKEIDS**T**LYKNLFV

SPNPDVW**T**SEQNPPY

EWPFLII**T**DLFLKSP

YEAQKRI**T**QVFELEI

KAVVNED**T**QGNVSQL

GFTFGTA**T**LASGGTG

PSEAIEI**T**APEGSFA

AESGIRY**T**QMPHVME

DALCLVL**T**LMNGGDL

KSLQAVI**T**NLTQGEE

LQPEPIN**T**PTHTKQQ

GTLVSHV**T**LRLLKPE

VNVRRDS**T**GMPLWLV

MFDERIF**T**GNKFTKD

KVLKWNF**T**TPRDEYI

DITENAA**T**VSWTLPK

SSIQEQI**T**KANEEFQ

GSMLVSW**T**PPLDNGG

FLPLFKA**T**INPQDHR

PRAEKED**T**AELGVHL

CLSVLSS**T**SLRLHSN

QSLLQLS**T**CDVESKR

HSADVFH**T**FRQLDRL

HQQALPG**T**HIPEEAQ

SAPYLLA**T**SCSDEKV

GLVVKAG**T**TVRFPAI

ALYSRIG**T**AEVEKPA

IALSSSE**T**TKHATNT

NRLVDYI**T**KTSCHLA

AEAQPEA**T**AKNLLHS

FFLYSKL**T**VDKSRWQ

CVSSIFW**T**QEVSQAL

ALQPSVL**T**MCNGGAG

YKDSRGL**T**PLYHSAL

EPKECTY**T**IPKLLEG

EALIKKL**T**KQDNLAL

AIQIASA**T**MPALSLN

QKEVEQQ**T**GLSVFLP

LEEKAAM**T**DAMVPRS

NVIRSIQ**T**DKREKYY

AGGHPLD**T**PHLPQEL

KVQIHHN**T**WLHFPGH

KAEPLAF**T**FSHDYKG

KLYIEAP**T**FDLQGSY

DCNKAAV**T**IQSKYRA

TILEPLF**T**ESESKIF

RHLKESQ**T**ESTNAIL

GDGFYLP**T**AGAPGSL

LSLSVEL**T**EAKLHHG

PSQFPSE**T**AFLEEFG

RVASNPY**T**WFTMEAL

PSPFDCS**T**DQEEKIE

TSESSIH**T**ITPSVVN

AELENIA**T**LCFKALE

EQVAWAL**T**KFPCLQL

KNYSLKY**T**KYTKKDT

QHPAASP**T**HPSAIRG

QNAQVCQ**T**NPEPPAT

IPVNEKD**T**LTYFIYS

XXXMAAA**T**VGRDTLP

IILGNPA**T**ELSVATH

KDNSRDE**T**FHFQCDC

YLDGNEI**T**AIQKHGI

LFGDVGD**T**GERLLYN

KKAEAVA**T**VVAAVDQ

NLELRNL**T**VELEQKI

VEDNKRF**T**RIPLVQL

EALHELR**T**AEEFKTE

QDLRVLP**T**IDLSTMP

LADDLKK**T**VTIRAGA

AVNNYHK**T**NPTGTQE

INKFAEY**T**HEKWAFD

TIRGDLE**T**SNLQDMQ

GEGECEQ**T**PEELEPV

LREAGSV**T**ALVQCVL

VARREQN**T**KRDIPDK

ELHICFE**T**TKSNEAM

LGHDEDN**T**RISSAGC

VVRDAVI**T**SNGTLPD

QMQLDDS**T**QLNSDLK

VKSVNSG**T**WIASNSE

RGMINVS**T**SSDMDHN

VRRQEDG**T**QPQVNGQ

LHPRAVS**T**SFGCRVN

NIAVILL**T**DLIIDHS

DYLRWVQ**T**LSEQVQE

CHALSGS**T**AELREDL

ESWLVPA**T**LPGPGES

HFHIRCF**T**CQVCGCG

QIQFSQC**T**DRIKDLE

NKPVSFL**T**AQQLQQL

WGWRIER**T**REGDSMA

PPPAPRP**T**SRYNSSS

WSWSEQN**T**NEQQPDV

NLGKKPG**T**ITTVGLL

LEVIYGD**T**DSIMINT

ARLMSAL**T**QLKERYS

TINIIIC**T**VDYLLRL

QKREAEE**T**AQKMAQL

PEKAKSP**T**KEEAKSP

DYESYEK**T**TRTSDVG

SGNIKNW**T**SLMAAVA

PPARLAK**T**PSSSSSQ

VDSSQKP**T**PSAREDP

MLCEEFV**T**SYKALKS

LVLEIEG**T**NVSTTYI

SQLKSDL**T**SRDDLIS

QTKVVPQ**T**VMATVPV

EDSGGGG**T**GVKKKRK

DQIIPRM**T**TQSPLNQ

KEIDNDF**T**NETKISP

AENGKEG**T**AKEGAPG

PRPLPRS**T**FARPAFN

YSLLVLA**T**DLGSPPQ

XXXMFKN**T**FQSGFLS

EERCSFA**T**VNIQLKG

TPLGPGS**T**SSLERRK

PSSLKGS**T**SLLVHPV

GYPNRDS**T**KYAEIYG

KVDQHEW**T**KCNTTPT

SNQPASV**T**DYQNVSF

KFPGPSH**T**ESYSWPP

PYLSEVL**T**FCLEVAR

CQDGSNL**T**ASQRAYA

KNLRSIS**T**SNQPRKK

KDVNERN**T**VKGFRKF

NCQDKLE**T**SLHVLNQ

RSAELGC**T**VDEVESL

HMHQNVP**T**EEQRTNK

KKQSKKG**T**EEKKKPS

MPDPVDP**T**TVTKTFK

IASNSEL**T**PSESLAT

RLQMFNV**T**PHHHQHS

RPDIKKA**T**QAAKQAS

SKKLKKA**T**GASKKSV

APKKPTS**T**KKPKTMP

ELKHHHV**T**LEASQKE

GWLLEEQ**T**RKKQKPL

EVIYDAK**T**VNAVVEF

SECGKGF**T**MKRYLIA

IRLPRKA**T**KTKKNYK

TPVPRVR**T**DYVVRPS

PVGKLTQ**T**SGKTTQT

QARPGEG**T**APLPPPG

TPLQPQA**T**TLPAQTL

KHVSRQD**T**LQQCQAW

QNEFLGC**T**DRLVITP

SLTQAAF**T**PIPVDTS

KDPRQQS**T**PQAANSP

KREKEQV**T**EPAAKVP

EKNQALN**T**DNYGHDL

PESIPLG**T**EDRISAS

QHKAVLL**T**GEQGTAK

DWVDGEP**T**EAEKDFW

IWIFAHV**T**QGQDPWI

YLCNAQI**T**MLQMLAL

KTDMLMI**T**RRGVLFV

EELGSAG**T**SARCPPQ

LEARLWE**T**EKICQLE

TDSRTSL**T**IENANRN

GPSGSGK**T**TVITILM

IREIQKV**T**KLDRDPA

DSLEEKY**T**EPVILNL

LWAALLV**T**FLAGCQA

FYYYWKK**T**PEAASSR

PYRDSKM**T**RILQDSL

VGEISVN**T**SVAFLPY

VVVGEIV**T**EEHLRRA

EFRIDSV**T**GAITVAK

RGASFGS**T**SVSGSEP

TTNKQTS**T**DGKEKTT

VEHLEDQ**T**VTEFDDA

EGSDIES**T**EPQKQCS

QAPPPQP**T**QPLQPKA

SKSSQIG**T**GDLKILT

NSQLARH**T**SPSVISD

DRSQGGE**T**LDECGLK

YYGLSGM**T**KNTFEVF

ENAEDDK**T**ENQIPQR

XXMEDSS**T**DTEKEEE

SEAINEL**T**VLVLVNT

MSLYASL**T**SEKVQSL

KAFCRMV**T**RKLETQK

QGEVEEQ**T**FKEKELD

LDDNKLL**T**LPNGERL

EGYYEAI**T**AVELKSR

SLHQEFS**T**KQKLLQN

VRDPVRD**T**PGTCHVT

KVEHVKV**T**VLDVPGP

KESDAMS**T**QDQHVLF

DSSNCSI**T**REHICVH

NLVNKDS**T**RRDSGAY

TVKVTEA**T**ITGLIQG

TIVPHGY**T**EVANIPV

KTADHFW**T**DPSADEL

ALGEKPI**T**FYRQALL

NSSNKSY**T**LNGTKTW

DVSGILA**T**QVYTWVD

NLDLDNE**T**FHIYQPQ

IKRKQQD**T**PGSPDHR

RHRINPE**T**DLPRKLS

GKDQEER**T**KAIEAFR

ELYEHWK**T**LMNRQDQ

VKGAAGR**T**DLLRLRR

LVDIQGL**T**FQVICSY

FNYLKDA**T**EVKVIIE

DVSRDGM**T**LTWYPPE

PNLSCKR**T**LSFQLLL

AVKGAPE**T**LHSMFSQ

PSGADSK**T**VKLKSPV

RILRDVH**T**ALSTPGS

PPALQAP**T**GVTPAPS

AIDGGKN**T**SGLPAEP

LDPECKC**T**DDRLWEA

SHREKPG**T**FYQQELP

IKRQVVE**T**MGLFHDM

HPPLQPL**T**GSAGQPS

MHFESFI**T**EKKQNIQ

PHIKPPP**T**TPIPQLP

ELAWHVK**T**ITITEME

GNPRVLD**T**SRSSISI

DNEGISS**T**VITFMTS

KVEQAVE**T**EPEPELR

RSQKSLW**T**IKSMSED

KKLKHRR**T**SETSISP

LSPVLPF**T**PAAMTQA

SSIYKED**T**VLHLCLR

QEIVQKR**T**FTKWINS

TPPQPSL**T**ASPSSRP

LEDEGEY**T**CMASNDY

FVIDGKK**T**TCKLKMK

SALAEGS**T**YVPYRDS

DGSRCFF**T**KSKTDYN

AAETARR**T**REFRSPP

KSLQWGI**T**SPLLRCD

PPFSPHP**T**LLSSTPK

TFDIEAV**T**LGTTSNN

AEEFNDE**T**NVRGRLF

FHAVAWG**T**PGFLTVM

VYLGRDK**T**MQLLEEL

ENLNKLM**T**NLKSTAP

PFQMDPE**T**GLITVRE

QREIERN**T**CKIKLFC

TCIHGVW**T**QLPQCVA

YLDFSVS**T**TGMLAGV

IPEPSNI**T**GNSITLT

NGVSVED**T**INLLASL

TSSAQKI**T**KSVKAPT

GSQFFIT**T**KPAPHLD

PTDEVPS**T**PLALEPS

SEADSSE**T**SVSEKEN

ERCIKEL**T**YQAEEDK

IYGKPIP**T**IQWIKGD

YDGGSPV**T**GYWLEMK

SYSENIA**T**LQFSRVE

HSKKVLN**T**GHPLVTS

VRPARRR**T**PSPDYDF

INIEHAL**T**FLRNRSI

AVSSFKK**T**SEMEASS

SGSQMGS**T**LSLAERP

DGERDPF**T**CPGFASQ

HGNNQDL**T**RNSSTPG

YYALGQA**T**GLDEEKR

NIVHLSK**T**HFPIVQN

LRRIANV**T**LIDLRGC

EYQCSSS**T**PSLNLTN

VAGQIRA**T**QQQHDFT

SSVHAQS**T**LESVISQ

ESVSSPP**T**LHMKTGL

ESCGSNS**T**NRSTAES

NEKVEEF**T**TFVKALA

VKWFKNG**T**EILKSKK

LACKVTG**T**PPIKITW

YWVMAFI**T**KTIKLVK

KKISAAQ**T**RLSELHD

ILRQLRQ**T**VEATNSM

LIRSASV**T**PKINQLC

PSVICEH**T**KQFQTIS

NKQQNFN**T**GIKDFDF

PYPSEAS**T**TFPESHL

DMAPLDH**T**SDKSLLD

QKIERVS**T**KVIRTLK

GHPAPPV**T**SARFPSW

ILNSLGL**T**ISVSPSD

GAEIEIR**T**YDLKANA

RKAAASA**T**VSVQVTD

VLGQLGI**T**KVFSNGA

VSMHKFL**T**IPRLEEL

LELYLVK**T**SDSGTYT

KKDGFNI**T**TSEKCNI

VPPSSHE**T**GFIQYLD

ASDKYRM**T**FIDSVAV

VDPSKFM**T**PYIAHSQ

GSAEGGP**T**ELTPATS

VQGNPGK**T**KKGELSI

EIEERAR**T**EIISTDN

QLRDVQE**T**VGLPVVF

VPHILRS**T**LPATALP

DLDIIQV**T**VPDPSPT

VDASFMQ**T**LWNMNPI

FPLPDEK**T**KKRIFQI

ENALQLG**T**PVLIENI

LLAILLK**T**TNRASDV

INNVRRM**T**SHPTLPY

EKNFSFL**T**KGANQHN

KQSQEEE**T**HRLEIEE

PAAIAAA**T**TTLAQEE

LEGKMKG**T**KVKTPEM

ARYSIEN**T**DSSSLLT

GWGSPQA**T**AMVLNNL

FGGLDVL**T**ISLGGIP

EEGLRAW**T**EGEKQQV

KAEKSSS**T**DQKDSKP

YSLWEHS**T**KNHLQLE

THEASKL**T**NHVELFK

SLFVRDA**T**RDHRGVY

XXXXXXM**T**TQAPTFT

DAYLVLH**T**AKTSRGF

YQVGKEF**T**SKMDLEK

KKEQCSS**T**QSKSNKT

QINKRYK**T**KDDIKST

REKVASD**T**EEADRTS

LGHLAKF**T**CEIQSAP

HVVFGSE**T**GEGENND

TAITVSD**T**VAVKETS

ITLPVAS**T**ASTSLVV

ELNRKIA**T**QEVLIRH

YQLSKEY**T**LDVYRLS

SLGGNAK**T**AIIANVH

MIKVRQY**T**VKEIREG

EFVSHPS**T**TCREQMY

DSGEKRV**T**ITEKSVA

SFLSSNS**T**FCNKQNV

DCKKLID**T**FGIRWPE

PNRTGAR**T**LADIKAK

VQQVHEE**T**ANTIQRK

KVAERKE**T**EFFSSLP

WAVDRCY**T**GSGRVAA

KKQKSIF**T**SVPDTPD

EARNDAG**T**ASCSIEL

EMKKSVH**T**QVNTNTT

QELYLFD**T**TGKHLYT

DQLNEQV**T**KLQQQLK

ARLREVQ**T**GWEDLRA

TIKAIKD**T**EAQKVST

NFKKENE**T**LKSDLNN

DVNVIEG**T**KAVLECK

SCPSEPN**T**AALQAAL

VLPRVRF**T**RHGKELK

AASRPDS**T**HMVEARQ

KSDVPPL**T**TKGKKAT

AAQLKPM**T**PFKRTRI

KINREKA**T**FKHPRTT

HCISLKD**T**NKRVESS

SPSADIG**T**TTEGDSI

PSRKSVL**T**DPAKLKK

FSVSICQ**T**LMQDQQT

AEELSYK**T**AHEDISQ

LFVIDPN**T**GVITTQG

LDLGSLN**T**SSHRTVP

WVNEKMK**T**ATDEAYK

ISVSPLA**T**SALNPTF

LSVSVVQ**T**PTTFQET

CSELSGN**T**KSRESIQ

VIFYRDK**T**EELMELT

SCGKKEY**T**IVVKVLD

IDQFLIA**T**FMQLRLI

QQQAFPV**T**EIDCAQD

VKGGVDV**T**LPRVEGK

NSSPDED**T**EVKTLKL

MATFECE**T**SEPFVKV

FYLNKMK**T**FEEPPFE

KIPSRFS**T**PEFTILN

LYHSEAF**T**VNFGDTE

AGCPVGQ**T**RTNTFEQ

VYWDYAM**T**IRLEEIV

PLKRLLN**T**WTNRYPD

SCQRDPN**T**GRLNGES

LILEGTF**T**GTPPISV

LQPTAGG**T**SKNEGTI

EAISLSE**T**VKKVEYD

RRGRKPK**T**VLNGILS

RNPEGSR**T**LSPSRGK

VHERDVN**T**DFLLILL

KQRKDKD**T**LKAEEKK

ELPLIFI**T**PLSDVKV

PKPAVAH**T**AEREQVI

SSGNCLH**T**DGNETNV

PPDRFKF**T**SVTDSLE

GTTVVKV**T**PDLKPTE

GIGPPVE**T**IQRTTAR

RMYLSCL**T**TSRSMTD

RARHRCQ**T**PHLLAAX

IEYLMGG**T**YRCTYTR

QPRPPPS**T**QAPSVNG

MIKLKRG**T**KGLVRLF

AGKASCT**T**HLYIKAP

SMALYNR**T**RRISQTS

DIKWVDD**T**HALGVFS

VLSNVEV**T**LNVLADS

RVFKIKS**T**VVVKDLV

RELEIAR**T**NQEHAAL

EDISPVQ**T**DALDSVR

LCLEPPN**T**HRVSSFL

XXXXMVL**T**LGESWPV

SLGVILY**T**LVSGSLP

VVVISGD**T**GCGKTTR

RVCSKRG**T**TESKEVL

DSEDESE**T**IPTDKTV

YGADHPY**T**ARELEAL

TSLHCAL**T**EGLEILS

LGGNSYV**T**WNYAWSK

PRVRKPK**T**TPTPRKM

QYHRGEM**T**ALREELK

NMVQRPS**T**VDFIDWV

NIEECSL**T**QQNKSTV

ILEILES**T**IEDYAQY

SEDDNFG**T**AQSNKAI

HQNVVAE**T**LSALENK

ESDSGDY**T**CIATNMA

WKYYCLP**T**GWANFGV

EYDPSKA**T**FVKVVPT

IYRGEEL**T**YDYKFPI

PGEKRKM**T**DSHVLEE

ARAYPEL**T**LAIFSEI

KCIDPGK**T**LPEALDY

KQLLEYQ**T**MLKGRCP

RLHQLHI**T**GKLPAGS

GDLPEAG**T**TRTFSVV

LNPAVLS**T**ASLGSSQ

MTAFADR**T**VVTIAHR

DDQELQI**T**DRIKIQK

ILGAKEG**T**QEFVHTE

DGEWVYN**T**FCIYKRC

AGAPPPC**T**AVTSDHL

ELPRHPP**T**FDNLYEY

QNYHLEC**T**ETQAWMR

PAIFGVL**T**KELLQLF

RGISKHS**T**SGTDEGE

LLHQGLS**T**SHFAWAF

KVDMKDV**T**KINREKA

AVQLNIF**T**AVLSALK

DTTKVSF**T**YDETAGM

KAVLQRK**T**ACQDTDE

TVQSKIL**T**TGENYKT

RRNCCYK**T**NSHGETV

PDCNFRV**T**DVIEGTE

RRGRAGA**T**SGELPVV

GKQPFLG**T**LQDNLIE

KDAASFL**T**RMKEKHG

CGNFDLK**T**INEMRTP

KPDDKAK**T**LSYEPLS

SYWSIFM**T**LLHFVAS

GDGPRSE**T**KEFTTPE

SRESKMQ**T**RILKQDL

CYIQHVD**T**GLWLTYQ

YIIEKRE**T**SRLAWAL

VSEPELA**T**LSGDMAQ

AYDPEDE**T**ILEEAKV

SSKRRPR**T**RAQKVEV

SPDSPSS**T**PTPSPVS

LNESTFD**T**QITKKMG

EVRKSAL**T**LVMGPLD

VYYSRSK**T**THIIATN

YVIVPLK**T**GLQEVEV

VGIDLGS**T**GDTTEAL

ALAVVDV**T**KRHVDLK

ERKQERE**T**FSSAEGT

APKAPEV**T**TVTKDSM

HYLHLRA**T**VVSIQRR

TKDKKII**T**NHQKKNL

SDLDSNQ**T**HSSGTVT

PFKAAPH**T**IHPPCVV

EGNDIRK**T**NVAQIRM

SPFSSPE**T**SRFGHHI

IQRAAES**T**EEVSSLR

ANGEIAY**T**IAEGNTG

YRKHYED**T**KANVHIP

GMMSDYG**T**ARRTGTP

MAHMFVS**T**NLGESFM

QKICGKK**T**AFDPSVD

GQDYEHV**T**MLRDKFR

RMEIKDQ**T**KCAASVG

EISSASR**T**CTFLPPF

RDGQVIS**T**STLPGVQ

KRFDFDY**T**TMHRIKL

DLKSLPD**T**ELMKDTA

EQNMIGM**T**PTVIAVH

PKDPVIE**T**IRNYFPE

QSLPASQ**T**HMRVTAS

AVALVFL**T**GCQAWEF

DTDENVG**T**PCPVSAF

HSNMQFS**T**FDRDADQ

GFSWIEV**T**ILKPHLQ

KLKPDPN**T**LCDEFKA

HHLGKEF**T**PSAQAAF

YILIGSS**T**SYCEIQG

LGANGRH**T**SSWLVTP

GRIGRLV**T**RAAFNSG

DSRFGSL**T**SNFKEFG

PENLPPL**T**ADFAEDK

LKSNSMV**T**LGCLVKG

FPESWLW**T**IEELKEP

EPKNLVK**T**NCELYEK

KAQEAGY**T**NATTTDQ

GREEEFF**T**YRSSVTY

SMILDIC**T**RYLGDVD

VKDSCVG**T**LVVKGDP

YRLVGEA**T**ISCWYSQ

YVVLARL**T**AQPAPSP

DEANQDF**T**NRINKLK

QDIIWDD**T**KQTVRVL

CQNDIML**T**ATCKSDG

LAKKYEA**T**LEKCCAE

NLLDNWD**T**LGSTVGR

SLKYNFV**T**PLTHMVV

DPPACYG**T**VLAEFQP