A Machine Learning Classification Model for Gold-Binding Peptides

Jose Isagani B. Janairo

Department of Biology, De La Salle University, 2401 Taft Avenue, Manila 0922, Philippines

Corresponding author email: jose.isagani.janairo@dlsu.edu.ph

SUPPORTING INFORMATION

R SCRIPT (CODE)

```
set.seed(1234)

library(Peptides)
library(dplyr)
library(caret)
library(ggpubr)
library(iml)

set.seed(1234)
df <- read.csv("datasetSI.csv", header = TRUE)
out <- kideraFactors(seq = df$Sequence)

output <-as.data.frame(do.call(rbind, out))
mother_dataset <- cbind(df, output)

mother_dataset$Class <- as.factor(mother_dataset$Class)

str(mother_dataset)

dataset <- mother_dataset[3:13]
```

```r
validation_index <- createDataPartition(dataset$Class, p=0.75, list=FALSE)
validation <- dataset[-validation_index,]
dataset <- dataset[validation_index,]
validation$Class <- as.factor(validation$Class)

control <- trainControl(method="cv", number=10, classProbs = FALSE)
metric <- "Accuracy"

fit.glm <- train(Class~., data=dataset, method="glm", metric=metric, trControl=control)
fit.glm

fit.cart <- train(Class~., data=dataset, method="rpart", metric=metric, trControl=control)
fit.cart

fit.knn <- train(Class~., data=dataset, method="knn", metric=metric, trControl=control)
fit.knn

fit.svmL <- train(Class~., data=dataset, method="svmLinear", metric=metric,
trControl=control)
fit.svmL

fit.svmR <- train(Class~.,  data=dataset, method="svmRadial", metric=metric,
trControl=control)
fit.svmR

fit.svmP <- train(Class~., data=dataset, method="svmPoly", metric=metric,
trControl=control)
fit.svmP
```

```
fit.ann <- train(Class ~ ., data = dataset, method = "nnet", metric=metric,
trControl=control)
fit.ann

fit.svmR1 <- train(Class~ KF2 + KF3 + KF4 + KF5 + KF6 + KF7 + KF9 + KF10,
data=dataset, method="svmRadial", metric=metric, trControl=control)
fit.svmR1

# predicting using best model
predictions <- predict(fit.svmR1, validation)
show <- confusionMatrix(predictions, validation$Class)
show

#Feature selection
x <- dataset[2:11]
predictor <- Predictor$new(fit.svmR1, data = x, y = dataset$Class)
imp <- FeatureImp$new(predictor, loss = "ce")
imp$results
plot(imp)

ggplot(data = imp$results, aes(x = imp$results$feature, y = imp$results$importance)) +
geom_segment( aes(x=imp$results$feature, xend=imp$results$feature, y=0,
yend=imp$results$importance), color="grey") +
  geom_point( color="orange", size=4) +




  theme_light() +
  theme(
    panel.grid.major.x = element_blank(),
```

```r
    panel.border = element_blank(),
    axis.ticks.x = element_blank()
  ) +
  xlab("Classification Variable") +
  ylab("Variable Importance")


ggplot(data = imp$results, aes(x = imp$results$importance , y =
reorder(imp$results$feature,imp$results$importance))) +
  geom_point(size = 5) +
    xlab("Variable Importance") + ylab("Classification Variable") + theme_bw(base_size =
14) +
  theme(
    panel.grid.major.x = element_blank(),
    panel.grid.minor.x = element_blank(),
    panel.grid.major.y = element_line(colour = "grey60", linetype = "dashed")
  )


library(cowplot)


a <-ggplot(data = fit.svmR$results, aes(x = sigma, y = Accuracy)) + geom_point(size =
5, alpha = 0.7, aes(colour = C)) + theme_classic(base_size = 14) +
theme(legend.position = "none", axis.title.x = element_blank(), axis.title.y =
element_blank()) + xlim(min = 0, max = 0.1) + ylim(min = 0.7, max = 0.85)
ggdraw() +draw_plot(b) + draw_plot(a, x = 0.11, y = 0.12, width = .35, height = .35)


a <-ggplot(data = fit.svmR$results, aes(x = sigma, y = Accuracy)) + geom_point(size =
5, alpha = 0.7, aes(colour = C)) + theme_classic(base_size = 14) +
theme(legend.position = "none", axis.title.x = element_blank(), axis.title.y =
element_blank()) + xlim(min = 0, max = 0.1) + ylim(min = 0.7, max = 0.85)
```

Table S1. Training and Test Dataset

Sequence	Intensity	Class
HFSSWETQQG	384446	A
WTHRDASTPW	383659	A
WYEKWQKANW	378302	A
WMETKWQARA	373760	A
GTWSEHQNGW	372844	A
ETWSMQQHEW	371294	A
WRAGQAQMQW	367860	A
WKPWMEPQHS	367310	A
AMQQQWEMSQ	363157	A
RWQIEEHFAP	360512	A
PEESQEGWMA	358214	A
TGEWGMQGIH	356816	A
EEPHWEEMAA	355438	A
WWKVANIHSK	345972	A
RHWHSWTWEI	344942	A
MNWKWGLESM	342485	A
NWTAKWTQTH	342381	A
HWIKIPPWMW	341846	A
HWKQKVHWWG	341349	A
WHKWWTHGHW	340044	A
KYWQMWMSWK	338940	A
KWQWKQAGAQ	337810	A
EHQQWKETWH	337270	A
GQWQWMDAGW	336102	A
QWTWKIQVMK	336075	A
TGHEPMEGWG	334770	A
HWKGEMHTDF	332748	A
PEEGPHSLWH	332184	A
EWVEAMGGHT	331057	A
WPAMGWNMEQ	330589	A
KWAIWEMKGH	330058	A
GETWETHYSE	329553	A
WVHKRLNWTV	329449	A
WSWPKVKSFW	329447	A
MLGWMHQSWQ	329003	A
NWKWQMKWTQ	328402	A
EWHVKWSEAI	328377	A
LAGVPMHWYT	328328	A
QEHLSEMWGE	327416	A
ASHQWAWKWE	327039	A

WSEETEMWPL	325272	A
DMVWHESWGI	323649	A
WWLQKWHGSH	323382	A
ENHSWGGGGA	323376	A
QRHSWGGGEA	322797	A
WKATWAKYEK	322695	A
TWHIMWRHAW	321358	A
WEAKEWLHNW	320623	A
NKGGGWQGPE	319988	A
WGWKWEHSEA	319819	A
TEKLWWHNGS	319782	A
QTEWWTYRKG	318708	A
AESEWSMMMW	318118	A
KLWEVSRGWG	317439	A
WMAHAVVSPA	317021	A
SMEIQWQKPS	316501	A
RWPWHWQEMM	316490	A
EKSSKSMESE	316344	A
MPWESHEISE	316121	A
VHRGWSEMVW	315620	A
GEHMEHMEWS	315460	A
EWSSQRRWGN	315172	A
SRQWQTWAAE	314740	A
AESEHEWEVA	314288	A
TQWHEWHWYQ	313682	A
SQYWEHEVHQ	312935	A
HGHMRWWWWQ	312299	A
HIRGEHKWNW	311701	A
EWKHPIWWER	310267	A
PWSWHEIYLS	310086	A
WMNWESSQKH	310018	A
GWWARWMWMQ	309848	A
WWERGAGRGW	309840	A
VLWRHEWAWK	309798	A
TKWWWMTRVQ	309764	A
WWGMTMSKEK	309480	A
YWASHKHWWW	308871	A
WGEWHSEGLH	308636	A
REWWKGFHMW	308628	A
GGWHKLDESE	308575	A
REWGWAKSGR	308462	A

WQENSMEENW	308362	A
WGMWSHAHTG	307131	A
WHWQHEMESW	307090	A
RRWAISLRMW	306510	A
WEQSHVTQWE	306318	A
GMEGHWSRWW	305771	A
WEFRSHIHMH	305745	A
FVEGKSWGGL	305238	A
NLWMQGGIML	304992	A
WWHMQGWAWH	304546	A
EWHHNWEEWM	304536	A
GYHHWWGQPW	304468	A
SMWGTTQGHT	304459	A
AHALWIWHKT	304010	A
WSWESPWEWH	303839	A
HWEGHSWTIA	303515	A
QWGRDTMWWM	303191	A
WGVGSEMHHS	302955	A
HKWMHIQWTT	302952	A
WQKHTKWWWM	302933	A
EEAGRWYSPK	302790	A
KGWSMWQWGW	302569	A
APSWVNSMEP	302503	A
WKWWNKEVAE	302479	A
AMQWRQWSGS	302423	A
GHHGWVWFMM	302327	A
QSGSGRSWMM	302180	A
NEIKWAQWQW	302015	A
ESMHSWHWYT	301875	A
KSWNVGGAEW	301554	A
NWELEEHSAS	301315	A
WHEDHAWRSL	301201	A
GMWHIEHIWL	301014	A
GKTWDPWSGH	301013	A
WHHWAQGWHG	300557	A
YEAVSTTWQS	300456	A
WWKHTATSSS	300162	A
WKHNKTWGSS	299746	A
YGAWKTRWWM	299229	A
AHKWGWAAEW	299107	A
VWAFNMMEHQ	298988	A

WTEWTHIGAE	298934	A
WESDTIKISH	298634	A
TTWHGFPWAG	298329	A
WREWSHRWPW	297938	A
YWYWPQMGWH	297926	A
WWKHSQVVAH	296970	A
IHWQVMAESE	296580	A
HWYTTSGGDS	296490	A
GWALMHDENG	296316	A
MVWSLHHASW	296265	A
NHKTPAWMLQ	296113	A
RWWEMTMESL	295888	A
SWWTLMTMGW	295451	A
WEWATKPKEG	295345	A
VHYGSQIEWG	295165	A
WLKWKEKRGM	294735	A
ETEWLGHETL	294621	A
SLMKWMALAE	294574	A
WEERWTGTYE	294531	A
KWQEPWWWAT	294477	A
FAQDDWLMRH	294377	A
LWAHSEKLGG	294095	A
SHWRKIRQMT	294014	A
SKQMGTKWRY	293954	A
EWEWKHHGWK	293890	A
EWHVMWTEAL	293814	A
LWMAGKIGPS	293726	A
WGESVKIIHT	293660	A
ETGHHIWEWM	293419	A
VSWHLGKYKG	293364	A
SQWPWGKETS	293245	A
EKQWVGQGAS	293225	A
NWTHWSTTQH	293033	A
GHWQEIGDGV	292661	A
HRHWEWEAWM	292499	A
EWLHKHWWQR	292416	A
RGEHWKAYWW	292327	A
WIEHAWSRHL	292235	A
GWTWQAKKMA	292078	A
GEKSYWTGRE	291909	A
WSQKRWEIGQ	291480	A

WETVRSMSWW	291399	A
EGSDHPSWNQ	291201	A
QHEVEMMWSL	291055	A
KWWEESALHE	290920	A
GWKWEGAPWE	290887	A
EQKAEWWTHW	290618	A
GWSGAHWSEW	290594	A
WEHIWHKQAW	290411	A
SWLQWKWRGH	290209	A
WMMWGWVHEI	290172	A
FKPAGIWWNQ	290147	A
MTSTNEPYWA	290084	A
RRRWEEEAHL	289915	A
WISWEKVAAV	289790	A
LHWSSISWQH	289406	A
ETWHSWSVKW	289177	A
HWWWEHEMEH	288862	A
MSWHWNMTGW	288613	A
EPNWHAWSEA	288453	A
HQWKESHQGE	288356	A
IWQKWSKETE	288277	A
LWSPWAPKML	288157	A
ATADWSPMDY	288099	A
ENWHEMAQRG	288096	A
WMAVDRTGSG	287617	A
WWSLEVQEPH	287554	A
WSAVQRWHTH	287451	A
WWHGWVQMGW	287377	A
KGHIWKTWWH	287325	A
PVWQGWWHEV	287037	A
THMQAWVIWE	286976	A
MWSHIWYRQS	286753	A
WNEWTHNWTW	286600	A
SEKYHRWGLG	286345	A
WWEHGWHWGQ	286329	A
STGHKVWWPG	286056	A
WQERMLEQRT	286002	A
IHMWWGYSIP	285945	A
LEVRKEWGAE	285842	A
YWAMGLHWWT	285686	A
HWYAWKELHN	285239	A

SEWWMSHAFE	284932	A
WSWTESATEQ	284916	A
REWAEKESWR	284784	A
HRGWETEAWK	284705	A
PHHAEETLEG	284628	A
WQHEEYWWHW	284628	A
WWKVREYINE	284443	A
HWWSHATLAM	284397	A
SIWKEQMESQ	284358	A
QQWPHEHEWW	284105	A
EWASEWVHER	283541	A
TWLHMGEYPG	283423	A
EGTQHSTWQE	283422	A
NLTGKWSWTR	283407	A
RQGEWPHTMW	283316	A
SHHAEEWMEH	283180	A
WSWHWWHQAK	282756	A
AQWWTMTWEE	282413	A
WMEETWSSMR	282400	A
WSQIHIHAMH	282160	A
WSHMHRVAIG	282092	A
SGARQWEQWK	282079	A
VTWHMGPYPH	281679	A
AWEAEHRQIS	281496	A
RWSSHEGIHS	281220	A
WSGWGNWNHK	281157	A
YWNVGEQWHH	281040	A
WWQMWPQESH	281011	A
EHQISESPWT	280996	A
AHWNDHGIWM	280898	A
GSSEWRQSWP	280887	A
YGHWSTNEKA	280858	A
WSIHWTEEAS	280692	A
HHTDQTWESG	280414	A
WFEPGSWGMD	280308	A
TQSWAHQRLQ	280282	A
HPWHYQEWGM	280119	A
GWGKNTISQW	279782	A
MTVWKMEHWG	279494	A
GWSTKISWEG	279483	A
WRGQQYTWHT	279476	A

WWREHMQERM	279320	A
PHWHWAHWIE	279311	A
WSWEWQMWAN	279295	A
TKEPGWMWDY	279231	A
SWIGKGEYKE	279089	A
HWKWHEEDQM	278947	A
GEGWEYEEAH	278866	A
MEEQAWKHHG	278818	A
KWYLSSRRMA	278589	A
WWTESETTAQ	278512	A
EWKRHKIVFW	278465	A
EWWTERNWHQ	278278	A
SWVWMSGHYG	278267	A
HPWAIHGMWR	278242	A
GMMEWAQPWK	277972	A
WWHWGQWQHM	277892	A
MWYSGPAYIS	277804	A
ESGEWHSGAH	277712	A
WHGWEWWGIH	277643	A
WVEQSTASEE	277634	A
HEAHMSTWEG	277457	A
NHTGQHWQIS	277116	A
MIEENKWRTT	276976	A
ETWEIWQQPT	276799	A
AHWQSESHAH	276674	A
MMQKWYMDQP	276517	A
MSTQMHHWAG	276468	A
GPWWAQIHGR	276243	A
HEQWMYWWWH	276138	A
AHAKWLWHMW	275891	A
LYLEWQWEKH	275809	A
WMEVRGHSKE	275558	A
WRIISHWTHH	275358	A
TTEKWTWAAW	275167	A
VEAWHEAATG	275140	A
EKTWARHGEA	274712	A
SWSTEMGNYN	274305	A
TWAITWMGWM	274098	A
WWHGSHVGKM	273993	A
GKWHHEGHWI	273979	A
MGVNFGWHAE	273877	A

HWMEPYHWWH	273853	A
GSLWEEAKWE	273820	A
NHEWWEVKWR	273764	A
AMGGPIWKEA	273709	A
EGSHWGESAA	273644	A
WVGQKWWHHW	273439	A
HEYGKMSEQW	273325	A
QSYWHETWGL	273149	A
PKQHDWDLKM	273120	A
EHGQFHHWWK	272725	A
GDWQTLRAEW	272654	A
TWWGQKVWWE	272324	A
GWWARTLSKR	272276	A
SWTWHELYKW	272258	A
VETWGWPMNE	271900	A
AHWHFEFMWW	271844	A
WEWEGQQITH	271674	A
TWHGMEWKSG	271586	A
GTWWEGQKEG	271585	A
IETMKWRSGW	271244	A
KWTSWASMSM	270984	A
EKKHQIYWWW	270872	A
LGMGWHEGHF	270871	A
WVEETWATGE	270795	A
RHVNWRWVPS	270664	A
LDWHGTHWLW	270661	A
SNMRQWGQWI	270644	A
QMGGVASGNW	270636	A
GKEYWMMPMW	270598	A
HEPTYWWGNE	270577	A
WEVEFEKHHP	270534	A
IWKAELHEKM	270472	A
WHWEIESEWW	270359	A
WKVGWGMWWW	270257	A
AWQAQHADHF	270236	A
WRLGMLWWSL	270180	A
GHWELTWWEE	270029	A
EIVRAKRWSG	269816	A
EHWPWSMWTM	269711	A
REWWPHSHSW	269674	A
WWTWESGQYQ	269661	A

DHQVQLKWMI	269612	A
HGSRNSWEMG	269586	A
HLWLGHDWSS	269578	A
TWWESHIWAH	269486	A
HEGWALNWWA	269334	A
TELTKHAGME	269064	A
NWFIGWQHMG	269048	A
WHHWEIEGGD	269043	A
MTHWEPMRAE	269043	A
SEIYTEEWKA	268924	A
WHRGTWWWPW	268920	A
GWWPSKSRES	268608	A
HAMYHWWANI	268603	A
PDYWQGEVHN	268544	A
LSEESHWPWQ	268292	A
MHWPHGHTWW	268279	A
WMLHHEQHGT	267810	A
GSEEGMWSHA	267799	A
SSWEHSGEGG	267606	A
QSGWQYATSW	267546	A
ITVASEWEPW	267316	A
WPQLRSEHGQ	267270	A
WQMGWHWGGM	267175	A
WMWQMTWIHQ	266812	A
WPESHWSWQY	266480	A
WWLGSAPMMP	266058	A
VITAHQTAWL	265830	A
SWTWHWREHH	265721	A
IWWMHWWHKE	265617	A
SALSHHEFTW	265177	A
GWHHHASFNT	265057	A
HVRKWEASKW	264998	A
HSTWWEWEQH	264733	A
DGSSTEVLWK	264675	A
SMEKSWGSHE	264360	A
WGWEWTSGHV	264174	A
AEWVPHHEGG	264076	A
SAEQQWKSRH	263801	A
EQIWWHGHGR	263518	A
WSRHRVVTSR	263507	A
AGHWEKHEVA	263444	A

VSSDEKHRSW	263382	A
HHHQWDAGTA	263304	A
AQMQGQWEVA	263194	A
SWWSEWGRHG	262989	A
EWLPWANSWS	262787	A
IHQEQQWAYE	262787	A
WSEWMWEARR	262784	A
MMWRGMEQEE	262738	A
GWQGIKKWWW	262532	A
SSQHWWVWHH	262428	A
HEYQPGEWWI	262418	A
MMEEKHHLEV	262372	A
NWWTHEGLIW	262370	A
WGAYAMLWKW	262338	A
HKWHYDQWGK	262265	A
GHALYANIGY	261737	A
HSVAFESQKW	261677	A
HEWMEWHQHE	261632	A
AIGWRGNKEW	261559	A
PRTRHWQWWP	261543	A
WMSKEWGEWE	261519	A
KISLIWSWGG	261296	A
GGWWEWHYTE	261197	A
GQGEMHWWWP	261041	A
SEIWEVKVHM	260929	A
AGHWQLHQVA	260899	A
THHTQHREER	260823	A
WGHWANESHE	260640	A
EESHSEWMWH	260547	A
HPMGVGEEPI	260412	A
KWYPGKAYHP	260322	A
HWWTAKWWIS	260309	A
PMWHEWWWEN	260225	A
HWSKHSWHIE	259961	A
NQQEWMYNVQ	259914	A
RWEWTEVLHD	259906	A
SFEWWHQIEH	259895	A
LGTMSSSWGH	259780	A
EKYHINIMWL	259745	A
WMLEREATWW	259724	A
NEWAHHGRQH	259715	A

KAHYTSIWYW	259617	A
ERMTAWRWWE	259598	A
EWMWWHTQGA	259354	A
ESREHLISTW	259262	A
WSYEASSDGA	259243	A
WTYGAWTQHR	259132	A
GQWMWSLHWH	259089	A
WVHMDKQWWV	259071	A
WWMVRQKHTM	258490	A
LGSWTHWMQW	258476	A
VTSLNTSAWS	258432	A
GMWGWKWQNG	258416	A
HLQHYEWWTG	258214	A
WTHEWVDLGT	258111	A
EWTKMALALS	257985	A
TSGWWHELEH	257855	A
WGHTARESHE	257741	A
RAWSAWEWWN	257696	A
VHWQTVLLVM	257645	A
ILRHAWKERT	257349	A
EWKTYRQWWT	256906	A
SWGRWKTHGS	256663	A
SWSQVMHKWV	256586	A
WSGPWWHWHG	256561	A
WHPEPALWWW	256384	A
SWEWWHLEYM	256267	A
TKWLWSGVEE	256260	A
LEWPKWDTGW	256192	A
RHHHHHWEEW	256096	A
SGRESWSWKT	256016	A
AWWGHHPEWN	255942	A
EWWEKWQESW	255916	A
WHMWWWHHIW	255760	A
PEEWHGEMGR	255732	A
PWESTGIQYK	255646	A
WLRSQMMQIG	255606	A
RLHWMWWTSL	255594	A
QHVRAIAWKE	255487	A
HGWHMKQETG	255446	A
SVEQSWHAQA	255412	A
SEKWWWWMRR	255197	A

QWVQALGEGS	255192	A
PAEPWWHHMW	255130	A
MGRWNWWGMG	255054	A
GTVWWHWTAS	254718	A
KTMWHEWWMG	254703	A
QEWESWAGHS	254544	A
TMWMAIEETM	254468	A
EKRWHATMEA	254396	A
TSSWWWGMTW	254357	A
GEMRLLWHME	254194	A
WPTEFSHWPK	254191	A
MWSWHWAEGG	254137	A
YWKHNHWSWW	254105	A
GAKYHWSARH	254093	A
WWHHSEWFTH	253999	A
RWHATRLWSK	253781	A
HTRLTSLQSE	253695	A
EKWWANHGEA	253655	A
HEGWAIRWTA	253642	A
HQYGLMPQNW	253586	A
HGVHWMGISL	253403	A
AHWQTEWHAH	253386	A
WESWWHWKTY	253231	A
LGAHHRAPEA	253099	A
AIWQQIHLWS	253005	A
TSEHGWHTGD	252915	A
PWTVMLKSSR	252830	A
GIQSWWGANK	252737	A
MEWWTHPTWH	252692	A
WEWHEWHWYE	252645	A
HKWVALPQEP	252310	A
GWWPHASIAK	252255	A
LGTKHEGMWS	252175	A
WLTHWIDHER	252131	A
SELIWTHREP	251878	A
ESLSWGSDEA	251804	A
MVSLEWWHWE	251750	A
GWDSSWGRWW	251657	A
HEEWSYWWWH	251365	A
QDSHWISWRV	251252	A
LVHGHESRWQ	251121	A

WPKIVWIGMW	251052	A
LWTEQPAIHQ	251046	A
RDWALTKNSW	251009	A
WNQSYNWSPF	250831	A
WQHLHIHGEQ	250797	A
WSAPWMWTGG	250776	A
QMEWQYASMS	250744	A
IRHSMEWWVT	250482	A
VLPSWHSSHG	250345	A
MPSHGKKSAE	249801	A
VHHRHGRWHI	249777	A
WEGHWHLNAS	249579	A
SMAKTKWSEE	249435	A
WTIDLWHGPG	249430	A
WLWWEMVRAV	249355	A
ELSSRKWMEG	249139	A
WEEMVLHHHS	249003	A
GWEGSYPGKT	248790	A
IEYGMSTEEW	248680	A
WTELHLIASH	248349	A
TAKWHHESWY	248346	A
WKWIWLQHEN	248249	A
PFWQGFEEEE	248212	A
IPQEPHWKWQ	248102	A
WKHLGRVAHE	247983	A
AWHHKVAAQE	247701	A
GEWAWWGKAW	247455	A
QSWSKRRHQW	247378	A
GGMQDPQGSW	247345	A
AMMKSMTHWA	247184	A
SEEWHHGSGN	247116	A
MIFWGSRRHG	247097	A
ARWEERSNQE	246977	A
HYWWGWWPRH	246789	A
WHMRWSHHHA	246597	A
NNYISGWEWM	246457	A
WIAWWSVMHK	246374	A
RSWEKSLEPI	246367	A
LYITKQGELY	246278	A
ARWEWAIMEA	246119	A
WSAEHKNEAK	246045	A

AWLHSVAREE	245935	A
KEYHRTQIEW	245700	A
WSNHMLLHKM	245504	A
LMWGWRKSES	245469	A
EYIWHHKRMI	245400	A
WLSWRLQVAE	244965	A
WHWAGWWAWM	244885	A
WWHLSWNHAW	244718	A
MYKGWEWEMH	244573	A
QFKMEFKAAE	244566	A
HTLRKYHWTG	244561	A
GTPGAGWMEW	244424	A
TLWMEESETE	244420	A
ATEWSEVIHR	244220	A
AEWEWTSSSY	244206	A
SKQGEGWLEI	244125	A
PSGEGQEWWI	243899	A
MKTGQWESHM	243744	A
SGKRWPGHGA	243689	A
EWARWMHMHW	243554	A
AWIHMVAREE	243435	A
SSWGWHELWE	243419	A
SMHQHNLEIQ	243303	A
WWWETSWEWH	243112	A
NWLWSQWMWT	242948	A
SEIYTEQWKA	242803	A
SNPVRKLMWS	242786	A
YTAKGLHWTT	242656	A
TMWKIVIMKW	242399	A
GEEGMEQEFE	242360	A
WKQVREHMLE	242303	A
MWWWWIRTKP	242295	A
LHEEEEWAYE	242111	A
FWMSMWTMQH	242079	A
WEWWAHERKE	241956	A
AGASWWKPHH	241807	A
WPSHSTHQAI	241759	A
YMSSVDWGYH	241753	A
HSEGIWTNWW	241633	A
GWAKAHREIG	241500	A
AHEWAGRLES	241494	A

NWDSYWSWNR	241482	A
QQMGMETKWG	241444	A
WMRHKIIHLK	241384	A
SGKEKAIWWT	241186	A
VSHWHWHPHA	241043	A
WSWKWSAWAL	240983	A
MYPAMQEAAW	240896	A
AWTWSWWKRE	240678	A
HLNHEHMWQW	240628	A
PHMEHWAGGW	240591	A
LILQSGHKEE	240362	A
AHWHKQKKWW	240190	A
MQWSKPGAKQ	239931	A
GGWGLIRQPE	239818	A
HGAWSARTMH	239697	A
WQAKDWLHRT	239672	A
WVQDFTAPEQ	239418	A
NKEWEDAHWP	239263	A
GWLEEKHSRK	239179	A
PSWEPHHWAG	239065	A
MESIMWAKLS	239063	A
LVVLRLRKWD	239020	A
SHIVWTVKGG	238828	A
WYWELHQWWS	238776	A
WQWMMMAMWR	238642	A
LHKWEKAAGE	238586	A
SKAKSLWPEE	238311	A
TEKTWHTISY	238232	A
REHLWADTQT	238210	A
KSVTKKEHWG	238155	A
LSLSHQSWLE	237816	A
NHWHVDIWRM	237803	A
IHTMMHMWRG	237574	A
VPGWGSGKHA	237504	A
HTTSALWWHP	237436	A
HGMYSPWWWN	237315	A
MDQGMHSLWH	237255	A
LGMEEHAWQK	237230	A
TWRLTKNEPG	237200	A
RTHPTWGYLW	237048	A
AMHDSATLWE	236927	A

HGHEWKHMYS	236805	A
ETDEHLLSWW	236740	A
HKKVHVWKKW	236585	A
DWIHKHTWRA	236432	A
EHQWEQITHT	236148	A
THLMWEYGLG	235892	A
WRHHPHWPHG	235679	A
AHMEVWPLHV	235626	A
HGHSDWWWWE	235605	A
YWYWSESGYH	235481	A
QWAPQSVHEA	235441	A
EGWSGRRELS	235436	A
WEAFEWMHQW	235362	A
PSEWIATHGH	235340	A
HIGPRMWHNT	235179	A
WSHEMGVGIL	235082	A
EFPIMGGAQE	235012	A
GWWTGWTKRH	235008	A
PKWMKWWWGH	234830	A
LRHTSEWWVW	234767	A
IFWVAMTEGT	234650	A
MGAARSRRWI	234446	A
QGSSTEVLWK	234383	A
PHMELEGAGN	234358	A
MSSMRSEAGE	234309	A
LREMEMSSSW	234229	A
NDWKHQHEST	233998	A
QGRWWWWWTT	233911	A
IEMTSHWLDS	233738	A
THLVWWVMHH	233725	A
WYGMWEMAQW	233710	A
AQWEWAIPER	233642	A
LWDQKTTSAP	233592	A
KMENGRQWSH	233588	A
NWWEFALWMG	233521	A
MFGSPLQEGL	233475	A
RDQQWKYRVD	233465	A
RWAMRSYHEA	233442	A
HVDFWGAWMW	233186	A
RRYHMGWQWK	232972	A
GWHGTTGDKS	232824	A

LEAGHETGGA	232770	A
GDGWAHRTTA	232760	A
AHEWGENWWW	232713	A
AKEWWAMEIW	232431	A
LKWGWRLSES	232025	A
KLTKIWTWGG	232011	A
SWWVSMMWWQ	231875	A
GWWWEAMFSA	231433	A
WSMLVWLHSW	231100	A
EWSLKAIAHM	231013	A
EWGSWWLGEK	230981	A
GGHKRTWTTD	230903	A
PMTESEMQWD	230866	A
MDMGEQRTAH	230742	A
IEAKLKGTWY	230561	A
TEQAWPESYQ	230554	A
WYEKWRKARW	230385	A
MDGTGSTWWW	230285	A
VKSTWIWWIH	230213	A
PTWQVHHREW	230204	A
MQRLWRWWEQ	229899	A
RWKWHSRQKK	229893	A
KVWFLHHAFW	229846	A
KHWGTAGSHE	229834	A
VEWWGWSSQE	229805	A
GGGNWRAGSA	229689	A
MEYHRWELEW	229575	A
ESVTSHSMAM	229532	A
WTHEPHVGLK	229488	A
KRNDDMSAHW	229343	A
VTKTTYWEHT	229293	A
MHWETWSHHW	229213	A
RKEKWDVLLK	228854	A
WSHEHEMGKE	228524	A
KAWWMAGKHP	228346	A
QSWWAKWDWE	228278	A
MLSSQSWWWQ	228093	A
VWSLQWSAWT	227785	A
RGATWWWWTT	227748	A
GWQTSWGRWW	227584	A
EITKWSLTWE	227394	A

FYWESWSTWP	227373	A
RTITMQWLWS	227370	A
PSWHHVTHMN	227065	A
EQELNMTMHG	226991	A
ASGKPWEYHS	226980	A
WLGLAWTRHK	226568	A
IQASKRSTWG	226469	A
HALELMGILS	226397	A
WPWMWWHVEE	226318	A
LGSKSSPWGH	226287	A
MHSEGIAWEM	226133	A
VIEYSGQFQW	226110	A
WWEWWHLGAE	225757	A
WPAWWRMYGM	225536	A
KHWGTAGTHE	225484	A
VWWQEMIDWW	225462	A
EMKGMTGWYP	225337	A
FGGRKLNESE	225179	A
WMHTKATKRK	225167	A
AIESWAKEHS	225151	A
PGILSNYEHG	225148	A
PWGHWSEHHS	225079	A
RLMEWSEHEN	225079	A
WKTLWKASAH	225063	A
WVHLRIRWTV	225055	A
EESESAWWEG	224998	A
HEYNKEEWTH	224665	A
EWEWWWYRMH	224462	A
WYWQIGQWWM	224093	A
QHENKGHWWL	223953	A
VMWFTWWGRA	223909	A
MWPTHQHYIS	223754	A
STGWAKGKWM	223717	A
ETLRHLHVKW	223702	A
LRAHSRSWQR	223683	A
AQTWKGMHMW	223607	A
TRQWPHRTMW	223207	A
ESGGRWEEWL	223111	A
GRWRSLRAET	222851	A
ITWALHHSWR	222839	A
GKLWPWLLRH	222649	A

WIWNMLNESW	222305	A
SWSHTRQTLH	222130	A
MFQGWTVTGG	222066	A
IAGYMMHWYS	221931	A
GQYRKEEWSH	221918	A
GSAHLHRDRE	221904	A
HSMPRWIEIE	221893	A
RMMHHMMDLQ	221670	A
RGKSKWMHSE	221581	A
LVVLRKDKWE	221495	A
IRWHESGTIW	221494	A
GREHEMWEML	221480	A
KGASMFGWWY	221472	A
WSEEWWTWSR	221221	A
HIDGYEWSSG	221182	A
WWINQAWWSW	221120	A
SWIHINNKEM	221091	A
IHTGYWQHFH	221077	A
WMMHGWTWWP	220948	A
AKEEKHWLEA	220872	A
HWSTNWKGKG	220800	A
TQSVRMLPWS	220788	A
LAMEKSGLMW	220568	A
AFASWLPHTT	220545	A
WWGKSLMLEL	220503	A
GKSRDWEGNG	220503	A
QISLWFLTWQ	220328	A
LVMIQWWGTE	220232	A
HSERKVTWEL	219882	A
MGWSWWTWGH	219832	A
WWWWKSWSHW	219789	A
AKQNDWELMD	219542	A
MWTERSKGRH	219412	A
AWHWIELEWW	219387	A
AMHHQEAWQW	219264	A
WWREYAWTEW	219193	A
EWWTSEEIEW	218988	A
KREPEMSSSW	218917	A
HEVATAGRSW	218899	A
KQRRVEEWNT	218828	A
QVPAWWEVHR	218764	A

GSWWSWSGWG	218675	A
WLAWGEWEKF	218626	A
KMEWHAPHGG	218185	A
SWAHSTKGSL	218115	A
QMEETGMEAH	217960	A
PHLQHRKWHS	217868	A
THLWWTHGHW	217636	A
THHWEHQEER	217572	A
WKSNKSTHHQ	217557	A
LWEMWWMEWG	217498	A
HRTHEFGSHW	217124	A
ERPNSLKWRM	217059	A
TGMYHDWGKH	217057	A
QHRWENHSHS	216487	A
LAKSHWRGHE	216460	A
KWEHSMEGGP	216454	A
WGWAETWASL	216425	A
IVHGHESRWD	216409	A
AKMHMAAEHS	216235	A
LWHWTWESVQ	216054	A
WAQWASWMYK	215853	A
HWPTEESSGH	215706	A
HSMSRWLELE	215645	A
FRDWNSWAAE	215371	A
KYHWLMWHEK	215359	A
WWSQSQTSAN	215250	A
SEEERWPIHG	215203	A
EMYHLQIMYL	215067	A
GPKAAGMFWA	215028	A
RTKAEWSMWA	214979	A
SLHGHQRHEM	214659	A
ASSMTSWHWA	214437	A
HTYASLEIGR	214367	A
MMHSQHGTHA	214362	A
HEKESHQEGK	213977	A
SHGILMELQH	213963	A
WWWETEWWRE	213684	A
EWWWAMWQWE	212933	A
EIQLTETSWW	212931	A
YEASIPRWSL	212753	A
EMFHEKYWWW	212699	A

TWLYARIHML	212690	A
KVVKNMQMWE	212402	A
GWEGWYSHFW	212367	A
QASHWARTHP	212188	A
KKTHEWWTQR	212177	A
WWWTWVEPWS	212138	A
WIVEWELTTT	211917	A
LWAHSQLLGG	211868	A
MLDGQETIQH	211843	A
SGLATMGHGA	211781	A
MYLWFEGGKY	211705	A
HWEWWWGRWW	211529	A
PTHGKEQYKE	211460	A
EGIETRGDHE	211223	A
AMFHFAAEIS	211171	A
MSHRFSIQFE	211054	A
YGIWTWQEFR	210899	A
EHVRAIAWME	210768	A
AEWVKHHQGG	210335	A
GGAIYARHGY	210300	A
SHERGAWAPI	210110	A
MPHSEHHWHA	210084	A
YEAVMSSTDF	210046	A
SQDMGVPDWE	210014	A
PMMWWTGLST	209909	A
QQSSSEELWW	209884	A
AYKIWMNMTK	209737	A
HHSVGHDVRW	209151	A
NSPGHHEELL	209018	A
WQQTYQWWTS	208972	A
SMEKTWGSHE	208953	A
HMTDQWGGQG	208952	A
LWYWWWHGEW	208761	A
NHWHWWWWEL	208575	A
PELTWWTLRA	208400	A
RMEQSGKEAH	208086	A
KGATRWTELE	208014	A
HMKWTWMKNH	207972	A
EYHWHHLAKH	207836	A
PMGYQEIQWD	207816	A
QWEMWWIEEL	207607	A

MGGEEASFGR	207561	A
IEATMQTWWH	207296	B
HVIWIEEQGW	207255	B
HMWRNSKDNT	207075	B
SAISHGQMSW	206846	B
WYWDHGRWTK	206822	B
VAMFSMSSWI	206530	B
RWFAGWWSWA	206248	B
KKEYDQHRWR	206076	B
GEAHKPSWQG	205866	B
ASAWWKTIWW	205778	B
TEKWMHAGPE	205661	B
MDHWNQHGKE	205416	B
KRSRHWQTSK	205121	B
YHMHWEKRLH	204885	B
HKWVAKSEGS	204879	B
NETAMSGHTY	204842	B
NHERKGHWWI	204776	B
DQMMMEQISS	204767	B
MTERSLSGGM	204766	B
PMEWTGRHQH	204624	B
QIRSVEREAM	204289	B
KHTHHPAWWM	204241	B
MWYKWWNNSA	204160	B
AKMLMKSHWA	203811	B
WSWSALEGWS	203664	B
EDLAETTSHS	203342	B
MSIDWLSAGH	203269	B
AWHWLEMEWW	203014	B
ATSTMWTLAQ	202972	B
KKYWWWYAGW	202703	B
KYMAKRQAAW	202372	B
KSPHHWYADM	202259	B
KHWEPWMHHT	201989	B
QKKHRIYTTW	201836	B
IGAWWARWSI	201577	B
SWFIQVNQKG	201287	B
WKATTALYEK	201086	B
GHWEEKHQGV	201033	B
WDLGSLWWTK	200830	B
ALEEKHTIQA	200811	B

MIGQRKSDIG	200502	B
GHWEALHHHM	200458	B
IWGWPWQMVN	200441	B
EDEEKGSWTK	200436	B
WMHLGRVAHE	200258	B
REWAHHEAQH	200246	B
MYYNGGEYHA	200133	B
RWRPYTMTRA	200094	B
WAHHMHTMHG	199386	B
LYYREEDYHA	199369	B
VWMWTYWEHW	199253	B
RGSHVRHWAL	199104	B
KVSRGSWHDV	198784	B
WKAHAYVPMA	198553	B
GGKRRMRGPW	198477	B
ASQWNEYWGT	198349	B
TRQMEAWVHE	198122	B
SWSVVFWIQP	198049	B
GWWMHAPHAL	197835	B
MSHHHRRLEL	197818	B
TTREYAWSET	197697	B
AHEWAGRIQS	197646	B
ATHTHQLQTW	197395	B
RSGKSWEYIT	197231	B
PLGEQMTQLG	197115	B
GQVASAEAMW	196973	B
HGKYMKWSTR	196890	B
PIGKPTHMWP	196598	B
SWWSWVQKTM	196452	B
TIAWGQTELK	195768	B
SVWTKIIRTW	195750	B
HIIVHVSIIW	195707	B
AMASTIMHST	195703	B
GWWWFKWWEG	195644	B
MRRWRPWAAD	195626	B
FEEQVGGWEW	195420	B
ADWWSKSTQQ	195280	B
WEWAWPTFEG	195050	B
YWYTMNKGWH	195028	B
ASSISARGLW	194956	B
FHHVWSVLGG	194933	B

MTMGPRNSIG	194826	B
AIQSWALQHS	194772	B
WEMWWHWLSY	194694	B
YFWHLGKYKG	194685	B
HQRTKYWWTH	194515	B
KTSTWHRSLK	194434	B
KHKWEMRRGE	194196	B
VIWEWYKKVS	194112	B
WHWNHEKQSW	193909	B
VHDGWWESVW	193659	B
AWDFPREWWQ	193599	B
WMWLWMATAI	193182	B
IIWEVPAGTE	193127	B
FRHHEKHMIL	193126	B
WSWSEPASER	192576	B
RGRVRLKWKI	192250	B
ASWQLSLQMH	192170	B
VISAGRSAWI	192162	B
RRKGMQSLWG	192082	B
WTHRRASTMT	191638	B
RTKHEWDHKE	191144	B
GSKPRQKMEH	191053	B
HHHLGAVWLR	190894	B
MELYHRWGIG	190806	B
LYVREEEYHA	190760	B
KMTESHQHSQ	190196	B
LGTKHEHPWS	190175	B
GHHEIHYVEP	189984	B
WWWWETAWGE	189925	B
RTKHETRHLE	189772	B
VEAWIERRWG	189738	B
PHGILKELDH	189584	B
ESWFQGNRET	189201	B
LDYHRSDHQT	189146	B
NMEMWQVKKM	188909	B
RHTWRGLKET	188780	B
WEVEMQLGGM	188669	B
TEEHTIWKWI	188642	B
SHHGEAWSGN	188567	B
HHNDRRTAYQ	188358	B
KFEQGNEWSH	188334	B

TVGDLTWGGT	188310	B
RHEPGWAHAW	187774	B
TDSTAHRRID	187655	B
GESTESHYPD	187609	B
STHTEHHWHR	187462	B
EEWATWGIAT	187081	B
EEESEYQEAG	186962	B
HMGRLHMVHH	186714	B
RGETWQVLWA	186713	B
AGRSEQRSSW	186626	B
AHPMDHQARE	186528	B
KEQPDEGWKA	186453	B
TRGRRYSWHT	186042	B
EMTTMTFETE	185950	B
TTEGHHSEQE	185757	B
GKWRRPLRRS	185699	B
IWGTHLTHQI	185637	B
FHLRASVHSR	185289	B
VALKKLMMWH	185041	B
MHTKQRKQFD	184927	B
NTGARSLWTK	184902	B
RGQSTRVIWA	184711	B
TFHHSSEEAM	184124	B
IWYWWSGGQT	184084	B
LYHTKRGEIY	183881	B
HIWEAKHIHS	183693	B
YSRVEQRTHG	183662	B
SENRHLEMSW	183650	B
KNRISRTWEQ	183417	B
MRYWDGDVHR	183285	B
MERRHIQKPW	183114	B
EAWTEFWSER	183047	B
DWEKTTHEDL	182883	B
KMEHIVSSKE	182867	B
KLFSRSWTWR	182750	B
HGRHMRPSEL	182717	B
EGWWWGVMWF	182622	B
WKAKGTRKEN	182620	B
SSQGGHMEDQ	182598	B
HHPVGHRVRW	182564	B
HQHLEQASRQ	182344	B

VHHRHGAWHI	182296	B
YLMKVRSGYH	182092	B
LTDSLDAGAD	181916	B
SYSASQEAAW	181868	B
ERHEHHIEQH	181790	B
RSGARFLWTL	181474	B
MHHAQQSLQG	181461	B
LNRKLSIVAW	181368	B
ETWGESLLRE	181232	B
ESMSESMAAE	181069	B
LERWSEWWWH	181041	B
RIHSKSWSMI	180862	B
MKSGMFRGHS	180556	B
SWMINYRDKG	180543	B
IMNQMHWKWR	180328	B
VHRGWSQKVW	179956	B
HGFEYSRGKG	179816	B
MSHRMPHRME	179748	B
ISMGVGEESL	179683	B
LLYTSWYAET	179333	B
TESGGEQWAL	179209	B
LLPERSQMHL	179140	B
HHEKALWHRS	177425	B
ELEVSKLKKW	177146	B
LHHKGEQYVS	177088	B
YEATLSRWSK	176566	B
QKGEVAMGRW	176420	B
SGSQSPKEGV	176327	B
VWWKEWWRWW	176206	B
DETEFWAEGM	175649	B
WMLHVWHGKT	175637	B
HWWSWFWNET	175554	B
QHRHPQSMTS	175407	B
WSGWGRWRHK	175124	B
TEQKVIHHHS	174761	B
LGASRTSELE	174614	B
SIHSLYTHSM	174399	B
SMTWEWYWWE	174095	B
PEPGGEDWAI	174025	B
GQAGLMSWRG	174001	B
KMRRGMKEIH	173353	B

QHRWWWWWWW	173228	B
SESQEDNHSG	172985	B
WLRVREHMLQ	172739	B
PSELSPTAAW	172718	B
AHWHKRKKTW	172676	B
QSQWIGGETI	172613	B
HLLVHVSLLW	172063	B
RLEEVAKGRW	171803	B
GAWWESWWGD	171530	B
ELWMWTMWWE	171496	B
TKEMWAAGKQ	171059	B
SWWEQLVTTE	170889	B
IIAHATKEAS	170848	B
GHHGNEHMSK	170671	B
GSQEGKTSHA	170545	B
WARWAKSLYI	170514	B
SSELTSWAAW	170174	B
NEHSHSKGEW	170164	B
AGKQYWMIHY	169991	B
WKRHLIIHIL	169736	B
KKQQLHHIEV	169502	B
QSWQIWRNMS	169432	B
IWGWILWHEL	169427	B
PGARDSQDWL	169422	B
GRVAPAEAKW	169415	B
TSMRRTEIEW	169370	B
KFEWAIELWL	169261	B
QAWIAAYARR	169090	B
QSIAHIHVKW	169030	B
SKWLWSGVQQ	168929	B
RHTWRGKKET	168766	B
SKHEESWEKG	168603	B
LAWLMSRIQE	168570	B
KRPTHWNHHE	168549	B
KETWSHKSTH	168535	B
KLKAEHSWRR	168329	B
RLEEGWNGKE	168232	B
STIDSTLESE	168104	B
TGLYWGEWFR	167969	B
PVEDMWGARA	167904	B
HALQIKEHLS	167893	B

ESTHMSMVLT	167865	B
ATQWPQVHHR	167753	B
WMRIRSQHGR	167640	B
IHHKGQRYVM	167639	B
RQSQMWAEGK	167317	B
GGAHYARHGY	167209	B
HSYASIQHGR	167082	B
VFTHVHHHSV	166832	B
RTNHQEHMAM	166699	B
IWSQNMAHHR	166358	B
SLLSESVHRH	166026	B
MHAIHNATGA	165663	B
HMSSTQTDRG	165562	B
KWKMKWSKQG	165518	B
QAWLAAYRRD	165275	B
KEQDVEEWQW	164951	B
IRRLISIVAW	164361	B
RSMGHHQDLL	164056	B
QYHWGHIALH	163889	B
AHWRPQSHAH	163876	B
MPEHLVTSKE	163739	B
QRDHRLSLGE	163329	B
LDMHLSAIHK	163214	B
AEQHTLAYTS	163185	B
SGTESSMEGV	162754	B
YHSHWGMRKH	162687	B
MQISWWSLAA	162289	B
VIMSWHSSHG	162202	B
EPWMNGRRET	162027	B
HMSTAKIMEM	161353	B
RHSELSMTMA	161341	B
LHTQMTKHHS	160598	B
AHSFEHEARE	160431	B
STLVREYHRQ	160259	B
HHSEQTEHTW	159713	B
GSMGAGTKQW	159496	B
WLAWWWVSIM	159181	B
GYAKAHRELG	159067	B
LHMESHMAAH	158934	B
AHAITHTHLT	158788	B
MISETISAAI	158755	B

ILYTSWYAES	158564	B
STFSQKGRWR	158351	B
GWHLHKKSKS	157821	B
KVSNGSWHQV	157744	B
YLGGKLDRLG	157488	B
ESWENFEIWE	157478	B
VATVLWITQW	157302	B
KQNMKTLVAW	157234	B
HMYFHGHHEP	157186	B
PTTTQRHKWT	157116	B
ASARWMKKRV	157052	B
LHHESHLRAI	157046	B
VHHGNMHAVH	157044	B
KKRIAWMGRW	156781	B
VHYGMQHQSE	156444	B
RHVRTRWVKP	156430	B
VMMRHDVMQR	156282	B
VASVIWHSRW	156139	B
KARKSSHGLW	156050	B
EHHGRQHLMI	155776	B
GKWGYRRTGK	155489	B
DKSDRKRHSN	155381	B
SQSRERRHSG	154979	B
RIEEGWRGKD	154477	B
EARQQLRGVE	154147	B
IESTKEQEEM	153952	B
PDEEESWALW	153859	B
HFYAMQSRKW	153781	B
SWWWENIFWW	153697	B
MLRGNESHRH	153278	B
PVWMWLLATE	153050	B
LYLETRTQLH	152997	B
WHAWWSVKHK	152922	B
ERHQGGHEDH	152725	B
EEMQPATTEG	152321	B
KTYMGKAYHM	152118	B
QWIMWARSTP	152036	B
AHTRRHGHWK	151903	B
LHSHHKASWK	151717	B
GHEKALTHRM	151585	B
VKTKSTWERA	151213	B

WANWAMSLYL	151205	B
LHMQFHKAAH	151186	B
QGWPGRAQIM	150550	B
WWELPLMLQI	149583	B
GMQRIVSSEI	149517	B
KIGWLHRMTQ	149489	B
MWPWGTAQGG	149309	B
AWNANHARHM	149052	B
RKYWHQFTGH	148926	B
QTIPAQMPME	148916	B
GEVGTLEHKH	148896	B
GSYHKELSYK	148753	B
RWTQKAISKG	148693	B
VMMRHEVSER	148322	B
HEHDSLHLYK	148295	B
GLLWMTLIRH	147942	B
AGQSEERSTW	147678	B
IEGHSIMTTH	147597	B
SGQWGKRGHH	147508	B
ESMTPQVNTT	147489	B
KLKAEHSWRD	147130	B
HHRGQHLTRW	147129	B
EENEPEHMQQ	147098	B
HHIKGAVWLR	147083	B
EHWTHRREKT	146920	B
RWSPHQGHHS	146780	B
YSRVGEDTHG	146780	B
MPSGDQMKSH	146542	B
RRRWEEERHK	146209	B
KAEEHGPHWE	146138	B
LFLSHRSTLE	146119	B
SNHHEKHMLL	146025	B
NTVNALGEGF	145933	B
AHVGKPQGMY	145860	B
THEHATSAHI	145778	B
RGTHVRHWAK	145759	B
VAKKMLPPWH	145653	B
ARKMGNAIKW	145584	B
HITVHIQVNW	145560	B
HITLGHRTPP	145493	B
MIRHIKMGMA	145418	B

THSWQKVAAV	145415	B
STGIKWRHAT	145400	B
HGAWSAASKH	145358	B
KSTRVGGAQW	145051	B
AQWVKHHNGG	144843	B
AEWYWSWWEE	144409	B
SDTKLKAKWA	144346	B
MPRQGPMELH	144040	B
PAQKEAWVGQ	143998	B
AHLDSKQHQR	143874	B
HHLMHAVYMN	143531	B
MRAFMMGRGH	142996	B
RILQSMEHER	142708	B
GSKAAGSSWA	142652	B
WEAYAKITLT	142627	B
GSAHLHRERE	142127	B
KVQELPWGGI	142052	B
RWKWHSRRLK	142043	B
LEGHTLPTTH	141991	B
SLNHLPSHSR	141931	B
SWRLSKRQMG	141845	B
IWGWMTRKVR	141813	B
EHTHWNHMRS	141369	B
TDHIHHHEQD	140974	B
IHIRFGHKEQ	140843	B
KGGEQASMER	140392	B
TSHRITHGMG	140282	B
RHWGSTTTQI	139444	B
GLTHHQGHTH	139345	B
SSIYRQYHRR	138875	B
HHGQMGHAAH	138515	B
MSEWALELWL	138375	B
MQHSEVLVHK	138291	B
PQIIWTHREM	138073	B
LQSWFGKPSH	138038	B
KMGTPRHGSL	137978	B
PWPDVLHLSV	137391	B
KMDTHAMHGG	136747	B
WLRSRKKRIG	136665	B
DMGMGRMSKL	136282	B
GWWSGTSKAH	134937	B

RETAELEMWR	134732	B
RLTSEHEADG	134705	B
GTSMAIWTHM	134607	B
KHLISLAHAQ	134478	B
VRLAILGHKH	134115	B
LRWLFSRLEE	133820	B
HGSEYSRHKG	133515	B
KRWMSTRKEG	133397	B
EHWKTMKTPK	133281	B
AAAWQQQAHI	133166	B
GWWMSKSRQM	132865	B
WIYEWELTST	132607	B
ASRKKAQTWR	132590	B
MMRGWSYSGG	132529	B
SMKAASEHET	132405	B
RQPALMEHPY	132035	B
RRKMGQAIKW	131407	B
YSIHRGSMST	131369	B
GKKQTARMWK	130148	B
YIEGKLRRLG	130089	B
KHEERLTASS	129945	B
WQGHWHIRAS	129894	B
QKLGKTGTYM	129722	B
QEMHWEQKAA	129411	B
TGGSQTSGHH	129235	B
KMESMIRQGL	129102	B
VWSWWYWGIW	129102	B
HTKSQEMMEH	128828	B
MADRRWIKAH	128353	B
KMSGRNKLFH	128046	B
GPLRILGLAK	127971	B
EHHGREHKML	127953	B
RWSAKWSRSH	127903	B
GKQGHWPRTW	127856	B
QEHRMAERHN	127669	B
KRHTRDGGLE	126756	B
GHAGKPYPAL	126451	B
RRKMMQNHSS	126426	B
RASHWARSHM	126392	B
QTNHTGLNQE	126214	B
KSHHHRRLQL	126184	B

KVHHIGTNWE	125839	B
EEHQPAGRHQ	125177	B
VNEHRRRVEG	125122	B
EENAITWIWA	124939	B
NHRTQRHPHM	124901	B
QMTQRKQHSQ	124879	B
KHRHIKKGMA	124820	B
KVREIMWGGH	124723	B
ASDSHLGHIR	124653	B
QKWSEHHLSK	124544	B
HKGALHKVHH	124485	B
EEQESEISEQ	124150	B
APRSHIGHHA	124120	B
TNWHQWHTYR	123265	B
AKRWRDWSGF	122975	B
HKYMHGHHQM	122672	B
QTHVLWFQAH	122640	B
SLFIQWEETE	122330	B
TLLWMVWLWS	122212	B
SHQRGATAMH	121924	B
RTITKRWLWP	121548	B
MSPVVKSHQK	121387	B
EAWIWANWIS	121311	B
IEAMLKGWWY	121260	B
HQMSLEQEQK	121183	B
MGAEMHPPTS	120788	B
SSHLRESASH	120764	B
TISHTHRHQR	120437	B
QWWAFWWEWI	119313	B
WWIGFAMKKM	119305	B
MKMGHGPRPR	118932	B
SLHRRRAFST	118914	B
AQSQWMKKKW	118884	B
RQSAQIQKWR	118104	B
ALGGNEATDT	118052	B
RLWKRGGHKL	117851	B
ARWAHFIRKW	117423	B
LEIHPEPRRR	117245	B
KRQQQMTAIW	117116	B
AENGSLAYSS	116789	B
AHHHGGSEQT	116403	B

KRWKTGLDPK	116209	B
RMTERKHHRQ	115530	B
GKRRIYSSQH	115347	B
HGRHKRMPQI	115330	B
AYIHSLRLFI	115172	B
SKHGHEDHES	115024	B
WEAYALISLS	114833	B
TRRPYRWSMM	114591	B
SKWKAHQETK	114482	B
ESVTTHTSAF	114449	B
GHAGLMYMAL	114019	B
MIKLTKAIAE	114011	B
MWTVKLLSSR	113397	B
MGTILLTNSH	113364	B
HNWLQSHRGQ	112780	B
RHSEQGTDHM	112774	B
KHWMKVSPMK	112233	B
KKSGGVFHLR	112150	B
IHWKLVPKKL	111836	B
SHWRLIRRKS	111818	B
EIQSVENEAF	111800	B
RSTAHWSASQ	110623	B
KQRAGHDKMW	110582	B
RHSEVWTKHV	110486	B
VAWVKWLWQW	110249	B
KSIRWLSAGH	109658	B
LKDLSYKRDK	109572	B
KGAHHRASEA	109232	B
TKSRLPTHHR	109140	B
LGRVAMADLL	108885	B
HHKTTGWMHK	108868	B
SSRQYAWPDT	108825	B
HAHSKQWTVS	108541	B
KEAHHEWHGA	108476	B
EHGEHHYVEK	108469	B
MHKEIQGAGR	108035	B
GTHIHKKSLS	107361	B
IHFHHKASTL	107314	B
MTMRVLHISV	107241	B
HMWAHHGKTR	107209	B
RNPSHEAKMY	107173	B

KRHSQVIVHL	107171	B
YKPMVRSGYH	107128	B
GLTEWLSDRG	106573	B
QAHDGGHQRH	106236	B
GSIWQEAKTQ	106138	B
QGNDSLESTG	105745	B
AGIMLTKGME	105491	B
GHTQAIHHHK	104718	B
HHRGYQWSPG	104372	B
QQKDMASSQG	104244	B
KAQKTSHGKW	103299	B
HEKSMHWLRS	103105	B
LGKQEHATRL	103005	B
MIRLSAAEIR	102749	B
HTSKHSWHIQ	102541	B
KGKQGSAGGW	102516	B
IWRRKSTSAM	102499	B
ESSLPGGAEE	102492	B
QHTKSKLSML	102118	B
RLQTERAGWK	102104	B
RSAATKHLHT	102049	B
RGFERGTRHK	101739	B
STLHMNVVAH	101404	B
GEVGTKEHMH	101354	B
QNEQKLDVGF	101106	B
LHGTLHRMSR	100939	B
IAKSHTRGHQ	100584	B
KMTGGVSHKR	100417	B
HSKRKKHMAM	99885	B
TSTQPMTQTH	99492	B
TTHETVRLGT	99274	B
KGLHSGSRNN	99112	B
SSIHKRYVAH	99018	B
TIMELEPYLW	98639	B
HEHQTLHKYM	98559	B
MQQWHGEKGR	98290	B
YHKHTELRIH	98231	B
KSMSREKYTA	98041	B
LWYLSSRRKA	97680	B
ERNHQKSQKH	97448	B
AYIHTKRKSL	96745	B

ATRKKRETWQ	96534	B
VHHNHGRWHK	95234	B
GHAHMSYTAK	95003	B
ESGHHHSESK	94961	B
ETHHGAMKRS	94723	B
HQHIERAMAN	94711	B
LTSHVHGGTK	93935	B
KQQRAWKHHG	93832	B
RSQWWSWALG	93484	B
RRQNKIRVGM	93333	B
TMHVHRHEAN	93175	B
TGGSETSGHH	92867	B
SMHHSFQQAK	92611	B
AEAKSSIKHH	91967	B
KTEHSMQGGM	91909	B
WSRHHHHAKH	91800	B
IATSLAEIHK	90868	B
TKWKIYIKKW	90790	B
TPHEHQKELE	90061	B
HVHSHEENGT	89965	B
EATSQMWSER	89896	B
VSSDIEVTEN	89893	B
RSMSLHRVLI	89844	B
SWHGKQTKSG	89658	B
VHEYMERKNS	89330	B
RWRMYTKSRA	89016	B
SETISQGMRA	88774	B
TQMRFHIHKH	88766	B
RKQLWRVIIL	88758	B
GWWAATLSLR	88716	B
TKKGESSSWK	88352	B
LQAGHDSGGA	88269	B
TKIQRQASSW	87763	B
ASSGHHKETR	87216	B
KSMVVKSHRK	86487	B
PKKAAPQHQT	86438	B
SSEEMHHKGV	86227	B
VKFGYGHGMV	86225	B
LGSHHHSRKH	85875	B
NKLMKRVRMF	85549	B
SRHHHHHERR	85535	B

IWFMTAMKKL	85378	B
LVKHRWWGSQ	84742	B
EIQYSKLKLT	84454	B
KSKGMARPHG	84408	B
YSAKRKLQHR	83970	B
TIHRLSSGMM	83813	B
HTKAEIHEKK	83731	B
EHGQHHYVQK	82804	B
GWIEQKGMRK	81675	B
LAQQHGKGSE	81338	B
SLMIDWEQSQ	80445	B
VQATHDAASG	80287	B
HDKPIQRQRL	80088	B
RGAMGKKKEA	79544	B
KGSHIISRMH	79060	B
LEAQLGKKFK	78865	B
RSKAEWSKTA	78791	B
QMKSKDVRPS	78724	B
VAIAHLEGKH	78493	B
KRFAHTRSSK	78258	B
SMEEPEKTKI	78025	B
TQIIGPHSKM	77985	B
AHVGPTEGTV	77700	B
TSDWSGHGAQ	77682	B
YGHTFTRQKA	77650	B
HTPAAHTTWA	77560	B
IEIHMEMARR	77462	B
EISSRKWKEG	76915	B
MRLARSERSH	76900	B
IEKETHEEGK	76389	B
IHWLIVMKLI	76224	B
LRHSRRGGIQ	75749	B
AKHRSATITQ	75572	B
RESALPGHSY	75416	B
KHQQRISAPP	75324	B
LEVRKETGAQ	75048	B
STAHMTLGSL	74911	B
KHRVASAQKM	74820	B
LSYSIHIIET	74519	B
GRSLSKHGSH	74339	B
WRQRKLQNRS	73683	B

SHKEHQKYIS	73648	B
RGKEHVVNDG	73634	B
ESHHGAKKRF	73398	B
EGPHTRGKAM	73016	B
GTTKTHHAEG	72848	B
RRMHSHKTAV	72196	B
SIITKVWITF	72055	B
MTQSTGHNYK	72023	B
TSAQHLRQAL	71327	B
WSYEASFRGA	71126	B
YLSKPSWERA	70607	B
ASSGHHKQTR	70564	B
MLRHRTRIKK	70092	B
GAKYHTSARH	69887	B
EQHLEGKETM	69187	B
AEAMTSLKHH	69186	B
GKMSALHKQK	67553	B
AGHTRIGNVA	67373	B
NRPHSHMTRV	67283	B
RQHIWARTRS	66799	B
KSSSWHRPLK	66593	B
TMSETNKTAR	66569	B
QEGLQRDQAH	66469	B
DGPEHVVQQH	66453	B
TNLETGSGGL	66143	B
QSARTKHKHT	65840	B
SGIKTQYEIG	65547	B
KIHPGEQYVT	65249	B
EQESQYRQAG	65129	B
IHGEPGIRAH	64980	B
RMSAGTPAMQ	64974	B
HTLTGEERQK	64805	B
GWQGSYMGKS	64421	B
DTSTLHQVKI	64238	B
RISELSFTSR	64144	B
YKGHFKQRMH	64064	B
GSEIRMHMRS	63845	B
SFSHLGTQTR	63769	B
WIKTKQKRGK	63743	B
EQTATWGIAS	63208	B
RAWHAAYARR	62751	B

SHKDASVHSQ	62067	B
RKEWRYAFKS	61006	B
MARKQATYGR	60982	B
AGLSETAAET	60844	B
EHGESVSKKL	60657	B
RQGIRRRRAH	60570	B
AHHHGGSQRT	60507	B
REGMGPLGET	60106	B
EREQLGSTSK	59849	B
VMGTGSGKHA	59161	B
HSGRMHSVHH	59005	B
QSTEQMIIRE	58502	B
TETLSEGPQA	58417	B
VGYEKRHRSE	57761	B
KSPEAMLGRG	57305	B
SQASPMHQGH	57189	B
HGLYKKTSSR	56337	B
GAGWQSQATL	55931	B
SKGVGRHEAR	55843	B
EWHGTTGRKS	55249	B
KQFHKSALIM	54749	B
TISGKEHRHQ	53998	B
QKRTHASKEA	53436	B
KYTDMWKPTK	53329	B
GHWNQIGRGV	52462	B
PLQKSAAELD	52388	B
GMIRILGLAL	52312	B
YSIHRGSKSS	52161	B
HHTKKHHKTRG	51932	B
VQEINRNVEH	51535	B
ERHLQGLQSK	50169	B
SSHKQESASH	50042	B
PTWERIVSTQ	49611	B
GEKAIITGLQ	49209	B
HVHSHQQRGT	49073	B
KPKPRQKYSA	48783	B
LSMQAKIGAG	48208	B
QGMHTRGLAK	48191	B
AKMTVRMKQK	46944	B
KLKQTHIPGE	46939	B
EHGETVSKKK	46849	B

RHHSLSTPKH	46744	B
RQTAHHEARH	45148	B
RDHHMQLSGR	45054	B
GTELRPHPDS	44889	B
RGLEMEDKAK	44839	B
MKKTTTGLPS	44403	B
RSHKARLKLQ	44347	B
VHTRSYIIVK	44267	B
KHGIKMHLSK	43057	B
QTSILAHAHK	42971	B
TMSHSTGRAI	42595	B
PRKADTEQSH	42489	B
EQEEPKQVGS	41584	B
TSEAGAGAGT	41570	B
EQAERTYMKL	41053	B
GWSTRALKKA	40868	B
HRGHQRAKAR	40574	B
AKKHHKKRIR	40464	B
EEHKENQEAH	40460	B
SKQQMQLSKH	40453	B
ERQHEKTEKI	40436	B
GQLAHHSGLR	39818	B
TKSEKFHTKL	38451	B
SSTGQESMTH	38250	B
KQRMRQGTLA	37948	B
VRQHRARVQG	37589	B
KKSEMEKDTR	37253	B
QWSPRARTGR	36961	B
LATSKAGLHK	36622	B
SSSHGKKTAE	36586	B
PKTGPSDGHS	36473	B
HSISGQQRRK	35946	B
VRKALMGHPI	35705	B
VHQYKERKRS	35705	B
ATQTHKHHLR	35669	B
PPRGGHKERR	34723	B
KKGSMRHGSL	34377	B
ENQIRLSKGE	34275	B
TQATSSHEHI	33540	B
GGGRWRAGPA	32970	B
SGGQKKQESE	31824	B

RSSHHSSQKE	31763	B
AHVGKSEGSV	31319	B
GAEEEKQHVG	31033	B
KKSHDTTSRR	30555	B
QHATHPSMGR	29593	B
RSLTDKLSSD	29406	B
WKQQSTSSKA	29027	B
RRTKHRHQSS	28853	B
VSAKRKKEHN	28607	B
NQHIPEKTGQ	28475	B
MVDRKTGARA	28165	B
RDEQSKHSKT	26618	B
ERKRMIITAK	26347	B
SWWPTVRKSK	25775	B
HVRKTEASKT	25538	B
RGRRSIQFTG	25486	B
RSRGMEHRRE	25444	B
GDQEKQRQKQ	25233	B
GNSKSMIHSH	24751	B
IHDISRSTEK	24447	B
RGLSKASLID	24414	B
HHKSSGWKHK	24377	B
TISWRIDVAQ	24184	B
SKKTETVHQH	23477	B
KSHTEKKRAQ	23192	B
ETSTGTSARG	22957	B
QPTGKSKVLS	22813	B
QTTGIHEEKM	21958	B
SRLESGSGGL	21783	B
QRSGQLARAE	21779	B
LLRISYLRRK	21440	B
GKWARMIRRP	20449	B
LPHSQKLAAR	20434	B
RGEKETAHAT	19366	B
HYAPAINGKG	19332	B
RGVRSRTVKM	18967	B
DSHKANLKKE	18952	B
RASMATESTR	18807	B
DSRGPEIRDE	18661	B
EAQEELRGVG	17861	B
NVKASTQVGR	17746	B

FKPGIGSNSR	17696	B
YQAVKPFSRM	17585	B
HGSRRPTQKG	17495	B
KHGLKPHKTK	17322	B
RTSDKAHSLG	16919	B
VKSGVGHGPV	16670	B
PTTPQTEAGE	16213	B
RKGKEAKSLL	16054	B
ERSGELANAG	15939	B
MSSTNRHKTT	15553	B
TKGSIASLRL	15444	B
QSGHHHSQSK	14498	B
KLSENSEMHK	14464	B
SLGSIAFIAI	13512	B
GEEEKENEKE	12772	B
TQWASKMKQG	12697	B
RSGASRITML	12563	B
QSGGRWQQTL	11392	B
RRYHKGTNTK	10993	B
IANDGGKGSQ	9941	B
QDRQMQHKRR	9774	B
KTLMLWFLRG	9660	B
KSMGHWYARK	8756	B
TIGKKSGKEH	6928	B
ALRRRWQLKR	6033	B
KVTKTIIASE	5067	B
SHATGNTQLK	3897	B
QAKRMIITAL	3101	B
STWTIKTKGT	3101	B
KLFQKQLRSR	3029	B
AMARTKKLRY	2271	B
RSMGSMFSRG	2179	B
KVSLSIIASQ	2009	B
SGRESTSTKS	1997	B
TKEKEMSGKR	1894	B
QGAPGKMKER	1561	B
VGGGRKGAVG	1409	B
AKKTVRKLRK	1258	B
GSTTQGNLQG	436	B
TAAGNANKNS	203	B

Table S2. External Validation Dataset

Sequence	Class
WAGAKRLVLRRE	A
MHGKTQATSGTIQS	A
LKAHLPPSRLPS	A
WALRRSIRRQSY	A
TGTSVLIATPYV	A
EQLGVRKELRGV	A
RMRMKMK	A
PPPWLPYMPPWS	A
AYSSGAPPMPPF	A
TGIFKSARAMRN	A
KHKHWHW	A
TSNAVHPTLRHL	A
RLNPPSQMDPPF	A
QTWPPPLWFSTS	A
GCGLFHAIAHFIHGGWHGLIHGWYG	A
KHKHWHWG	A
HHHHHHH	B
MHMHMHM	B
RHRHRHR	B
YHYHYHY	B
WHWHWHW	B
KHKHKHK	B
AHAHAHA	B
GHGHGHG	B
QHQHQHQ	B
IHIHIHI	B
NHNHNHN	B
VHVHVHV	B
SHSHSHS	B
THTHTHT	B
EHEHEHE	B
DHDHDHD	B
LHLHLHL	B
FHFHFHF	B
PHPHPHP	B

MEGQYKSNLLFT	B
VAPRVQNLHFGA	B