

Citation: Xu Y, Yu S, Zou J-W, Hu G, Rahman NABD, Othman RB, et al. (2015) Identification of Peptide Inhibitors of Enveloped Viruses Using Support Vector Machine. PLoS ONE 10(11): e0144171. doi:10.1371/journal.pone.0144171

Editor: Massimiliano Galdiero, Second University of Naples, ITALY

Received: April 24, 2015

Accepted: November 13, 2015

Published: December 4, 2015

Copyright: © 2015 Xu et al. This is an open access article distributed under the terms of the <u>Creative</u> <u>Commons Attribution License</u>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: These authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

RESEARCH ARTICLE

Identification of Peptide Inhibitors of Enveloped Viruses Using Support Vector Machine

Yongtao Xu^{1,2}, Shui Yu¹, Jian-Wei Zou³, Guixiang Hu³, Noorsaadah A. B. D. Rahman^{4,5}, Rozana Binti Othman^{5,6}, Xia Tao⁷, Meilan Huang¹*

1 School of Chemistry and Chemical Engineering, Queen's University Belfast, David Keir Building, Stranmillis Road, Belfast, Northern Ireland, United Kingdom, 2 School of Basic Medical Sciences, Xinxiang Medical University, Xinxiang, Henan, China, 3 School of Biotechnology and Chemical Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo, China, 4 Department of Chemistry, Faculty of Sciences, University of Malaya, Kuala Lumpur, Malaysia, 5 Drug Design & Development Research Group, University of Malaya, Kuala Lumpur, Malaysia, 6 Department of Pharmacy, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, 7 State Key Laboratory of Organic-Inorganic Composites, Beijing University of Chemical Technology, Beijing, China

* m.huang@qub.ac.uk

Abstract

The peptides derived from envelope proteins have been shown to inhibit the protein-protein interactions in the virus membrane fusion process and thus have a great potential to be developed into effective antiviral therapies. There are three types of envelope proteins each exhibiting distinct structure folds. Although the exact fusion mechanism remains elusive, it was suggested that the three classes of viral fusion proteins share a similar mechanism of membrane fusion. The common mechanism of action makes it possible to correlate the properties of self-derived peptide inhibitors with their activities. Here we developed a support vector machine model using sequence-based statistical scores of self-derived peptide inhibitors as input features to correlate with their activities. The model displayed 92% prediction accuracy with the Matthew's correlation coefficient of 0.84, obviously superior to those using physicochemical properties and amino acid decomposition as input. The predictive support vector machine model for self-derived peptides of envelope proteins would be useful in development of antiviral peptide inhibitors targeting the virus fusion process.

Introduction

Fusion process is the initial step of viral infection, therefore targeting the fusion process represents a promising strategy in design of antiviral therapy [1]. The entry step involves fusion of the viral and the cellular receptor membranes, which is mediated by the viral envelope (E) proteins. There are three classes of envelope proteins [2]: Class I E proteins include influenza virus (IFV) hemagglutinin and retrovirus Human Immunodeficiency Virus 1 (HIV-1) gp41; Class II E proteins include a number of important human flavivirus pathogens such as Dengue virus (DENV), Japanese encephalitis virus (JEV), Yellow fever virus (YFV), West Nile virus (WNV), hepatitis C virus (HCV) and Togaviridae virus such as alphavirus Semliki Forest virus (SFV); Class III E proteins include vesicular stomatitis virus (VSV), Herpes Simplex virus-1 (HSV-1) and Human cytomegalovirus (HCMV). Although the exact fusion mechanism remains elusive and the three classes of viral fusion proteins exhibit distinct structural folds, they may share a similar mechanism of membrane fusion [3].

A peptide derived from a protein-protein interface would inhibit the formation of that interface by mimicking the interactions with its partner proteins, and therefore may serve as a promising lead in drug discovery [4]. Enfuvirtide (T20), a peptide that mimicks the HR2 region of Class I HIV-1 gp41, is the first FDA-approved HIV-1 fusion drug that inhibits the entry process of virus infection [5–7]. Then peptides mimicking extended regions of the HIV-1 gp41 were also demonstrated as effective entry inhibitors [8, 9]. Furthermore, peptides derived from a distinct region of GB virus C E2 protein were found to interfere with the very early events of the HIV-1 replication cycle [10]. Other successful examples of Class I peptide inhibitors include peptide inhibitors derived from SARS-CoV spike glycoprotein [11–13] and from Pichinde virus (PICV) envelope protein [14]. Recently, a peptide derived from the fusion initiation region of the glycoprotein hemagglutinin (HA) in IFV, Flufirvitide-3 (FF-3) has progressed into clinical trial [15].

The success of developing the Class I peptide inhibitors into clinical use has triggered the interests in the design of inhibitors of the Class II and Class III E proteins. e.g. several hydrophobic peptides derived from the Class II DENV and WNV E proteins exhibited potent inhibitory activities $[\underline{16}-\underline{20}]$. In addition, a potent peptide inhibitor derived from the domain III of JEV glycoprotein and a peptide inhibitor derived from the stem region of Rift Valley fever virus (RVFV) glycoprotein were reported $[\underline{21}, \underline{22}]$. Examples of the Class II peptide inhibitors of enveloped virus also include those derived from HCV E2 protein $[\underline{23}, \underline{24}]$ and from Claudin-1, a critical host factor in HCV entry $[\underline{25}]$. Moreover, peptides derived from the Class III HSV-1 gB also exhibited antiviral activities $[\underline{26}-\underline{31}]$, as well as those derived from HCMV gB $[\underline{32}]$.

Computational informatics plays an important role in predicting the activities of the peptides generated from combinatorial libraries. In silico methods such as data mining, generic algorithm and vector-like analysis were reported to predict the antimicrobial activities of peptides [33-35]. In addition, quantitative structure-activity relationships (QSAR) [36-40] and artificial neural networks (ANN) were applied to predict the activities of peptides [41, 42]. Recently, a support vector machine (SVM) algorithm was employed to predict the antivirus activities using the physicochemical properties of general antiviral peptides [43]. However, the mechanism of action of antiviral peptides is different from antimicrobial peptides; in fact, various protein targets are involved in the virus infection. e.g. HIV-1 virus infection involves virus fusion, integration, reverse transcription and maturation, etc. Thus it is difficult to retrieve the common features from general antiviral peptides to represent their antiviral activities. Virus fusion is mediated by E proteins. Although E proteins are highly divergent in sequence and structure, they share a common pathway of membrane fusion dynamics. i.e. E proteins experience significant conformational change to form a-trimer-of-hairpin, which drives the fusion of viral membrane and host membrane [44]. The antiviral peptides derived from enveloped proteins function by *in situ* binding to their respective accessory proteins, disrupting forming of the trimer-of-hairpin and membrane fusion, and therefore inhibiting the virus infection. In view of the important role of E proteins in virus fusion process and common mechanism of action of self-derived peptides, we developed a SVM model to predict the antiviral activities of self-derived peptides using sequence-based statistical scores as input features. The sequencebased properties were calculated by a conditional probability discriminatory function which indicates the propensity of each amino acid for being active at a specific position. Our model exhibited remarkably higher accuracy in predicting the activities of self-derived peptides, compared to the previous models developed for general antiviral peptides using classical

physicochemical properties as descriptors [43]. The method would be useful in identification of entry inhibitors as a new generation of antiviral therapies.

Methods

Data collection

202 peptide virus entry inhibitors of enveloped viruses were collected, among them, 101 are active peptides and 101 are non-active peptides. These peptides comprised the 75p+75n training set of SVM models. The remaining 26 active peptides and 26 non-active peptides inhibitors were used as the test set.

Amino acid composition. Amino acid composition is the fraction of each amino acid in a peptide. The fraction of the 20 amino acids was calculated using the following equation:

Fraction of amino acid X = Total number of X / peptide length

Physicochemical properties

Five physicochemical properties were used in SVM models. Isoelectric point (PI), Molecular weight (MW) and Grand average of hydropathicity (GRAVY) [45] were calculated using the Protparam tool implemented in Expasy web server. Solvent accessibility and secondary structure features were calculated using SSpro and ACCpro packages implemented in the SCRATCH protein predictor server [46].

Sequence-based statistical scoring function. The knowledge-based statistical function is developed from the concept of residue-specific all-atom probability discriminatory function (RAPDF) [47]. RAPDF is a structure-based statistical scoring function. It is based on the assumption that averaging over different atom types in experimental conformations is an adequate representation of the random arrangements of these atom types in any compact conformation. Here we developed a sequence-based statistical scoring function, where we presume that averaging over different amino acid sequences with experimental validated inhibitive activities is an adequate representation of the random amino acid sequences with any inhibitory activity. The basis of this assumption is that the peptides share a common mechanism of action, i.e. the peptides derived from E proteins bind competitively to their partner proteins, disrupt the forming of a-trimer-of-hairpin, and therefore inhibit the virus membrane fusion.

The sequence-based scoring function is described in the following form:

$$S(\lbrace q_a^i \rbrace) = -\ln \frac{P(q_a^i | C)}{P(q_a^i)}$$

$$\tag{1}$$

Here, $q_a^i \in \{active\}$.

 $P(q_a^i|C)$ is the probability of observing amino acid *i* in an active peptide sequence;

 $P(q_a^i)$ is the probability of observing amino acid *i* in any peptide sequence, active or nonactive. They are approximately estimated using the following forms:

$$P(q_a^i|C) \equiv \frac{N_{obs}(i,a)}{N_{obs}(a)} \tag{2}$$

$$P(q_a^i) \equiv \frac{N_{obs}(i)}{N_{total}} \tag{3}$$

 $N_{obs}(i,a)$: The number of observed amino acid *i* within active peptides.

 $N_{obs}(i)$: The number of observed amino acid *i* within active peptides and non-active peptides.

 $N_{obs}(a)$: The number of observed amino acid types within active peptides.

 N_{total} . The number of observed amino acid types within active peptides and non-active peptides.

Similarly, we employed a dataset of experimentally verified non-active peptides in developing the statistical function, where $q_a^i \in \{inactive\}$.

For a given amino acid sequence, 20 columns of input are generated, corresponding to the occurrence of twenty natural amino acids at each position. Each column is assigned a value of N^* (-log-*likelihood*), where N is the number of amino acid and -log-*likelihood* is derived from the statistical function score. Each of the features thus combines the propensity of the amino acid for being active or non-active with the corresponding amino acid composition.

Below is an example of calculating the statistical scores for a given peptide sequence: The amino acid order for SVM input features is set as:

ACDEFGHIKLMNPQRSTVWY.

If the amino acid sequence of an active peptide inhibitor is:

DCPNGPWVWVPAFCQAVGWG,

the statistical N values of the sequence would be:

2,2,1,0,1,3,0,0,0,0,0,1,3, 1,0,0,0,3,3,0

The scores in the statistical function library based on the active peptide inhibitors are decided by Eq (1): -0.0856, 0.5057, 0.4740, 0.4133, -0.0856, -0.0856, 0.6439, 0.2508, 0.9440, -0.4670, 1.8603, 0.1330, 0.2261, -0.0115, 0.2761, 0.3288, 0.0479, -0.1207, 0.0079, 0.6816,

Therefore, the 20 SVM input features for the sequence would be: -0.1712, 1.0114, 0.4740, 0, -0.2568, -0.2568, 0, 0, 0, 0, 0, 0.1330, 0.6783, -0.0115, 0, 0, 0, -.3621, 0.0237, 0.

SVM Parameter Optimization

SVM models combined with radial basis function (RBF) kernel parameters were developed using the C-SVC module in LIBSVM (version 3.1) [48, 49] and executed under the Matlab interface. The performance of SVM depends on two parameters, gamma -g and $\cot -c$ [50]. The default value is 1 for -c and 1/k for -g, where k is the number of input entries. Various pairs of (*c*, *g*) values were converted to exponential values (i.e. $2^x; 2^y$) and optimized using cross-validation and the pair with the best cross-validation accuracy was selected.

5-fold cross validation was performed to evaluate the performance of SVM models. In the evaluation process, dataset was partitioned randomly into five equally sized subsets. The training and testing were carried out five times, each time four distinct subsets being used as training sets and the remaining subset as test set. The results were averaged over all five rounds of validation. The following equations were used to evaluate the prediction quality of the SVM models [48, 51]:

$$Sensitivity = \left[\frac{TP}{TP+FN} \right] * 100$$

$$Specificity = \left[\frac{TN}{TN+FP} \right] * 100$$

$$Accuracy = \left[\frac{(TP+TN)}{TP+TN+FP+FN} \right] * 100$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP+FP) * (TP+FN) * (TN+FP) * (TN+FN)}}$$

In the above equations, TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives. Matthew's correlation coefficient (MCC) reflects the performance of the model. It ranges between -1 to 1 and a larger MCC value indicates a better prediction.

Results and Discussion

SVM learning algorithm is a powerful machine learning method that has been widely used in pattern recognition and classification. SVM trains a dataset of experimentally validated positive and negative samples and generates a classifier to classify unknown samples into two distinct categories (positive or negative).

Collection of dataset

We performed an exhaustive literature search on self-derived peptide inhibitors of enveloped proteins and collected experimentally validated peptides derived from the three classes of E proteins. For those peptides with overlapping segments, only one peptide sequence was kept. 202 peptides were found, among them, 101 are active peptides and 101 are non-active peptides (<u>Table 1</u>). 75 active peptide inhibitors and 75 non-active peptides (75p+75n) of E proteins were used as the training dataset in SVM learning; the remaining 26 active and 26 non-active peptides (26p+26n) were used as the test set.

SVM input features. Three SVM models were developed using different features as input descriptors, namely physicochemical properties (denoted as EAPphysico), amino acid composition (EAPcompo) and statistical scoring function amino acid composition (EAPscoring).

Knowledge-based statistical functions are rooted in the Bayesian (conditional) probability formalism and derived directly from properties observed in the known folded proteins [52–54]. In knowledge-based scoring function, it was presumed that averaging over different atom types in experimental conformations is an adequate representation of the random arrangements of these atom types in any compact conformation [55]. Because the three classes of E proteins have different structural folds, it is difficult to retrieve a structure-based feature that is relevant to their antiviral activities. Generally speaking, any property associated with folded proteins can be converted into an energy function [56]. Since amino acid sequence determines the structural folds and properties of proteins/peptides, we presumed that a sequence-based statistical scoring function averaging over different amino acid sequences exhibiting inhibitive activities is an adequate representation of the random combinations of all twenty amino acid exhibiting any activity. In this approach, a peptide sequence derived from E protein is represented by twenty features each corresponding to the propensity of observing each of the twenty natural amino acids to be either active or non-active. A vector space of twenty sequence-based statistical scores was used as the EAPscoring input entries in the SVM learning.

We also built a SVM model using physicochemical properties as input features. Because of the feature of membrane fusion process, it was suggested that functional regions in glycoproteins need to be solvent accessible, hydrophobic and flexible [57]. Actually the majority of known peptide entry inhibitors share a common physicochemical property of being hydrophobic and amphipathic with a propensity for binding to lipid membranes [58]. Therefore, here the properties of E peptide inhibitors were described by five physicochemical parameters: PI, MW, GRAVY index (positive and negative GRAVY values indicate hydrophobic and hydrophilic peptides, respectively), solvent accessibility (exposed or buried) and secondary structure features (propensity for adopting α -helix, β -sheet or turn structure). These physicochemical features in

Table 1. Experimentally validated peptide inhibitors of E proteins.

	Active peptides	Non-active peptides	Ref
ніх	DCPNGPWVWVPAFCQAVGWG	SPLGFGSYTMTKIRDSLHLV	[9]
16p+31n	GPWVWVPAFCQAVGWGDPIT	ANGSRIPTGERVWDRGNVTL	
	LCDCPNGPWVWVPAFCQAVG	CGTCVRDCWPETGSVRFPFH	
	PNGPWVWVPAFCQAVGWGDP	CRANGSRIPTGERVWDRGNV	
	PTGERVWDRGNVTLLCDCPN	CSCRANGSRIPTGERVWDRG	
	RGNVTLLCDCPNGPWVWVPA	DLEAVPFVNRTTPFTIRGPL	
	RIPTGERVWDRGNVTLLCDC	ELSEWGVPCVTCILDRRPAS	
	TLLCDCPNGPWVWVPAFCQA	ETGSVRFPFHRCGTGPRLTK	
	WDRGNVTLLCDCPNGPWVWV	GAPASVLGSRPFDYGLKWQS	
	WVWVPAFCQAVGWGDPITHW	GLTGGFYEPLVRRCSELMGR	
		GNQGRGNPVRSPLGFGSYTM	
		GSRIPTGERVWDRGNVTLLC	
		HWSHGQNQWPLSCPQYVYGS	
		KCPTPAIEPPTGTFGFFPGV	
		LGSSDRDTVVELSEWGVPCV	
		LSCPQYVYGSVSVTCVWGSV	
		PFDYGLKWQSCSCRANGSRI	
		PPINNCMPLGTEVSEALGGA	
		QAVGWGDPITHWSHGQNQWP	
		RCGTGPRLTKDLEAVPFVNR	
		SKIDVWSLVPVGSASCTIAA	
		SPLGFGSYTMTKIRDSLHLV	
		SWFASTGGRDSKIDVWSLVP	
		TCILDRRPASCGTCVRDCWP	
		TEVSEALGGAGLTGGFYEPL	
		TGTFGFFPGVPPINNCMPLG	
		TKIRDSLHLVKCPTPAIEPP	
		TTPFTIRGPLGNQGRGNPVR	
		VGSASCTIAALGSSDRDTVV	
		VRRCSELMGRRNPVCPGYAW	
		VSVTCVWGSVSWFASTGGRD	
	LSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQ		[10]
	SGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARIL		
	NNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQ		
	YTSLIHSLIEESQNQQEKNEQELLELD		
	WMEWDREINNYTSLIHSLIEESQNQQEKNEQELL		
	YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWF		[8]
DENV		RWRHLKKMQRLQPRNPNWPGQFWVHYNW	[17]
5p+9n	FWFTLIKTQAKQPARYRRFC	MVIVQHQWMQIMRWPWQPE	
	RQMRAWGQDYQHGGMGYSC	QQCFRFPALRKKATYTRFWI	
		YPENLEYRVYIIPHPGEEHH	
		WHIVEPIVIEKDRPVNYEWE	14.01
			[19]
	MAILGDTAWDFGSLGGVFTSIGKALHQVFGATY		18
		AWEVHIQWELDEFEFWEFGADIQGSNWI	[00]
DEINV-DEI			20
2n+0n			
WNV			[16]
7n+13n			[10]
7 P + 1311		TGPEEDGRDTRD	[4 0]
	CDVIALLCHINT	NTTHYRVIRI TIG	[10]
	CDVIALLACHINT	DTBACDVIALI	

(Continued)



Table 1. (Continued)

GUVALLECHUNTESE TRACDVALLECHUNT DTRACDVALLECHUNT DTRACDVALLECHUNT DTRACDVALLECHUNT GUVALLECHUNT DTRACDVALLECHUNT GUVALLECHUNT GUVALLECHUNT GUVALLECHUNT HEAD GUVALLECHUNT GUVALLECHUNT GUVALLECHUNT HEAD		Active peptides	Non-active peptides	Ref
 DTRACEVALLECHLAT DTRACEVALLE DTRACEVALE DTRAC		CDVIALLCHLNTPSF	TRACDVIALLECHLNT	
DTRACDVPLL CDVALL CDVALL DTRACUMPLE CDVALLCOLVITSF CDVALLCOLVITSF CDVALLCOLVITSF CDVALLCOLVITSF HCMV WEIHHINKFADAYSYSRVIGGTVFVA [2] HFFATSTCOVVVERTANLAMLITTARSKYPY [2] HFFATSTCOVVVERTANLAMLITTARSKYPY CPPTDATWOREGP HFFATSTCOVVVERTANLAMLITTARSKYPY CPPTDATWOREGP HFFATSTCOVVVERTANLAMLITTARSKYPY CPPTDATWOREGP HSV KTSSEFARLOFTY HSV CPPTDATWOREGP HSV CPPTDATWOREGP HSV CPPTDATWOREGP HEVPLEVTTRHEK DARDAMBRIFARAYN DUKLPANAATTTSR DEVENKARGVCRST EVENKARGVCRST EVENKARGVCRST EVENKARGVCRST EVENKARGVCRST EVENKARGVCRST EVENKARGVCRST EVENKARGVCRST EVENKARGVCRST HSV-gH AAHLDALVAEFLOGRVLTTPVVHRALFYASAVLROPFLAGVPSA HSV-ghT APSIPSTOVALLEPSTOVALARGAN APSIPSTOVALST COVENTIONARGAN HSV-ghT APSIPSTOVALLEPSTOVALLEPSTOVALLEPSTOVALAGON Bg-zh0 GLASTLTFWARTWILLEPSTOVALLEPSTOVALLEPSTOVALLEPSTOVALAGON APSIPSTOVERTALLEPSTOVALLEPSTOVALLEPSTOVALLEPSTOVALLEPSTOVALAGON CPPTEGAPSVFSSOVERSTALLEPSTOVALLEPSTOVALLEPSTOVALLEPSTOVALAGON HSV-ghT AAHLDALVAEFLOGRVLTPVVHALLEPSTOPVAAA CPPTGAPSVFSSOVES			DTRACDVIALLECHLNT	
<pre>CDVALL OTATALA OTATALA OTATALA OTATALA OTATALA CDVALLACENTIN CDVALLACENTIN CDVALLACENTIN CDVALLACENTIN CDVALLACENTIN CDVALLACENTIN DTRACDVALLECHTI OTACDVALLECHTI OTAC</pre>			DTRACDVIPLL	
DTAPLA CDVALECHNTSP CDVALECHNTSP CDVALECHNTSP CDVALECHNTSP CDVALECHNTSP HCMV WEIRDISTWUTRETANLAMUTTTARSKYPY [2] HFFATSTOUVYDEPTNGTMRNASYG [3] HFFATSTOUVYDEPTNGTMRNASYG CYSRUVGEGP [3] HFFATSTOUVYDEPTNGTMRNASYG CYSRUVGEGP [3] HSV KTSSIEFARLCFTY CYSRUVGEGP [3] HSV KTSSIEFARLGFTY CYSRUVGEGP [3] HEVPLEYTRHEK DARDAMORFARYN [3] DURTARDAMART COMMELCANADAMORF [3] HEVPLEYTRHEK DARDAMORFARYN [3] DURTARDAMART COMMELCANADAMART [3] HEVDLEVTRHEKSGL [3] [3] HEVDLEVTRHEKSGL COMMELCANADAMART [3] DPREMIKKKYNATTSP COMELCANADAMART [3] DPREMIKKSCHARADAMART [3] [3] SUMMELCANADAMART [3] [3] MSV-pH AMELDALVERT [3] [3] SUMMELCANADAMART [3] [3] [3]			CDVIALL	
 COVALLACEUMTSEG COVALLACEUMT COVALLACEUMT COVALLACEUMT COVALLECHUNT COVALLECHUNT COVALLECHUNT COVALLECHUNT PERPENDINGTRADAYSSYSTWOGTVEYA PERPENDINGTRADAYSSYSTWOGTVEYA PERPENDINGTRADAYSSYSTWOGTVEYA PERPENDINGTRADAYSSYSTWOGTVEYA PERPENDINGTRADAYSSYSTWOGTVEYA COVARDAYSTRADAY COVARDAYSTRADAYSTRADAY COVARDAYSTRADA			DTRAPLAI	
DOWALLEGHANT DOWALLEGHANT INT COMMUNICADAYSYSRWGCTYFA INTERACUMALLEGHANT INTERACUM			CDVIALLACHLNTPSF	
HANV WEHHINKFAGAYSYSRWGGTYFVA (2) HIPPATSTGOVYSPANALTITTARSKYPY (2) HIPPATSTGOVYSPYGTONALTITTARSKYPY (2) HIPPATSTGOVYSPYGTONALTITTARSKYPY (2) HIPPATSTGOVYSPYGTONALTITTARSKYPY (2) HIPPATSTGOVYSPYGTONALTITTARSKYPY (2) HIPPATSTGOVYSPYGTONALTITTARSKYPY (2) HIPPATSTGOVYSPYGTONALTITTARSKYPY (2) HIPPATSTGOVYSPYGTONALTITTARSKYPY (2) HIPPATSTGOVYSPYGTONALTITARSKYPY (2) HIPPATSTGOVYSPYGTONALTITTARSKYPY (2) HIPPATSTGOVYSPYGTONALTITTARSKYPY (2) HIPPATSTGOVYSPYGTONALTITALSKYPY (2) HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPYGTONALTITALSKYPY HIPPATSTGOVYSPY HIPPATSTGOVY			CDVIALLECHLNT	
HOM WEIHINGFADAYSYSTRYOGTYFVA (29) WESGEVYA FEANLAMAUTTRABKYY V HEFASTSGOVVUSPFYAGTNINASYFG HEFADTSGOVVUSPFYAGTNINASYFG HEFADTSGOVVUSPFYAGTNINASYFG HEVELEVYTRHEIK HSV KTSGOVVERGPNAA HSV KTSGOVEFYAGGYVF CYSRLVSFYGGOTFAG HEVELEVYTRHEIK OMERVFANDAMORFARYN HSV TTSFLAUTMENK OMERVFANDAMORFARYN HSV HTSFLAUTMENK DOHETMENKA DO			DTRACDVIALLECHLNT	
 dp-0n WHSRGSTWL/RETAILLABULTIT/ARSKYPY HFATSTGOV/SPFWGTNABURSYFG FFIFAVTTWSDFGGPNAA KTTSSIEFALGOFY CYRPTGAT/VOFEOP GHRDYFFGGGYVYF CYRPTGAT/VOFEOP GHRDYFFGGGYVYF CYRPTGAT/VOFEOP GHRDYFFGGGYVYF CYRPTGAT/VOFEOP GHRDYFFGGGYVFF CYRPTGAT/VFRA DUKYNPSR/EARNY TUKYNPSR/EARNY TUKYNPSR/EAR	нсму	WEIHHINKFAQAYSSYSRVIGGTVFVA		[32]
HFATSTGDVVISPFYNDTRENASYFG FHIPPNTYNDSOFGRPNAA HSV KTTSSEFARLOFTY CPPTOATVOFED (1) 49x200 GHRVFTGGGVVF CVSRPLSVFRHEK DADADADDFARBYN HEVVPLEVTRHEK DADADADDFARBYN HEVVPLEVTRHEK DADADADDFARBYN DDHETDMELKFANAA DDLYMPSREAHBY DDHETDMELKFANAA DDLYMPSREAHBY DDHETDMELKFANAA DDLYMPSREAHBY DDHETMELKFANAA DDHETMELKFANAA DDHETMELKFANAA DDYNFSRVEAHBY DDHETMELKFANAA DDYNFSRVEAHBY DDHETMELKFANAA DDYNFSRVEAHBY DDHETMELKFANAA DDYNFSRVEAHBY HT HEVSDEGGUK FADDTWAADANAA HEVVELSVTFUHLAFTAN HEVSDEGGUK FADDTWAADANAA HEVVELSVTFUHHALFYASAVLROPFLAGVPSA TVGFFANAATD AVOFLESTABLEFNGTVHLAFTANAKKK UPYSSSVPSTALLEFNGTVHL FTN HSV-pTM AAHUDALYAEFLGGRVLTT HSV-pTM AAHUDALYAEFLGGRVLTT HSV-pTM SSOPSTALLEFNGTVHLAEFTAGAVLROPFLAGVPSA TVGLAATTOPVAAAGG FTEGAPSVPSDVPSTALLEFNGTVHLAGTAX SSUPSTALLEFNGTVHLAETTAGTXKKK UPHRGTVHLAFTOPVAAAG SSUPSTALLEFNGTVHLAGTTAKKK TVGLAATTOPVAAAGG FTEGAPSVSSDVPSTALLEFNGTVHLAGG FTEGAPSVFSDVPSTALLEFNGTVHLAGGA HSV-HB VTVSOVFOHRYSOFAGE HSV-HB TARDLEARALAGGALAFT HSV-pTM AAHUDALYAEFLGGRVLTTFVHLAETTXKK TVGLAATTOPVAAAGG FTEGAPSVSSDVPSTALLEFNGTVHLAGGA FTUFAGGAVYF FTAGGAVYF FTAGGAVAFAGAATTSTFTF	4p+0n	WHSRGSTWLYRETANLNAMLTITTARSKYPY		
HEIPENTTIKSDEGRENAA HSV KTTSSEERALGETY CPPTGATVOFEGP (81) 4p-20n GHRPYFEGGGYVFE CYSRUVSFRYEDGG HEVVPLEYTHRIEK DARDAMDAFF TTPKETVAWDWYPKR DOG(KOARGAMDRIF TTPKETVAWDWYPKR DOG(KOARGAMDRIF DUKTWAGSIVELENTY DIKLYPASTVELENTY DI		HFFATSTGDVVYISPFYNGTNRNASYFG		
HSV KTTSSEFARLOFTY CPPTCATWOREOP (51) 49-201 GHRYTFOGGVYF CYSPICOG HEVPULEVTRHEIK DARDAUDHFARTYN TTPKFTVAWDWVPKR DOHCTOMELKPANAA DOHCTOMELKPANAA DOHCTOMELKPANAA DOHCTOMELKPANAA DOHCTOMELKPANAA DOHCTOMELKPANAA DUKTNSPRVPAFHRY DMELKPANAATTSR DMATVAGHATLEH EVDDNNKKSVPFT EVDDNNKKSVPFT EVDDNNKKSVPFT EVDDNNKKSVPFT EVDDNNKKSVPFT EVDSNNKSVPFT HSV-gH AAHLDALYAEFLGGRVLTTPVVHRALFYASAVEOPFLAGVPSA HSV-gH CAASTLTRWATVALHAF HSV-gH AAHLAFOTOPVAAAROFLA APSIVSSDVPSTALLEPROTVHLLAFDTOPVAAA HSV-gH VTSGRUNDAG HSV-GH VTSGVWFFG HSV-gH AAHLAFOTOPVAAAROFLA HSV-GH AAHLAFOTOPVAAAROFLA HSV-GH AAHLAFOTOPVAAAROFLA HSV-GH VTSGVWFFG HSV-gH AAHLAFOTOPVAAROFLA HSV-GH TTHLAFDTOPVAAROFLA HSV-GH VTSGVWFFG HSV-GH TTHLAFDTOPVAAROFLA HSV-GH TTHLAFDTOPVAAROFLA HSV-HB VTVSGVWFFG HSV-GH VTVSGVWFFG HSV-GH SVERINTSSFEARLOFTVNHG HSV-GH SVERINTSSFEARLOFFYNGTVHLAFDTOPVAAROFLA HSV-HB VTVSGVWGHRYSGFRGF TVHLAFDTOPVAAROFLA HSV-HB VTVSGVWGHRYSGFRAGFF TVLTGTOPVAAROFLA HSV-HB VTVSGVWGHRYSGFRAGFF TVLTGTOPVAAROFLA HSV-HB VTVSGVWGHRYSGFRAGFF TVHLAFDTOPVAAROFLA HSV-HB VTVSGVWGHRYSGFRAGFF TVHLAFDTOPVAAROFLAGULAFELTFN HSV-HB VTVSGVWGHRYSGFRAGFF TVHLAFDTOPVAAROFLAGULAF		FFIFPNYTIVSDFGRPNAA		
4p-20n GHRIPTTEGGGYUTF CYSRLUSERYEDGG introvpeurymetik Darabadipiranyn introvpeurymetik Darabadipiranyn introvpeurymetik Darabadipiranyn introvpeurymetik Dictronaalian introvpeurymetik Dict	HSV	KTTSSIEFARLQFTY	CPPPTGATVVQFEQP	[<u>31</u>]
HEVPULEVTRHEIK DARDANDRIFARITYN TTRKETVAWDWVPKR DOIGKOARDANDRIF TTRKETVAWDWVPKR DOIGKOARDANDRIF TTRKETVAWDWVPKR DOIGKOARDANDRIF EVIDIKINKGVCRST EVIDIKINKGVCRSTALLEPNGVKIKLAPDTOPVAAAA EVIDIKINKGVCRSTALLEPNGVKIKLAPDTOPVAAAA EVIDIKULAPDTOPVAAARGVC EVIDIKINKGVCRSTALLEPNGVKIKLAPDTOPVAAAA EVIDIKULAPDTOPVAAARGVC EVIDIKULAPDTOPVAARGVC EVIDIKULAPDTOPVAARGVCE EVIDIKULAPDTOPVAARGVCE EVIDIKULAPDTOPVAARGVCE EVIDIKULAPDTOPVAARGVCE EVIDIKULAPDTOPVAARGVCE EVIDIKULAPDTOPVAARGVCE EVIDIKULAPDTOPVAARGVCE EVIDIKULAPDTOPVAARGVC	4p+20n	GHRRYFTFGGGYVYF	CYSRPLVSFRYEDQG	
ITRKFTVANUOWYPKR DOCIGKCAREDAMORIF DUKYINGSRVEAFHAY DUKYINGSRVEAFHAY DUKYINGSRVEAFHAY DUKYINGSRVEAFHAY DIMELKPANAATRISH DIMELKANAATRISH DIM		HEVVPLEVYTRHEIK	DARDAMDRIFARRYN	
DOHETOMELIKPANAA DUKYNPSRVAAFHAY DUKYNPAAA DUKYNPAAAA DUKYNPAAAAATTSR DWELKPANAATTSR DWELKPANAATTSR DWELKPANAATTSR DWELKPANAATTSR DWELKPANAATSR EVPLSRVOLGDCIGK FADDTVIHADANAA HYNDMLGRVJAWOCE LEVYTHHEIMSGLL PYPFEVULGRVGK PYPFEVULGRVGK HYNDR PYPFEVULGRVGK PYFKATTAYYKDYTV TYSTFIDLITINLED ATSPETEVUKINAK OV CIVEEVUKINAK GULST AAHLDALYAEFLGGRVLTFPVYHRALFYASAVLROPFLAGVPSA TWLATRGLLSRSPATYNPOINTGEVULGCDAAL (29) HSV-pH AAHLDALYAEFLGGRVLTFPVYHRALFYASAVLROPFLAGVPSA TWLATRGLLSRSPATYNPOINTGEVULGCDAAL (29) ATSPETEVUKINAK GLASTLEWAYNALIAR AAHLDALYAEFLGGRVLTFPVYHRALFYASAVLROPFLAGVPSA TWLATRGLLSRSPATYNPOINTGEVULGCDAAL (29) HSV-pTM APSYSSDVPSTALLEPNGTVIHLLAFDTOPVAAIA GASTLEWAYNALIAR SSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAK TVIHLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAK TVIHLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAK TVIHLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLLAFDTOPVAAK TVIHLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLAFDTOPVAAK TVIHLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHLAFDTOPVAAK TVIHLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHILAFDTOPVAAK TVIHLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHILAFDTOPVAAK TVIHLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHILAFDTOPVAAK TVIHLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHILAFDTOPVAAK TVIHLAFDTOPVAAIA KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNGTVIHI KKSSOVPSTALLEPNG		TTPKFTVAWDWVPKR	DCIGKDARDAMDRIF	
DUKYUPSRVZAFINY DUKKYANATITISR DUKYUPSRVZAFINY DUKYUPSRVZAFINY DUKYUPSRVZAFINY DUKYUPSRVZAFINY DVKYZAGHATLEFH DVKYKAKKYKNPTP EVDKINAAGQCRST EVPLSRVDLGDCGK FADIOTVHADANAA HVNOMLGRVJANKCE LEVYTHREKDSGLL PVFEEVIDKINAKG PVFEEVIDKING PVFEEVIDKINAKG PVFEEVIDKI			DDHETDMELKPANAA	
DMELRAPANASHTSR DNATVAGHATLREH DPKPKKNKKPKNPP EVIDKINAGGVGRST EPYPERVDLGDCIGK FADIDTVIHADANAA MVWDMLGRVAJAWCE LEVYTRHENDSGLL PYPEFEVIDKINAKG PYFECVIDKINAFECVIDKI PYFECVIDKINAFECVIDKI PYFECVIDKINAFECVIDKI PYFECVIDKINAFECVIDKI PYFECVIDKINAFECVIDKI PYFECVIDKINAFECVIDKI PYFECVIDKINAFECVIDKI PYFECVIDKINAFECVIDKI PYFECVIDKINGF PYFECV			DLKYNPSRVEAFHRY	
DIVATVAAGHATILEHH DIVERVEKINKKEPKINETP EVIDKINAKGVEKIST EVIDKINAKGVEKIST EVIDKINAKGVEKIST EVIDKINAKGVEKIST EVIDKINAKGUSSIST EVIDKINAKAG PVFEI PVFEEVIDKINAKAG PVFEI PVFEEVIDKINAKAG PVFEI			DMELKPANAATRTSR	
DPKPKKINKRWPTP EVICKINAKGVCRST EVPLSRVDLGOCIGK FADIDTVIHADANAA EVPLSRVDLGOCIGK FADIDTVIHADANAA HVNDMLGRVALAWCE LEVYTRHEIVRSGLL PVPKEKATMYKOVTV PVKFKATMYKOVTV TVSTFIDLINTMLED AVOPLLSRTVLELVV ACHLIDALYAEFLOGRVLTTPVVHRALFYASAVLROPFLAGVPSA TWLATRGLIRSPGRYVYFSPSASTWPVGIWTTGELVLGCDAAL 289-20 GLASTLTRWAHYMALIRAF RLTGLATSGFAFVNAAHANGAVCLSDLIGFLAGVPSA HSV-gH AAHLIDALYAEFLOGRVLTTPVVHRALFYASAVLROPFLAGVPSA TWLATRGLIRSPGRYVYFSPSASTWPVGIWTTGELVLGCDAAL 291-20 GLASTLTRWAHYMALIRAF RLTGLATSGFAFVNAAHANGAVCLSDLIGFLAGVPSA HSV-gH AAHLIDALYAEFLOGRVLTT HSV-gTM APSVFSSDVPSTALLLFPNGTVHLAFDTOPVAAIA GPTCARSPKSSDVPSTALLLFPNGTVHLAFDTOPVAAIA GPTCARSPKSSDVPSTALLLFPNGTVHLAFDTOPVAAIA GPTCARSPKSSDVPSTALLLFPNGTVHLAFDTOFVAAIA GPTCARSPKSSDVPSTALLLFPNGTVHLAFDTOFVAAIA GPTCARSPKSSDVPSTALLLFPNGTVHLAFDTOFVAAIA KKSSDVPSTALLFPNGTVHLAFDTOKKK LIPNOTVHLAFDTOPVAAIAPGF KKSTALLLFPNGTVHLAFDTOFVAAKK TVHLAFDTOPVAAIAPGFLAA APSVFSSDVPSTALLLFPNGTVHLAFDTOFVAAIA GPTGARSPKSVRGTALLFPNGTVHLAFDTOFVAAIA GPTGARSPKSDVFSTALLLFPNGTVHLAFDTOFVAAIA GPTGARSPKSDVFSTALLLFPNGTVHLAFDTOFVAAIA KKSTALLLFPNGTVHLAFDTOFVAAKK TVHLLAFDTOPVAAIAPGF KVLTPAJLAGAJAAGIA FVUCHARVLOFP FVUCHARVLOFP FVUCHARVLOFPNUTHLAFDTOFVAAKK TVHLLAFDTOPVAAIAPGFLAA ABSVLTPAJLIFNGTVHLAFDTOFVAAKKK TVHLLAFDTOFVAAIA KSTALLFPNGTVHLAFDTOFVAAKKK TVHLLAFDTOFVAAIA KSTALLFPNGTVHLAFDTOFVAAKKK TVHLLAFDTOFVAAIAPGFLAA CONGHRSOFFALLFPNGTVHLAFDTOFVAAIAKK TVHLLAFDTOFVAAIAPGFLAA KSTALLFPNGTVHLAFDTOFVAAKKK TVHLAFDTOFVAAIAPGFLAA CONGHRSOFFALLFPNGTVHLAFDTOFVAAKKK TVHLAFDTOFVAAIAPGFLAA CONGHRSOFFALLFPNGTVHLAFDTOFVAAKKK TVHLAFDTOFVAAIAPGFLAA CONGHRSOFFALLFPNGTVHLAFDTOFVAAKKK TVHLAFDTOFVAAIAPGFLAA CONGHRSO			DNATVAAGHATLREH	
 EVIDKINAKQVERT EVIDKINAKQVERT EVIDKINAKQVERT FADIDTVIHADANAA HVNDMLGRVAIAWCE LEVYTHRIEKOSGLL PVPFEEVIDKINAKQ PVPFEVIDKINAKQ PVFEVIDKINAKQ PVFEVIDKINAKA PVFEVIDKINAKA				
HSV-gH AAHLDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA AAHLDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA AAHLDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA AAHLDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA AAHLDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA AAHLDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA AAHLDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA AAHLDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA AAHLDALYAEFLGGRVLTT HSV-JTM APSVFSSDVPSTALLLFPNGTVHLLAFDTOPVAAIA GPTEGAPSVFSSDVPSTALLLFPNGTVHL SSDVPSTALLFPNGTVHLLAFDTOPVAAIA SSDVPSTALLFPNGTVHLLAFDTOPVAAIA SSDVPSTALLFPNGTVHLLAFDTOPVAAIA SSDVPSTALLFPNGTVHLLAFDTOPVAAIA HSV-HB VTVSQVWFGHRYSQFMGIF FVLATGDFVAAIAPGFLASASA HSV-HB VTVSQVWFGHRYSQFMGIF FVLATGDFVAAIAPGFLASASA AGHLTPNGTVHLLAFDTOPVAAIAPGFLASASA STALLFPNGTVHLLAFDTOPVAAIAPGF AGAUT AP4-10 STALLFPNGTVHLLAFDTOPVAAIACKK TVHLLAFDTOPVAAIAPGFLASASA AGUTAGUAALJSTEFLGGRVLT HSV-HB VTVSQVWFGHRYSQFMGIF FVLATGDFVMAJAGGLASASA AGUTAGUAFULDFVFL AFALTPNGTVHLLAFDTOPVAAIACKK TVHLLAFDTOPVAAIAPGFLASASA AGUTAGUAFULDFSFI HSV-HB VTVSQVWFGHRYSQFMGIF FVLATGDFVMISPFGVPGPR AGUTAGUAFULDFSFI AGUTAGUAFULDFSFI AGUTAGUAFULDFSFI AGUTAGUAFULDFSFI AGUTAGUAFULDFSFI AGUTAGUAFULDFSFI AGUTAGUAFULDFSFI AGUTAGUAFULAFDLAFLAGGLAAT AGDNATVAAGHATLFRELFROJLALFINGT AGDNATVAAGHATLFRELFLEROSLALHALGYOLAFV LOLEARLOHLVAELEFN AGDNATVAAGHATLFRELFROJCALAGUAFULAFUL AGDNATVAAGHATLFRELFROJCALFULAFULAFUL AGDNATVAAGHATLFRELFROJCALSCLONHELTLTWAEK HSV-HB SUFFICHTIN AGDNATVAAGHATLFRELFROJCALSCLONHELTLTWAEK AGUNATVAAGHATLFRELFROJCALSCLONHELTLTWAEK HSV-HB SUFFICHTON ALHALGYDAFULAFULAFULAFULAFULAFULAFULAFULAFULAFUL			EVIDKINAKGVCRST	
FADIOTVIADANAA HVNDMLGRVAJAWCE LEVYTRHEIKOSGLL PVPFEVIDKINAKG PVPFEVIDKINAKG PVFFKATMYKKOTV TVSTFIDLITMLED APTSPGTPGVAATO AYOPLLSNTLAELYV CIVECVDARSVPYD SP-2n GLASTLTRWAHYNALIRAF ALIDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA TWLATRGLLESPGGRVVTSPSASTWPVGWTTGELVLCCDAAL SP-2n GLASTLTRWAHYNALIRAF AHLIDALYAEFLGGRVLTTPVVHRALFYDSAVLROPFLAGVPSA TWLATRGLLESPGGRVVTSPSASTWPVGWTTGELVLCCDAAL Sp-2n GLASTLTRWAHYNALIRAF RLTGLLATSGFAFVNAAHAAGACLSDLLGFLAHSRALAG Sp+2n GLASTLTRWAHYNALRAF RLTGLATSGFAFVNAAHAAGACLSDLLGFLAHSRALAG Sp+2n GLASTLTRWAHYNALRAF RAHUDALYAEFLGGRVLTT Sp5vFSSDVPSTALLLFPNGTVIHLAFDTOPVAAIA Sp5vFSSDVPSTALLLFPNGTVHLAFDTOPVAAIA Sp17n TVIHLLAFDTOPVAAKK STALLFPNGTVIHLLAFDTOPVAAKK KKSSDVPSTALLLFPNGTVHLAFDTOPVAAKK STALLFPNGTVHLAFDTOPVAAKK STALLFPNGTVHLAFDTOPVAAKK STALLFPNGTVHILAFDTOPVAAKK SVF			EYPLSRVDLGDCIGK	
HVNDBULGVAARVCE LEVYTRHEIKSGGL PVFFEEVIDKINAKG PVFFEEVIDKINAKG PVFFEEVIDKINAKG PVFFEEVIDKINAKG PVFFEEVIDKINAKED APTSPGTPGVAAATO AYOPLESD APTSPGTPGVAAATO AYOPLSD APTSPGTPGVAAATO AYOPLSD APTSPGTPGVAAATO AYOPLSD APTSPGTPGVAATO AYOPLSD APTSPGTPGVAATO AYOPLSD APTSPGTPGVAATO AYOPLSD APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATO APTSPGTPGVAATAGALAFCGGVVTT HSV-pTM APSVFSSDVPSTALLIFPNGTVIHLLAFDTGVAAIA GPTEGAPSVFSSDVPSTALLIFPNGTVIHLLAFDTGVAAIA GPTEGAPSVFSSDVPSTALLIFPNGTVIHLLAFDTGVKKK LFPNGTVIHLLAFDTGVVAAIAPGF KKSSDVPSTALLIFPNGTVIHLLAFDTKKK GTVIHLLAFDTGPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLLAFDTKKK GTVIHLLAFDTGPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLLAFDTKKK GTVIHLLAFDTGPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLLAFDTKKK GTVIHLLAFDTOPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLLAFDTKKK GTVIHLLAFDTOPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLAFDTKKK GTVIHLLAFDTOPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLAFDTGVVAKK TVIHLLAFDTOPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLAFDTGVVAKK TVIHLLAFDTOPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLAFDTGVVAKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLAFDTGVVAKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLAFDTGVVAKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLAFDTGVVAKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLIFPNGTVIHLAFDTGVVAKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLFPNGTVIHLAFDTGVVAKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLFPNGTVIHLAFDTGVVAKKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLFPNGTVIHLAFDTGVVAKKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLFPNGTVIHLAFDTGVVAKKKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLFPNGTVIHLAFDTGVVAKKKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLFPNGTVIHLAFDTGVVAKKKK TVIHLAFDTOPVAAIAPGF KKSSDVPSTALLFPNGTVIHLAFDTGVVAKKKK TVIHLAFDTOPVAAIAPGF KKSSDVF KKSSDVF KV-HB VTVSOVFGHRYSOFAGIF KKSSDVF KKS				
LEVYTHEKOSGL PVFFEKVIKNAKG PVFFEKVIKNAKG PVFFEKVIKNAKG PVFFEKVIKNAKG PVFFEKVIKNAKG PVFFEKVIKNAKG PVFFEKVIKNAKG PVFFEKVIKNAKG PVFFEKVIKNAKG PVFFEKVIKNAKG PVFFEKVIKNAKG PVFFEKVIKNAKG APSPGTPGVAAATO AVOPLLISNTLAELVV GUVEEVDARSVYPYD HSV-9H AAHLIDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA TWLATRGLLRSPGRVVYFSPSASTWPVGIWTTGELVLGCDAAL (2) AAHLIDALYAEFLGGRVLTT HSV-9H AAPSVFSSDVPSTALLLFPNGTVHLLAFDTOPVAAIA GPTEGAPSVFSSDVPSTALLLFPNGT HLAFDTOPVAAIAPGFLAA APSVFSSDVPSTALLLFPNGTVHLLAFDTKK KKSSDVPSTALLLFPNGTVHLLAFDTKK KKSSDVPSTALLLFPNGTVHLLAFDTKK KKSTALLLFPNGTVHLLAFDTKK HSV-HB VTVSOVWFGHRYSOFMGIF HSV-HB VTVSOVWFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVFRFKSSDVFSTALLLFPNGTVHLLAFDTKK HSV-HB VTVSOVWFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVERSTFFTIN PCTVGHRRYFTFGGGYVF YVLANGGFLAYOPLALAGIQA HSV-HB VTVSOVWFGHRYSOFMGIF FVLATOPVAAIAPSTFT HSV-HB VTVSOVWFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVWFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVWFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVWFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVWFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVMFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVMFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVMFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVMFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVMFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVMFGHRYSOFMGIF FVLATOPVAAIAPGFLASA APSVER HSV-HB VTVSOVMFGHRYSOFMGIF FVLATOPVAAIAPGCLASA APSVER HSV-HB VTVSOVMFGHRYSOFMGIF FVLATOPVAAIAPGCLASA APSVER HSV-HB VTSOLATAPOLICHEREGSLALHALGYOLAFV LQLEARLOHLVALLEREGSLAHALGYOLAFV LQLEARLOHLVAELER AGDNATVAAGHATLREHLRDIKAENTDAN TAGDDAANAVAACALILINNA AGDNATVAAGHATLREHLRDIKAENTDAN TAGDDAANAVAACALINNALITTALARADHOKTLFWL AGDNATVAAGHATLREHLRDIKAENTDAN TAGDDAANAVAACALINNALINTDIEN HSV-GBL HSV-GBL HSV-GBL HSV-GBL HSV-GBL HSV-GDL HSV-GDL HSV-GDL HSV-GDL HSV-GDL HSV-GBL HSV-GDL HSV-GDL HSV-GDL HSV-GDL HSV-GDL HSV-GDL HSV-G			HVNDMLGRVAIAWCE	
PVPEEVUKINKAG PVKRATMYYKDVTV TVSTFIDLNITMLED APTSPGTFGVAAATQ AYOPLISNTLAELYV CIVEEVDARSVYPYD HSV-gH AAHLIDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA TWLATRGLIRSPGRVVYFSPSASTWPVGIWTTGELVLGCDAAL [25] 3p+2n GLASTLTRWAHYNALIRAF RITGLLATSGFAFVNAAHANGAVCLSDLLGFLAHSRALAG [27] AAHLIDALYAEFLGGRVLTT GYPEGAPSVFSSDVPSTALLLFPNGTVIHLAFDTOPVAAIA GPTEGAPSVFSSDVPSTALLLFPNG [27] Sp+7n TVHLLAFDTOPVAARGFGLAA APSVFSSDVPSTALLLFPNGTVIHLAFDTKK [27] KKSSDVPSTALLLFPNGTVIHLLAFDTKK GYPEGAPSVFSSDVPSTALLLFPNGTVIHLAFDTKK [27] KKSSDVPSTALLLFPNGTVIHLLAFDTKK GYPEGAPSVFSSDVPSTALLLFPNGTVIHLAFDTOPVAAR [27] SSDVPSTALLFPNGTVIHLAFDTKK GYPEGAPSVFSSDVPSTALLFPNGTVIHLAFDTOPVAAR [27] KKSSDVPSTALLFPNGTVIHLAFDTKK GYPEGAPSVFSSDVPSTALLFPNG [27] KKSSDVPSTALLFPNGTVIHLAFDTKK FVIGTOPVAAR [27] KKSSDVPSTALLFPNGTVIHLAFDTOPVAAKKK FVIHLAFDTOPVAAR [27] KKSSDVPSTALLFPNGTVIHLAFDTOPVAAKKKK TVHLLAFDTOPVAAR [28] KKSSDVPSTAULFPNGTVIHLAFDTTKK TVHLAFDTOPVAAR [29] APHAN SVERIKTSSIEFARLOFTVNHQ VGSFRRSSDASTFFTT			LEVYTRHEIKDSGLL	
HSV-HB VTVSQUVFGHRYSQFMGIF FVLATGDFVXHQ HSV-HB VTVSQUVFGHRYSQFMGIF QSQLALHALGYQLAFV HSV-HR TARLQLEARLQHUVAELERGSLALHALGYQLAFV HSV-HR ALHUZQLAFTLGGRVTTFVVHQLALARADHOKTLFWL HSV-PTM ALHUZQLAFTLGGRVLTT HSV-PTM TARLQLEARLQHUVAELQMATKACHALARADHOKTLFWL HSV-PTM TARLQLEARLQHUVAELQMATKACHALARADHOKTLFWL HSV-PTM TARLQLEARLQHUVAELQMATKACHALARADHOKTLFWL HSV-PTM TARLQLEARLQHUVAELQMACLONHELTLWNEARK (50) HSV-PTM TARLQLEARLQHUVAELQMACLONHELTLWNEARK (50) HSV-PTM TARLQLEARLQHUVAELQMACLONHELTLWNEARK (50) HSV-PTM TARLQLEARLQHUVAGHALARADHOKTLFWL HSV-PTM TARLQLEARLQHUVAGHALARADHOKTLFWL HSV-PTM TARLQLEARLQHUVAGHALARADHOKTLFWL HSV-PTM TARLQLEARLQHUVAGHALARADHOKTLFWL HSV-PTM TARLQLEARLQHUVAGHALARADHOKTLFWL HSV-PTM TARLQLEARLQHUVAGHALARADHOKTLFWL HSV-PTM TARLQLEARLQHUVAGHALARADHOKTLFWL HSV-HR TARLQLEARLQHUVAGHALTALARADHOKTLFWL TACQDRAMAVAKAGHALTLREHLRDIKAENTDAN TACAQARAMAVAKAGHALTREHLRDIKAENTDAN TACAQARAMAVAKAGHALTALARADHOKTLFWL HSV-HR TARLQLEARLQHUVAGHATLREHLRDIKAENTDAN TACAQARAMAVAKAGHATLREHLRDIKAENTDAN TACAQARAMAVAKAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARAAQATLSKRDRAGH TACAQARAMAVAKAGHATLREHLRDIKAENTDAN TACAQARAMAVAKAGHATLREHLRDIKAENTDAN TACAQARAMAVAKAGHATLREHLRDIKAENTDAN TACAQARAMAVAKAGHATLREHLRDIKAENTDAN				
APTSPETPGVAAATQ AYQPLLSNTLAELYV CIVEEVDARSVYPYD HSV-gth AAHLIDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA TVLATRGLRSPGRVVFSPSASTWPVGIWTTGELVLGCDAAL [26] 3p+2n GLASTLTRWAHYNALIRAF RLTGLLATSGFAFVNAAHANGAVCLSDLLGFLAHSRALAG AAHLIDALYAEFLGGRVLTT HSV-gTM PSVFSSDVPSTALLLFPNGTVIHLLAFDTQVAAIA GPTEGAPSVFSSDVPSTALLLFPNG TVIHLAFDTOPVAAIAPGFLAA APSVFSSDVPSTALLLFPNG SDVPSTALLLFPNGTVIHLLAFDTKKK LFPNGTVIHLLAFDTQPVAAIAPGF KKSSDVPSTALLLFPNGTVIHLLAFDTKKK GTVIHLLAFDTQPVAAIA STALLFPNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIA KKSSDVPSTALLLFPNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIAQGFLAS STALLFPNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIAQGFLASA KKSTALLFPNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIAPGFLASA KKSTALLFPNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIAPGFLASA KKSTALLFPNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIAPGFLASA KKSTALLFPNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIAPGFLASA KKSTALLFPNGTVIHLLAFDTQPVAAKKK TVIHLAFDTQPVAAIAPGFLASA KKSTALLFPNGTVIHLLAFDTQPVAAKKK TVIHLAFDTQPVAAIA KKSTALLFPNGTVIHLLAFDTQPVAAKKK TVIHLAFDTQPVGYRE (28] 4p+4n SVERIKTTSSIEFARLOFTVNHIQ YGGSFRFSSDAISTFTTN TARJGLSALGTGVGFV SVERIKTTSSIEFARLOFTVNHIQ AGDNATVAAGHATLREHLRDIKAENTDAN TARGAGLARJOHLSTLFVL AGDNATVAAGHATLREHLRDIKAENTDAN TAAGDARANAVAKAGLHDLNIETCHVL AGDNATVAAGHATLREHLRDIKAENTDAN TAAGDARANAVAKAGLHDLNIETCHVL HSV-gBh SIEFARLOFTVNHIOGRVAIAWCELONHELTLWNEARK (S)				
Arisper provide and a second a second and a second a second and a second and a second				
HTUPLESNITJAELTY CIVEEVDARSVYPYD HSV-gH AAHLIDALYAEFLGGRVLTTPVVHRALFYASAVLRQPFLAGVPSA TWLATRGLLRSPGRYVYFSPSASTWPVGIWTTGELVLGCDAAL [26] 3p+2n GLASTLTRWAHYNALIRAF AAHLIDALYAEFLGGRVLTT HSV-pTM APSVFSSDVPSTALLLFPNGTVIHLLAFDTOPVAAIA GPTEGAPSVFSSDVPSTALLLFPNG [27] 6p+7n TVIHLLAFDTOPVAAIAPGFLAA APSVFSSDVPSTALLLFPNGTVIHL SSDVPSTALLLFPNGTVIHLLAFDTKKKK LPPNGTVIHLLAFDTOPVAAIAPGF KKSSDVPSTALLLFPNGTVIHLLAFDTKKKK IPPNGTVIHLLAFDTOPVAAIA STALLFPNGTVIHLLAFDTQPVAAKKKK TVIHLAFDTOPVAAIA KKSTALLLFPNGTVIHLLAFDTQPVAAKKKK TVIHLLAFDTOPVAAIA STALLFPNGTVIHLLAFDTOPVAAKKKK TVIHLAFDTOPVAAIA KKSTALLLFPNGTVIHLLAFDTOPVAAKKKK TVIHLAFDTOPVAAIA STALLFPNGTVIHLLAFDTOPVAAKKK HSV-HB VTVSOVWFGHRYSOFMGIF HSV-HB VTVSOVWFGHRYSOFMGIF HSV-HB VTVSOVWFGHRYSOFMGIF HSV-HB VTVSOVWFGHRYSOFMGIF Ap+4n SVERIKTTSSIEFARLOFTYNHIQ YGGSFRFSSDAISTTTIN PCTVGHRRYFTFGGGYVFF YAYSHOLSRADITTVSTFI HSV-HR TARLOLEARLOHLVAEILERBOSLALHALGYOLAFV LOLEARLOHLVAEILER AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRADLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN TAAGQARANAVAKAGLHDLNIETDTERNH VEGQLGENNELRLTRDAI BOYREEOLGERALOHLVAEILERBOSLALHALGYOLAFV UCGGUNGELATSVALATRADLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN TAAGQARANAVAKAGLHDLNIETDTERNH VEGQLGENNELRLTRDAI BVY-gBh SIEFARLOFTYNHIORHVNDMLGRVAIAWCELQNHELTLWNEARK				
CVCEEULARSVTPTD CVCEEULARSVTPTD BSV-gH AAHLIDALYAEFLGGRVLTTPVVHRALFYASAVLROPFLAGVPSA TWLATRGLLRSPGRYVYFSPSASTWPVGIWTTGELVLGCDAAL [26] 3p+2n GLASTLTRWAHYNALIRAF RLTGLLATSGFAFVNAAHANGAVCLSDLLGFLAHSRALAG [26] AAHLIDALYAEFLGGRVLTT HSV-pTM APSVFSSDVPSTALLLFPNGTVIHLLAFDTQPVAAIA GPTEGAPSVFSSDVPSTALLLFPNGTVIHL [27] Gp+7n TVIHLLAFDTOPVAAIAPGFLAA APSVFSSDVPSTALLLFPNGTVIHLLAFDTKKKK [27] SSDVPSTALLLFPNGTVIHLLAFDTKKKK GTVIHLLAFDTOPVAAIA [28] KKSSDVPSTALLLFPNGTVIHLLAFDTQPVAAKK TVIHLLAFDTOPVAAIA [28] STALLFPNGTVIHLLAFDTQPVAAKK TVIHLLAFDTOPVAAIA [28] HSV-HB VTVSQVWFGHRYSOFMGIF FVLATGPLYVMSPFYQYRE [28] HSV-HB VTVSQVWFGHRYSOFMGIF FVLATGPLYVGYGYTE [29] HSV-HB VTVSQVWFGHRYSOFMGIF FVRGHTGFVVCYQGYTGFPR [29] HSV-HB VTVSQVWFGHRYSOFMGIF FVRGHTGFVVCYQGYTGFPR [29] HSV-HB VTVSQVWFGHRYSOFMGIF FVRGHTGFVVCYQGYTGFPR [29] HSV-HB VTVSQVWFGHRYSOFMGIF FVRGHTGFVVCYQGYTGFPR [29] HSV-HB VTVSQVWFGHRYSOFMALLGPLAGQLAFV LOLEARLOHLVAEILER [29] 4p+9n ALHALQGVLAFVLOSPSAY YOFHLVLHEALRAQALSROLLGRELAOELVAELAT [29]				
AAHLDALTAEFLOORVETTPVMAALTAF TASAVENUPPLAGVPSA TWEATNEGLEASPERTVTSPSASTMPV0101116EUV2020AAL [23] 3p-2n GLASTLTRWAHYNALIRAF RSAVENUPPLAGVPSA RLTSGFAFVNAAHANGAVCLSDLLGFLAHSRALAG AAHLDALYAEFLGGRVLTT HSV-pTM APSVFSSDVPSTALLLFPNGTVIHLLAFDTQPVAAIA GPTEGAPSVFSSDVPSTALLLFPNG TVIHLLAFDTQPVAAIAPGFLAA APSVFSSDVPSTALLLFPNGTVIHL SSDVPSTALLLFPNGTVIHLLAFDTKKK LFPNGTVIHLAFDTQPVAAIA KKSSDVPSTALLLFPNGTVIHLLAFDTKKK GTVIHLLAFDTQPVAAIA STALLFPNGTVIHLLAFDTQPVAAKKK TVIHLAFDTQPVAAIA KKSSTAULFPNGTVIHLLAFDTQPVAAKKKK TVIHLAFDTQPVAAIA KKSTAULFPNGTVIHLLAFDTQPVAAKKKK TVIHLAFDTQPVAAIA KKSTAULFPNGTVIHLLAFDTQPVAAKKK TVIHLAFDTQPVAAIA KKSTAULFPNGTVIHLAFDTQPVAAKK [1] HSV-HB VTVSQVWFGHRYSQFMGIF FVLATQDFVMAPFYGYRE [2] 4p+4n SVERIKTSSIEFARLQFTYNHIQ YGGSFRFSSDAISTTFTN PCTVGHRRYFTFGGGVVYF YVLANGGFLAYQPLSNT YAYSHQLSRADITTVSTFI HSV-HR TARLOLEARLQHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER [2] 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALARDHQKTLFWL AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATRNDLHRAARADAGHLSKRDRAGH ATLREHLRDIKAENTDAN ATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLRTALARADHQKTLFWL SIEFARLQGTYNHORHVNDMLGRVAIAWCELQNHELTLWMEARK				[06]
AAHLIDALYAEFLGGRVLTT HSV-pTM APSVFSSDVPSTALLLFPNGTVIHLLAFDTQPVAAIA GPTEGAPSVFSSDVPSTALLLFPNG PSVFSSDVPSTALLLFPNGTVIHLLAFDTQPVAAIA GPTEGAPSVFSSDVPSTALLLFPNGTVIHLAFDTQPVAAIAPGF (KKSSDVPSTALLLFPNGTVIHLLAFDTKKKK LFPNGTVIHLLAFDTQPVAAIAPGF KKSSDVPSTALLLFPNGTVIHLLAFDTKKKK GTVIHLLAFDTQPVAAIAPGF KKKSSDVPSTALLLFPNGTVIHLLAFDTKKKK GTVIHLLAFDTQPVAAIA STALLFPNGTVIHLLAFDTQPVAAKKKK TVIHLAFDTQPVAAIA STALLFPNGTVIHLLAFDTQPVAAKKKK TVIHLAFDTQPVAAIA STALLFPNGTVIHLAFDTQPVAAKKKK GTVIHLAFDTQPVAAIA STALLFPNGTVIHLAFDTQPVAAKKKK GTVIHLAFDTQPVAAIA STALLFPNGTVIHLAFDTQPVAAKKKK GTVIHLAFDTQPVAAIA STALLFPNGTVIHLAFDTQPVAAKKK GTVIHLAFDTQPVAAIA STALLFPNGTVIHLAFDTQPVAAKK GTVIHLAFDTQPVAAIA STALLFPNGTVIHLAFDTQPVAAKK GTVIHLAFDTQPVAAIA STALLFPNGTVIHLAFDTQPVAAKK GTVIHLAFDTQPVAAIA SVERIKTSSIEFARLQFTYNHQ YQGSRFSSDAISTTFTN FVRGHTGFVYCYGYTF YYLANGGFLAYQPLSNT YAYSHOLSRADITTSTFI FVRGHTGFVYCYGYTGPPR HSV-HR TARLQLEARLOHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER AGDNATVAAGHATLREHLRDIKAENTDAN ATLREHLRDIKAENTDAN ATAGDAGARAVAKAGLHDLNIETTENH VEGQLGERNELRLTRDA	nov-gn			[20]
AAALDAL TAEFGGAVUTI APSYFSDVPSTALLLEPNGTVIHLLAFDTQPVAAIA GPTEGAPSVFSSDVPSTALLLEPNG [27] 6p+7n TVIHLLAFDTQPVAAIAPGFLAA APSVFSSDVPSTALLLEPNGTVIHL [27] 6p+7n TVIHLLAFDTQPVAAIAPGFLAA APSVFSSDVPSTALLLEPNGTVIHL [27] 6p+7n TVIHLLAFDTQPVAAIAPGFLAA APSVFSSDVPSTALLLEPNGTVIHL [28] 6p+7n TVIHLLAFDTQPVAAIA GTVIHLLAFDTQPVAAIA [28] 6p+7n STALLEPNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIA [28] 6p+7n STALLEPNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIAPGFLAASA [28] 6p+7n STALLEPNGTVIHLLAFDTQPVAAKK STALLEPNGTVIHLAFDTQPVAAIAPGFLAASA [28] 6p+7n SVERIKTTSSIEFARLOFTYNHIQ YGGSFRFSDAJSTFFTN [28] 6p+4n SVERIKTTSSIEFARLOFTYNHIQ YGGSFRFSDAJSTFFTN [28] 6p+4n SVERIKTTSSIEFARLOFTYNHIQ YGGSFRFSDAJSTFFTN [29] 6p+4n SVERIKTTSSIEFARLOFTYNHIQ YGGSFRFSDAJSTFFTN [29] 6p+4n SVERIKTTSSIEFARLOFTYNHIQ YGGSFRFSDAJSTFFTN [29] 6p+4n SVERIKTSSDAJSTFFTN FVRGHTGFYVCYGYTGPR [29] HSV-HR TARLQLEARLQHLVAEILEREQSLALHALGYQLAFV LQL	sp+zn		REIGELAISGFAFVNAAHANGAVCESDELGFLAHSRALAG	
AFSVF3SDVFSTALLLEFNG (2) 6p+7n TVIHILLAFDTQPVAAIAPGFLAA APSVFSSDVPSTALLLEFNGTVIHL SSDVPSTALLLEFNGTVIHLLAFDTKKKK LEPNGTVIHLLAFDTQPVAAIAPGF KKSSDVPSTALLLEFNGTVIHLLAFDTKKKK EPNGTVIHLLAFDTQPVAAIAPGF KKSSDVPSTALLLEFNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIA STALLLEFNGTVIHLLAFDTQPVAAKKK TVIHLLAFDTQPVAAIA KKSTALLEFNGTVIHLLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLEFNGTVIHLLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLEFNGTVIHLLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLEFNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLEFNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIAPGFLASA KKSTALLEFNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIAPGF KKSTALLEFNGTVIHLAFDTQPVAAKKKK TVIHLLAFDTQPVAAIA KKSTALLEFNGTVIHLAFDTQPVAAKKKK TVIHLLAFDTQPVAAIAPGF KKSTALLEFNGTVIHLAFDTQPVAAKKKK TVIHLLAFDTQPVAAIAPGF KKSTALLEFNGTVIHLAFDTQPVAAKKKK TVIHLLAFDTQPVAAIAPGF KKSTALLEFNGTVIHLAFDTQPVAAKKKK TVIHLAFDTQPVAAIAPGFLASA KKSTALLEFNGTVIHLAFDTQPVAAKKKK TVIHLAFDTQPVAAIAPGF KKSTALLEFNGTVIHLAFDTQPVAAKKKKKK TVIHLAFDTQPVAAIAPGFLASA KKSTALLEFNGTVIHUAFDTQPVAAKKKKK TVIHLAFDTQPVAAIAPGFLASA YAYSOVWFGHRYSOFMGIF FVLATQDFVMSPYGYRE YAYSHQLSRADITTVSTFI FVLATQGFVMSPYGYRE KALALGQLAFVLDSPSAY YOFHL				[27]
APSYTSUPSTALLLFPNGTVIHLLAFDTKKK APSYTSUPSTALLFPNGTVIHL SSDVPSTALLLFPNGTVIHLLAFDTKKK LFPNGTVIHLAFDTKKK GTVIHLLAFDTQPVAAIA STALLLFPNGTVIHLLAFDTQPVAAKKKK TVIHLLAFDTQPVAAIA KKSSDVPSTALLLFPNGTVIHLLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLLFPNGTVIHLLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLLFPNGTVIHLLAFDTQPVAAKK [70] HSV-HB VTVSQVWFGHRYSQFMGIF FVLATGDFVYMSPFYGYRE [28] 4p+4n SVERIKTTSSIEFARLQFTYNHIQ YGGSFRFSSDAISTTFTTN PCTVGHRRYFTFGGGYVYF YYLANGGFLIAYQPLLSNT YAYSHQLSRADITTVSTFI HSV-HR TARLQLEARLQHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER [29] 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT RARRSLLIASALCTSDVAAATNADLRTALARADHQKTLFWL TSDVAAATNADLRTALARADHQKTLFWL AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN ATLREHLRDIKAENTDAN HSV-9Bh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK [20]	60.70			21
KKSSDVPSTALLLFPNGTVIHLLAFDTQPVAAKKK GTVIHLLAFDTQPVAAIA STALLLFPNGTVIHLLAFDTQPVAAKKKK TVIHLLAFDTQPVAAIA KKSSDVPSTALLLFPNGTVIHLLAFDTQPVAAKKKK TVIHLLAFDTQPVAAIA KKSTALLLFPNGTVIHLLAFDTQPVAAKKKK TVIHLLAFDTQPVAAIA KKSTALLLFPNGTVIHLLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLLFPNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLLFPNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLLFPNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLFPNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLFPNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLFPNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLFPNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIA KKSTALLFPNGTVIHLAFDTQPVAAKK TVIHLLAFDTQPVAAIA HSV-HB VTVSQVWFGHRYSQFMGIF FVLATGDFVYMSPFYGYRE [28] 4p+4n SVERIKTSSIEFARLOFTNHIQ YQGSFRFSSDAISTFTTN [29] YSHQLSRADITTVSTFI FVRGHTGFVYCYGYTGFPR [29] HSV-HR TAALQLEARLOHLVAEILEREQSLALHALGYQLAFV LOLEARLOHLVAEILER [29] 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSROLILGRELAQELVAELAT [29] 4p+9n ALHALGYGLAFVLDSPSAY YQFHLVLHEALRAQALSROLILGRELAQELVAELAT [29] 4p+9n ALHALGYGLAFVLDAN HATCSLAFALATSVALATRNDLLLRWAARDAQTILSKRDRAGH AGDNATVAAGHATLREHLRDIK	op+/11			
STALLLEP NGTVIHILLAFDTQPVAAKKK GTVIHILLAFDTQPVAAIA STALLLEPNGTVIHILLAFDTQPVAAKKK TVIHILLAFDTQPVAAIA KKSTALLLEPNGTVIHILLAFDTQPVAAKKK TVIHILLAFDTQPVAAIA KKSTALLLEPNGTVIHILLAFDTQPVAAKK TVIHILLAFDTQPVAAIA SHVLTAPALTFNLTDFVPILALAGIQA SHVLTAPALTFNLTDFVPILALAGIQA HSV-HB VTVSQVWFGHRYSQFMGIF FVLATGDFVYMSPFYGYRE [28] 4p+4n SVERIKTTSSIEFARLQFTYNHIQ YGGSFRFSSDAISTTFTTN PCTVGHRRYFTFGGGVVYF YYLANGGFLIAYQPLLSNT [29] YAYSHQLSRADITTVSTFI FVRGHTGFVYCYGYTGFPR HSV-HR TARLQLEARLQHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER [29] 4p+9n ALHALGYQLAFVLDSPSAY YOFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] 4p+9n ALHALGYQLAFVLDSPSAY YOFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] 4p+9n ALHALGYQLAFVLDSPSAY YOFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLILRWAAARDAQTILSKRDRAGH [29] AGDNATVAAGHATLREHLRDIKAENTDAN ATLREHLRDIKAENTDAN TAAGDARANAVAKAGLHDLNIETDTERNH VEGQLGENNELRLTRDAIE [20] [20] [20] HSV-gBh SIEFARLQFTYNHIORHVNDMLGRVAIAWCELQNHELTLWNEARK [20]				
Initial Find Mark TMILLA FOTO VAAIAN KKSTALLEPNGTVIHLLAFDTOPVAAKK TVIHLLAFDTOPVAAIAPGFLAASA SHVLTAPALTFNLTDFVPILALAGIQA SHVLTAPALTFNLTDFVPILALAGIQA HSV-HB VTVSQVWFGHRYSQFMGIF FVLATGDFVYMSPFYGYRE [28] 4p+4n SVERIKTTSSIEFARLQFTYNHIQ YGGSFRFSSDAISTTFTTN [29] YAYSHOLSRADITTVSTFI FVRGHTGFVYCYGYTGFPR [29] HSV-HR TARLQLEARLQHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER [29] 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLIGRELAQELVAELAT [29] 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLIGRELAQELVAELAT [29] AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLTRALARADHQKTLFWL SDVAAATNADLRTALARADHQKTLFWL [29] AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLILRWAAARDAQTILSKRDRAGH [29] MATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLTRANG [29] KKSTALLFPNITOR YGGJGENNELRTRDAI [29] KKSTALLFPNITOR YGGJGENNELRITRDAI [29] ALHALGYQLAFVLDSPSAY YGPHLVLHEALRAQALSROLILRWAAARDAQTILSKRDRAGH [29] AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLTRWAAARDAQTILSKRDRAGH [29]				
INTELECT INSTRUCTION OF ALL PERFORMANCE SHULTAPALTFNLTDEVENDATION OF ALL PERFORMANCE SHVLTAPALTFNLTDEVENDATION OF ALL PERFORMANCE SHVLTAPALTFNLTDEVENDATION OF ALL PERFORMANCE HSV-HB VTVSQVWFGHRYSQFMGIF FVLATGDFVYINSPFYGYRE [28] 4p+4n SVERIKTTSSIEFARLQFTYNHIQ YGGSFRFSSDAISTTFTTN [29] PCTVGHRRYFTFGGGYVYF YYLANGGFLIAYQPLLSNT [29] YAYSHQLSRADITTVSTFI FVRGHTGFVYCYGYTGFPR [29] HSV-HR TARLQLEARLQHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER [29] 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] 4p+9n ALHALGYQLAFVLOSPSAY YGGERALGYLAATNADLRTALARADH				
HSV-HB VTVSQVWFGHRYSQFMGIF FVLATGDFVYMSPFYGYRE [28] 4p+4n SVERIKTTSSIEFARLQFTYNHIQ YGGSFRFSSDAISTTFTTN [28] PCTVGHRRYFTFGGGYVVF YYLANGGFLIAYQPLLSNT [29] YAYSHQLSRADITTVSTFI FVRGHTGFVYCYGYTGFPR [29] HSV-HR TARLQLEARLQHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER [29] 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] Ap+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] Ap+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] Ap+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] Ap+9n ALHALGYQLAFVLDSPSAY YGFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] Ap+9n ALHALGYQLAFVLDSPSAY YGFHLVLHEALRAQALSRQLILGRELAQELVAELAT [29] AgDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN AGDNATVAAGHATLREHLRDIKAENTDAN ATLREHLRDIKAENTDAN YGGQLGENNELRLTRDAIE UVREEQLGENNELRLTRDAI UVREEQLGGERATGLNLNI				
Ap+4n SVERIKTTSSIEFARLQFTYNHIQ YGGSFRFSSDAISTTFTTN 4p+4n SVERIKTTSSIEFARLQFTYNHIQ YGGSFRFSSDAISTTFTTN PCTVGHRRYFFFGGGYVYF YYLANGGFLIAYQPLLSNT YAYSHQLSRADITTVSTFI FVRGHTGFVYCYGYTGFPR HSV-HR TARLQLEARLQHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLIIGRELAQELVAELAT RARRSLLIASALCTSDVAAATNADLRTALARADHQKTLFWL TSDVAAATNADLRTALARADHQKTLFWL AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN TAAGDARANAVAKAGLHDLNIETDTERNH VEGQLGENNELRLTRDAI GENNELRLTRDAI HSV-gBh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK	HSV-HB	VTVSQVWEGHBYSQEMGIE	EVI ATGDEVYMSPEYGYRE	[28]
PCTWIN FORUME FORUME FUNDING FORUME FUNDING PCTVGHRRYFFFGGGYVYF YYLANGGFLIAYQPLLSNT YAYSHQLSRADITTVSTFI FVRGHTGFVYCYGYTGFPR HSV-HR TARLQLEARLQHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT RARRSLLIASALCTSDVAAATNADLRTALARADHQKTLFWL TSDVAAATNADLRTALARADHQKTLFWL TSDVAAATNADLRTALARADHQKTLFWL AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN HSV-HB SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK GUN GUN	4n+4n		YGGSEBESSDAISTTETTN	[20]
YAYSHQLSRADITTVSTFI FVRGHTGFVYCYGYTGFPR HSV-HR TARLQLEARLQHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT RARRSLLIASALCTSDVAAATNADLRTALARADHQKTLFWL TSDVAAATNADLRTALARADHQKTLFWL GUVAATNADLRTALARADHQKTLFWL AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN HSV-HR YEGQLGENNELRLTRDAI TAAGDARANAVAKAGLHDLNIETDTERNH VEGQLGENNELRLTRDAI GENNELRLTRDAI DVREEEQLGERATGLNLNI	-191-111	PCTVGHBBYFTFGGGYVYF	YYLANGGELIAYOPLLSNT	
HSV-HR TARLQLEARLQHLVAEILEREQSLALHALGYQLAFV LQLEARLQHLVAEILER [2] 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT RARRSLLIASALCTSDVAAATNADLRTALARADHQKTLFWL TSDVAAATNADLRTALARADHQKTLFWL AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN YEGQLGENNELRLTRDAI VEGQLGENNELRLTRDAI GENNELRLTRDAI DVREEEQLGERATGLNLNI HSV-gBh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK [30]		YAYSHQI SBADITTVSTFI	EVBGHTGEVYCYGYTGEPB	
4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLIGRELAQELVAELAT 4p+9n ALHALGYQLAFVLDSPSAY YQFHLVLHEALRAQALSRQLIGRELAQELVAELAT RARRSLLIASALCTSDVAAATNADLRTALARADHQKTLFWL TSDVAAATNADLRTALARADHQKTLFWL AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH VEGQLGENNELRLTRDAN TAAGDARANAVAKAGLHDLNIETDTERNH VEGQLGENNELRLTRDAIE GENNELRLTRDAIE BUNETERLTRDAI DVREEEQLGERATGLNLNI	HSV-HB	TABLOL EABLOHLVAEILEBEQSLALHALGYQLAEV		[29]
RARRSLLIASALCTSDVAAATNADLRTALARADHQKTLFWL TSDVAAATNADLRTALARADHQKTLFWL AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN ATLREHLRDIKAENTDAN TAAGDARANAVAKAGLHDLNIETDTERNH VEGQLGENNELRLTRDAIE GENNELRLTRDAI GENNELRLTRDAI HSV-gBh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK ISIO	4p+9n	ALHALGYOLAFVLDSPSAY	YQFHLVLHEALRAQALSRQLILGRELAQELVAELAT	[20]
AGDNATVAAGHATLREHLRDIKAENTDAN HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH ATLREHLRDIKAENTDAN TAAGDARANAVAKAGLHDLNIETDTERNH VEGQLGENNELRLTRDAIE GENNELRLTRDAI DVREEEQLGERATGLNLNI HSV-gBh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK [30]	,	RARRSLLIASALCTSDVAAATNADLRTALARADHQKTLFWL	TSDVAAATNADLRTALARADHQKTLFWL	
ATLREHLRDIKAENTDAN TAAGDARANAVAKAGLHDLNIETDTERNH VEGQLGENNELRLTRDAIE GENNELRLTRDAI DVREEEQLGERATGLNLNI HSV-gBh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK [30]		AGDNATVAAGHATLREHLRDIKAENTDAN	HATCSLAFALATSVALATRNDLLLRWAAARDAQTILSKRDRAGH	
TAAGDARANAVAKAGLHDLNIETDTERNH VEGQLGENNELRLTRDAIE GENNELRLTRDAI DVREEEQLGERATGLNLNI HSV-gBh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK			ATLREHLRDIKAENTDAN	
VEGQLGENNELRLTRDAIE GENNELRLTRDAI DVREEEQLGERATGLNLNI HSV-gBh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK [30]			TAAGDARANAVAKAGLHDLNIETDTERNH	
GENNELRLTRDAI DVREEEQLGERATGLNLNI HSV-gBh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK [30]			VEGQLGENNELRLTRDAIE	
DVREEEQLGERATGLNLNI HSV-gBh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK [30]			GENNELRLTRDAI	
HSV-gBh SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK [30]			DVREEEQLGERATGLNLNI	
	HSV-gBh	SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHELTLWNEARK		[30]



	Active peptides	Non-active peptides	Ref
11p+0n	SIEFARLQFTYNHIQRHVNDMLGR		
	VAIAWCELQNHELTLWNEARK		
	FARLQFTYNHIQRHVNDMLGR		
	FARLQFTYNHIQRHVRDMEGR		
	YNHIQRHVNDMLGR		
	YNHIQRHVNDMLGRVAIAWCE		
	YNHIQRHVNDMLGRVKKAWEE		
	FARLQFTYNHIQRHVNDMLGRVAIAWCE		
	FARLQFTYNHIQRHVNDMLGRVKKAWEE		
	SIEFARLQFTYNHIQRHVNDMLGRVAIAWCELQNHE		
JEV	ATSSANSKA		[21]
1p+0n			
RVFV	SGSWNFFDWFSGLMSWFGGPL		[22]
2p+0n	WNFFDWFSGLMSWFGGPLK		
HCV	MANAGLQLLGFILAFLGWIGAIVS		[25]
CLDN-1	MANAGLQLLGFILAFLGW		
11p+0n	LLGFILAFLGWIGAIVST		
	FILAFLGWIGAIVSTALP		
	AFLGWIGAIVSTALPQWR		
	GWIGAIVSTALPQWRIYS		
	GAIVSTALPQWRIYSYAG		
	MANAGLQLLGFILAFL		
	MANAGLQLLGFILAFLGWIG		
	MANAGLQLLGFILAFLGWIGAI		
	MANAGLQLLGFILAFLGW		
SARS	MWKTPTLKYFGGFNFSQIL		[11]
11p+6n	ATAGWTFGAGAALQIPFAMQMAY		
	GYHLMSFPQAAPHGVVFLHVTW		
	GVFVFNGTSWFITQRNFFS		
	AACEVAKNLNESLIDLQELGKYEQYIKW		
	PTTFMLKYDENGTITDAVDC		[12]
	YQDVNCTDVSTAIHADQLTP		
	QYGSFCTQLNRALSGIAAEQ		
	IQKEIDRLNEVAKNLNESLI		
	NGIGVTQNVLYENQKQIANQFNKAISQIQESLTTTSTA	FKLPLGINITNFRAILTAFS	[13]
	IQKEIDRLNEVAKNLNESLIDLQELGK	VLYNSTFFSTFKCYGVSATK	
		PALNCYWPLNDYGFYTTSGI	
		RDVSDFTDSVRDPKTSEILD	
		SNNTIAIPTNFSISITTEVM	
		GIGVTQNVLYENQKQIANQF	
FF-3	VEDTKIDLWSYNAELL		[15]
1p+0n			
PICV	GHTLKWLLELHFNVLHVTRHIGARCKT		[14]
5p+0n	HLIASLAQIIGDPKIAWVGK		
	HYNFLIIQNTTWENHCTYT		
	PGGYCLEQWAIIWAGIKCF		
	LNLFKKTINGLISDSLVIR		
HCV	VSGIYHVTNDCSNSSIVY		[24]
4p+0n	PSQKIQLVNTNGSWHINR		
	DYPYRLWHYPCTVNFTVF		
	YLYGIGSAVVSFAIKWEY		

* The sequences in bold were used in the 75p+75n training set; the rest sequences were used in the 26p+26n test set.

doi:10.1371/journal.pone.0144171.t001

Data set	Model	Sensitivity	Specificity	Accuracy	МСС
EAP	EAPphysico	79.37	71.26	74.67	0.5
(default parameters)					
	EAPcompo	66.99	87.23	73.33	0.5
	EAPscoring	100	92.59	96	0.92
EAP	EAPphysico	80	72.94	76	0.52
(optimized parameters)					
	EAPcompo	94.67	94.67	94.67	0.89
	EAPscoring	100	97.4	98.67	0.97

TADIE 2. FETTOTTIATICE OF THE AVED EU ATHE EAFDIEU THOUEIS TAITHING SELV	Table 2. I	Performance of the AV	Ppred and EAP	pred models train	ina set V ^{75p+75}
--	------------	-----------------------	---------------	-------------------	-----------------------------

doi:10.1371/journal.pone.0144171.t002

the SVM learning. A third SVM model EAPcompo was also built where the fractions of amino acids in a peptide were used as input features in the machine learning process.

SVM training. The SVM models were trained using the experimentally validated 75p+ 75n data sets. During 5-fold cross validation, the training set was randomly partitioned into four subsets with equal size of (15p+15n) and a remaining subset (15p+15n). Three SVM models were built using sequence-based statistical scores, physicochemical properties and amino acid composition, respectively. The performances of the three models are shown in <u>Table 2</u>. It can be seen that the EAPscoring model performed best among the three models during 5-fold cross validation. A "grid-search" combined with cross-validation was adopted to search for the optimal parameters -c and -g in SVM models [49]. The result of the grid search is shown in the support information (<u>S1 File</u>). It is shown that the performances of three EAP models during 5-fold cross validation have been improved significantly using the optimized parameters (<u>Table 2</u>).

Evaluation of the predictive efficiency of SVM models on independent test set

The performance of the SVM models was evaluated using an independent dataset of experimentally validated peptides that were not contained in the learning dataset (<u>Table 1</u>). In the EAPphysico model where physicochemical properties of peptides were used as input features, an accuracy of 65% with a MCC value of 0.31 was observed (<u>Table 3</u>). In the EAPcompo model where amino acid composition features were used, the predictive accuracy and the MCC value are slightly higher. When the sequence-based statistical function scores were used as input in the EAPscoring model, a remarkable accuracy of 92% was achieved with a MCC value of 0.84. Thus the sequence-based statistical scores developed in the present research are predominantly superior to the conventional physicochemical properties or amino acid decomposition features in identifying active peptides derived from enveloped proteins.

Table 3.	Performance of AVP	pred and EAP	pred models on i	ndepende	nt test set V ^{26p+2}	5n

Model	Features	Sensitivity	Specificity	Accuracy	мсс
AVPpred	AVPmotif	100	50.98	51.92	0.14
	AVPphysico	72.22	61.76	65.38	0.32
	AVPcompo	63.16	57.58	59.62	0.20
	AVPalign	92.86	65.79	73.08	0.52
EAPpred	EAPphysico	68.18	63.33	65.38	0.31
	EAPcompo	72.41	78.26	75	0.5
	EAPscoring	92.3	92.3	92.3	0.84

doi:10.1371/journal.pone.0144171.t003

Comparison of the predictive efficiency of the AVP and EAP Models

AVPpred is a web server for prediction of the activities of general antiviral peptides (AVPs) based on a number of experimentally validated positive and negative data sets [43]. The peptide inhibitors employed in AVPpred target a variety of biological targets involved in virus infection. In contrast, the self-derived peptides of enveloped proteins being studied in the present research competitively bind to E proteins so as to mediate the virus fusion process. Because the self-derived peptides share similar mechanism of action, it is feasible to retrieve common features from them to build predictive SVM models. In order to evaluate the performance in predicting peptide inhibitors of the enveloped virus, we compared the AVPpred models with our EAPpred models using an independent 26p+26n dataset as test set. The results are shown in Table 3.

Four different features were employed in the AVPpred models, namely conserved motif search using MEME/MAST, amino acid composition, sequence alignment using BLAST and physicochemical parameters including secondary structure, charge, size, hydrophobicity and amphiphilic character [43]. When the AVPmotif model was used to predict the activities of the self-derived peptide inhibitors, it performed rather poorly with accuracy of 52% and MCC of 0.14. This is not surprising because AVPmotif was developed based on 20 general antiviral peptide motifs. However, the self-derived peptide inhibitors may not share a conserved motif with the general antiviral peptides since the latter interact with various biological targets with different mechanisms of action. In the AVPalign model, the peptide sequences were classified into active and non-active databases and the query peptide sequences were matched against the active and non-active databases using the BLAST program. Compared with AVPcompo and AVPphysico, AVPalign performed better with a predictive accuracy of 73% and MCC value of 0.52. Fusion mechanism is highly conserved among related viruses and entry of viruses into host cells has been inhibited by peptides derived from various regions of envelope glycoproteins [59]. Self-derived peptides would inhibit interactions of their original domain by mimicking its mode of binding to partner proteins [4]. Because similar sequences are often associated with similar structure and function, the sequence-based property AVPalign would account for the activities of the self-derived peptide inhibitors which regulate the virus fusion by mimicking the binding to E proteins.

In the AVPphysico model, 25 best performing physicochemical properties were selected out of the 544 properties to build the SVM model [43]. Antiviral peptide inhibitors are generally amphiphilic [60] and the activities of peptide entry inhibitors are dependent on their interfacial hydrophobicity [58]. Therefore we only employed five physicochemical properties reflecting hydrophobicity, solvent accessibility and secondary structure features as SVM input features. It was demonstrated that the accuracy and MCC of EAPphysico is comparable to that of AVP-physico model, indicating the five properties used in current modeling building are critical for their activities.

The MCC value of the AVPcompo models is 0.20, indicating that the antiviral activities of the peptides are related to amino acid composition. When the amino acid composition was used as input, the predictive accuracy of the EAPcompo model was higher than that of the AVPcompo model, indicating the peptide inhibitors of E proteins employed in the training set is sufficient to represent the contribution of amino acid composition to their inhibitive activities. In the EAPcompo model, the preference of the amino acid composition was ranked as: P, R, Q, D, F, W, E, L, T, I, N, H, Y, C, A, S, M, V, K, G (Fig 1). The role of arginine-arginine pairing and its contribution to protein-protein interactions has been investigated by computational approaches [61]. The higher abundance of R at protein-protein interfaces compared to K may be attributed to the formation of cation- π -interactions and the greater capacity of the





doi:10.1371/journal.pone.0144171.g001

guanidinium group in R to form hydrogen bonds (compared to K) [62–64]. Furthermore, it was suggested that the interface regions are enriched in aliphatic (L, V, I, M) and aromatic (H, F, Y, W) residues and depleted in charged residues (D, E, K) with the exception of arginine [62, 65–69]. This is in agreement with our amino acid composition analysis, where higher population of aliphatic Leu residue as well as aromatic residues Trp and Phe was observed, whereas positively charged Lys was hardly observed. The predominant occurrence of proline and glutamine residues is characteristic for the unique protein-protein interactions for E proteins. e.g. a conserved proline-rich motif was suggested to be engaged in monomer-monomer interactions in Dengue E proteins [70]. A conserved glutamine-rich layer is involved in the extensive H-bond network in HIV-1 gp41 E proteins [71]. Thus the preference of the amino acid composition identified from the EAPcompo model is generally in accordance with the predominant residues involved in protein-protein interactions, manifesting the amino acid composition of the self- derived peptide inhibitors are closely related to their potential activities in mediating the protein-protein interactions in the virus fusion process.

Because the antiviral activities of peptides are dependent on amino acid composition, we presume amino acid composition discriminated by the propensity of their activities would be an intrinsic feature in the self-derived peptide inhibitors which share a common mechanism of action. When statistical function scores were employed in the SVM model (EAPscoring), a remarkable predictive accuracy of 92% with an ideal MCC value of 0.84 was achieved, significantly better than any AVP models. The logarithm form of the discriminatory function (Eq 1) can be deemed as the pseudo energy of the system. In our previous study, we suggested that the stability of proteins is related to their *in situ* binding potential to the partner regions [72]. The prominent performance of EAPscoring model indicates the sequence-based stability feature of self-derived peptides may reflect their potential of binding to E proteins so as to regulate the virus entry process.

Conclusions

We developed three SVM models using physicochemical properties, amino acid composition and statistical discriminative function as input features. The prediction accuracy and the MCC value of the EAPphysico model where five physicochemical properties were employed are comparable with the previous AVPphysico model where 25 physicochemical properties were used. The AVPcompo and EAPcompo models demonstrated that the activities of antiviral peptides are dependent on amino acid composition. A sequence-based scoring function was developed for the self-derived peptide inhibitors of E proteins. The outperformance of the EAPscoring models supports our hypothesis that an intrinsic feature, represented by the propensity of each amino acid for being active in self-derived peptides, is responsible for the activities of the peptides to regulate virus fusion by mimicking the binding to their accessory proteins. The sequence-based statistical scoring function would be useful in development of novel antiviral therapies to target the initial step of viral infection.

Supporting Information

S1 File. Parameters optimization by Grid-research combined with 5-fold cross validation. x-axis is log2^g, y is log2^c and z-axis represents accuracy(%) (Figure A) Parameters Optimization for EAPphysico model. (Figure B) Parameters Optimization for EAPcompo model. (Figure C) Parameters Optimization for EAPscoring model. (DOCX)

Acknowledgments

The authors are grateful for the computing resources from QUB high performance computing Centre. The authors declare no conflict of interest.

Author Contributions

Conceived and designed the experiments: YX MH. Performed the experiments: YX MH JZ GH SY. Analyzed the data: YX MH JZ GH SY NR RO XT. Wrote the paper: MH.

References

- 1. Teissier E, Penin F, Pécheur EI. Targeting cell entry of enveloped viruses as an antiviral strategy. Molecules. 2011; 16: 221–250.
- 2. Backovic M, Jardetzky TS. Class III viral membrane fusion proteins. Curr Opin Struct Biol. 2009; 19: 189–196. doi: 10.1016/j.sbi.2009.02.012 PMID: 19356922
- Kielian M, Rey FA. Virus membrane-fusion proteins: more than one way to make a hairpin. Nat Rev Microbiol. 2006; 4: 67–76. PMID: <u>16357862</u>
- London N, Raveh B, Movshovitz-Attias D, Schueler-Furman O. Can self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions? Proteins. 2010; 78: 3140–3149. doi: <u>10.</u> <u>1002/prot.22785</u> PMID: <u>20607702</u>
- Qureshi NM, Coy DH, Garry RF, Henderson LA. Characterization of a putative cellular receptor for HIV-1 transmembrane glycoprotein using synthetic peptides. Aids. 1990; 4: 553–558. PMID: <u>1974767</u>
- Wild C, Dubay JW, Greenwell T, Baird T Jr, Oas TG, McDanal C, et al. Propensity for a leucine zipperlike domain of human immunodeficiency virus type 1 gp41 to form oligomers correlates with a role in virus-induced fusion rather than assembly of the glycoprotein complex. Proc Natl Acad Sci USA. 1994; 91: 12676–12680. PMID: <u>7809100</u>
- Wild C, Greenwell T, Matthews T. A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virusmediated cell-cell fusion. AIDS Res Hum Retroviruses.1993; 9: 1051–1053. PMID: 8312047
- Liu S, Jing W, Cheung B, Lu H, Sun J, Yan X, et al. HIV gp41 C-terminal heptad repeat contains multifunctional domains. Relation to mechanisms of action of anti-HIV peptides. J Biol Chem. 2007; 282: 9612–20. PMID: <u>17276993</u>
- Egelhofer M, Brandenburg G, Martinius H, Schult-Dietrich P, Melikyan G, Kunert R, et al. Inhibition of human immunodeficiency virus type 1 entry in cells expressing gp41-derived peptides. J Virol. 2004; 78:568–75. PMID: <u>14694088</u>
- Koedel Y, Eissmann K, Wend H, Fleckenstein B, Reil H. Peptides derived from a distinct region of GB virus C glycoprotein E2 mediate strain-specific HIV-1 entry inhibition. J Virol. 2011; 85: 7037–7047. doi: <u>10.1128/JVI.02366-10</u> PMID: <u>21543477</u>
- Sainz B Jr, Mossel EC, Gallaher WR, Wimley WC, Peters CJ, Wilson RB, et al. Inhibition of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) infectivity by peptides analogous to the viral spike protein. Virus Res. 2006; 120: 146–55. PMID: <u>16616792</u>

- Zheng BJ, Guan Y, Hez ML, Sun H, Du L, Zheng Y, et al. Synthetic peptides outside the spike protein heptad repeat regions as potent inhibitors of SARS-associated coronavirus. Antivir Ther. 2005; 10: 393–403. PMID: <u>15918330</u>
- Yuan K, Yi L, Chen J, Qu X, Qing T, Rao X, et al. Suppression of SARS-CoV entry by peptides corresponding to heptad regions on spike glycoprotein. Biochem Biophys Res Commun. 2004; 319:746–52. PMID: <u>15184046</u>
- Spence J. Design and Characterization of Glycoprotein-derived Peptide Inhibitors of Arena Virus Infection. PhD. Dissertation, Tulane University. 2013.
- Autoimmune Technologies, Safety, Tolerability, and PK of Escalating Doses of Flufirvitide-3 Dry Powder for Inhalation in Healthy Subjects. 18 Nov 2013. Available: <u>http://clinicaltrials.gov/ct2/show/</u> NCT01990846?term=Flufirvitide&rank=2%202013.
- Hrobowski YM, Garry RF, Michael SF. Peptide inhibitors of dengue virus and West Nile virus infectivity. Virol J. 2005; 2: 1–10.
- Costin JM, Jenwitheesuk E, Lok SM, Hunsperger E, Conrads KA, Fontaine KA, et al. Structural optimization and De Novo design of dengue virus entry inhibitory peptides. PLOS Neglected Tropical Diseases. 2010; 4: 1–11.
- Bai F, Town T, Pradhan D, Cox J, Ashish Ledizet M, Anderson JF, et al. Antiviral Peptides Targeting the West Nile Virus Envelope Protein. J Virol. 2007; 81: 2047–2055. PMID: <u>17151121</u>
- Schmidt AG, Yang PL, Harrison SC. Peptide inhibitors of dengue virus entry target a late-stage fusion intermediate. PLoS Pathog. 2010; 6: e1000851. doi: 10.1371/journal.ppat.1000851 PMID: 20386713
- Alhoot M, Rathinam A, Wang S, Manikam R, Sekaran S. Inhibition of Dengue Virus Entry into Target Cells Using Synthetic Antiviral Peptides. International Journal of Medical Sciences. 2013; 10: 719– 729. doi: 10.7150/ijms.5037 PMID: 23630436
- Li C, Zhang L, Sun M, Li P, Huang L, Wei J, et al. Inhibition of Japanese encephalitis virus entry into the cells by the envelope glycoprotein domain III (EDIII) and the loop3 peptide derived from EDIII. Antiviral Research. 2012; 94: 179–183. doi: 10.1016/j.antiviral.2012.03.002 PMID: 22465300
- Koehler J, Smith J, Ripoll D, Spik K, Taylor S, Badger C, et al. A Fusion-Inhibiting Peptide against Rift Valley Fever Virus Inhibits Multiple, Diverse Viruses. PLOS Neglected Tropical Diseases. 2013; 7: 1– 11
- Liu R, Tewari M, Kong R, Zhang R, Ingravallo P, Ralston R. A peptide derived from hepatitis C virus E2 envelope protein inhibits a post-binding step in HCV entry. Antiviral Res. 2010; 86: 172–179. doi: <u>10.</u> <u>1016/j.antiviral.2010.02.316</u> PMID: <u>20156485</u>
- 24. Sabahi A. Early Events in Hepatitis C Virus Infection: An Interplay of Viral Entry, Decay and Density. PhD. Dissertation, Tulane University. 2008.
- Si Y, Liu S, Liu X, Jacobs J, Cheng M, Niu Y, et al. Human Claudin-1–Derived Peptide Inhibits Hepatitis C Virus Entry. HEPATOLOGY. 2012; 56: 507–515. doi: <u>10.1002/hep.25685</u> PMID: <u>22378192</u>
- Galdiero S, Falanga A, Vitiello M, D'Isanto M, Cantisani M, Kampanaraki A, et al. Peptides containing membraneinteracting motifs inhibit herpes simplex virus type 1 infectivity. Peptides. 2008; 29: 1461– 1471. doi: 10.1016/j.peptides.2008.04.022 PMID: 18572274
- Galdiero S, Falanga A, Vitiello M, D'Isanto M, Collins C, Orrei V, et al. Evidence for a role of the membrane- proximal region of herpes simplex virus type 1 glycoprotein H in membrane fusion and virus inhibition. Chembiochem. 2007; 8:885–895. PMID: <u>17458915</u>
- Galdiero S, Vitiello M D'Isanto M, Falanga A, Cantisani M, Browne H, et al. The identification and characterization of fusogenic domains in herpes virus glycoprotein B molecules. Chembiochem. 2008; 9: 758–67. doi: 10.1002/cbic.200700457 PMID: <u>18311743</u>
- Galdiero S, Vitiello M, D'Isanto M, Falanga A, Collins C, Raieta K, et al. Analysis of synthetic peptides from heptad-repeat domains of herpes simplex virus type 1 glycoproteins H and B. J Gen Virol. 2006; 87:1085–97. PMID: 16603508
- Cantisani M, Falanga A, Incoronato N, Russo L, De Simone A, Morelli G, et al. Conformational modifications of gB from herpes simplex virus type 1 analyzed by synthetic peptides. J Med Chem. 2013; 56:8366–76. doi: 10.1021/jm400771k PMID: 24160917
- Akkarawongsa R, Pocaro NE, Case G, Kolb AW, Brandt CR. Multiple peptides homologous to herpes simplex virus type 1 glycoprotein B inhibit viral infection. Antimicrobial Agents and Chemotherapy. 2009; 53: 987–996. doi: <u>10.1128/AAC.00793-08</u> PMID: <u>19104014</u>
- **32.** Melnik LI, Garry RF, Morris CA. Peptide inhibition of human cytomegalovirus infection. Virol J. 2011; 8: 1–11.
- Torrent M, Nogues VM, Boix E. A theoretical approach to spot active regions in antimicrobial proteins. BMC Bioinformatics. 2009; 10: 1–9.

- Fjell CD, Jenssen H, Cheung WA, Hancock RE, Cherkasov A. Optimization of antibacterial peptides by genetic algorithms and cheminformatics. Chem Biol Drug Des. 2011; 77: 48–56. doi: <u>10.1111/j.1747-</u> <u>0285.2010.01044.x</u> PMID: <u>20942839</u>
- Lata S, Mishra NK, Raghava GP. AntiBP2: improved version of antibacterial peptide prediction. BMC Bioinformatics. 2010; 11(Suppl 1): S19. doi: <u>10.1186/1471-2105-11-S1-S19</u> PMID: <u>20122190</u>
- Cherkasov A, Jankovic B. Application of 'inductive' QSAR descriptors forquantification of antibacterial activity of cationic polypeptides. Molecules. 2004; 9: 1034–1052. PMID: <u>18007503</u>
- Frecer V. QSAR analysis of antimicrobial and haemolytic effects of cyclic cationic antimicrobial peptides derived from protegrin-1. Bioorg Med Chem. 2006; 14: 6065–6074. PMID: <u>16714114</u>
- Taboureau O, Olsen OH, Nielsen JD, Raventos D, Mygind PH, Kristensen HH. Design of novispirin antimicrobial peptides by quantitative structure-activity relationship. Chem Biol Drug Des. 2006; 68: 48–57. PMID: 16923026
- Jenssen H, Fjell CD, Cherkasov A, Hancock RE. QSAR modeling and computer-aided design of antimicrobial peptides. J Pept Sci. 2008; 14: 110–114. PMID: <u>17847019</u>
- Fjell CD, Jenssen H, Hilpert K, Cheung WA, Pante N, Hancock RE, et al. Identification of novel antibacterial peptides by chemoinformatics and machine learning. J Med Chem. 2009; 52: 2006–2015. doi: <u>10.1021/jm8015365</u> PMID: <u>19296598</u>
- Frecer V, Ho B, Ding JL. De novo design of potent antimicrobial peptides. Antimicrob Agents Chemother. 2004; 48: 3349–3357. PMID: <u>15328096</u>
- Jenssen H, Lejon T, Hilpert K, Fjell CD, Cherkasov A, Hancock RE. Evaluating different descriptors for model design of antimicrobial peptides with enhanced activity toward P. aeruginosa. Chem Biol Drug Des. 2007; 70: 134–142. PMID: <u>17683374</u>
- Thakur N, Qureshi A, Kumar M. AVPpred: collection and prediction of highly effective antiviral peptides. Nucleic Acids Res. 2012; 40 (Web Server issue): W199–204. doi: <u>10.1093/nar/gks450</u> PMID: 22638580
- White JM, Delos SE, Brecher M, Schornberg K. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. Crit Rev Biochem Mol Biol. 2008; 43: 189–219. doi: 10.1080/10409230802058320 PMID: 18568847
- Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 1982; 157: 105–132. PMID: <u>7108955</u>
- Cheng J, Randall A, Sweredoski M, Baldi P. SCRATCH: a Protein Structure and Structural Feature Prediction Server. Nucleic Acids Research. 2005; 33 (web server issue): W72–76. PMID: <u>15980571</u>
- Samudrala R, Moult J. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J Mol Biol. 1998; 275: 893–914.
- 48. Cortes C, Vapnik V. Support-Vector Networks. Machine Learning. 1995; 20: 1–31.
- Fan R-E, Chen P-H, Lin C-J. Working set selection using second order information for training SVM. Journal of Machine Learning Research. 2005; 6: 1889–1918,
- Hsu C-W, Chang C-C, Lin C-J. A Practical Guide to Support Vector Classification. Initial version: 2003 Last updated: 15 April 2010; 1–16. Available: <u>http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf</u>
- Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta. 1975; 405: 442–451. PMID: <u>1180967</u>
- Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci. 2002; 11: 2714–2726. PMID: 12381853
- Lu H, Skolnick J. A distance-dependent atomic knowledge-based potential for improved protein structure selection. Proteins. 2001; 44: 223–32. PMID: 11455595
- Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006; 15: 2507–2524. PMID: 17075131
- Moult J. Comparison of database potentials and molecular mechanics force fields. Curr Opin Struct Biol. 1997; 7: 194–199. PMID: 9094335
- Ngan S-C, Hung L-H, Liu T, Samudrala R. Scoring functions for de novo protein structure prediction revisited. Methods in Molecular Biology. 2007; 413: 243–282.
- 57. Galdiero S, Falanga A, Tarallo R, Russo L, Galdiero E, Cantisani M, et al. Peptide inhibitors against herpes simplex virus infections. J Pep Sci. 2013; 19: 148–158.
- Badani H, Garry RF, Wimley WC. Peptide entry inhibitors of enveloped viruses: the importance of interfacial hydrophobicity. Biochim Biophys Acta. 2014; 1838: 2180–2197. doi: <u>10.1016/j.bbamem.2014</u>. 04.015 PMID: <u>24780375</u>

- Pelay-Gimeno M, Glas A, Koch O, Grossmann TN. Structure-Based Design of Inhibitors of Protein– Protein Interactions: Mimicking Peptide Binding Epitopes. Angew Chem Int Ed Engl. 2015; 54: 8896– 8927. doi: <u>10.1002/anie.201412070</u> PMID: <u>26119925</u>
- Vigant F, Santos NC, Lee B. Broad-spectrum antivirals against viral fusion. Nat Rev Microbiol. 2015; 13: 426–37. doi: <u>10.1038/nrmicro3475</u> PMID: <u>26075364</u>
- Vondrášek J, Mason PE, Heyda J, Collins KD, Jungwirth P. The molecular origin of like-charge arginine —Arginine pairing in water. Journal of Physical Chemistry B. 2009; 113: 9041–9045.
- Glaser F, Steinberg DM, Vakser IA, Ben-Tal N. Residue frequencies and pairing preferences at proteinprotein interfaces. Proteins, 2001; 43: 89–102. PMID: <u>11276079</u>
- 63. Bahadur RP, Chakrabarti P, Rodier F, Janin JA. Dissection of Specific and Non-specific Protein-Protein Interfaces. Journal of Molecular Biology. 2004; 336: 943–955. PMID: <u>15095871</u>
- **64.** Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. Proteins: Structure, Function and Genetics. 2002; 47: 334–343.
- Jones S, Thornton JM. Principles of protein-protein interactions. Proc Natl Acad Sci USA.1996, 93: 13–20. PMID: 8552589
- Tsai C-J, Lin SL, Wolfson HJ, Nussinov R. Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. Protein Science. 1997; 6: 53–64. PMID: <u>9007976</u>
- Conte LL, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. Journal of Molecular Biology. 1999; 285: 2177–2198. PMID: <u>9925793</u>
- Janin J, Séraphin B. Genome-wide studies of protein-protein interaction. Current Opinion in Structural Biology. 2003; 13: 383–388. PMID: <u>12831891</u>
- Bahadur RP, Zacharias M. The interface of protein-protein complexes: Analysis of contacts and prediction of interactions. Cellular and Molecular Life Sciences. 2008; 65: 1059–1072. PMID: <u>18080088</u>
- Gadkari RA, Srinivasan N. Prediction of protein-protein interactions in dengue virus coat proteins guided by low resolution cryoEM structures. BMC Struct Biol. 2010; 10:17. doi: <u>10.1186/1472-6807-10-17</u> PMID: <u>20550721</u>
- Suntoke TR, Chan DC. The fusion activity of HIV-1 gp41 depends on interhelical interactions. J Biol Chem. 2005; 280:19852–19857. PMID: <u>15772068</u>
- Xu Y, Rahman N ABD, Othman RB, Hu P, Huang M. Computational Identification of Self-inhibitory Peptides from Envelope Proteins. Proteins: Structure, Function and Bioinformatics. 2012; 80: 2154–2168.