AVP-IC_{50}Pred: Multiple machine learning techniques based prediction of peptide antiviral activity in terms of half maximal inhibitory concentration (IC_{50})

Abid Qureshi[1], Himani Tandon[1] and Manoj Kumar[1,*]

[1]Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research, Sector 39-A, Chandigarh-160036, India

* To whom correspondence should be addressed: manojk@imtech.res.in

Supporting Data

Figure S1. Two sample logos of eight N and C- terminal residues of 97 highly effective peptides (IC_{50} < 1 µM) and an equal number of least effective peptides (IC_{50} > 100 µM). Amino acids are colored as per their (i) Charge, (ii) Hydrophobicity, (iii) Surface exposure, (iv) Flexibility and (v) Disorder

Figure S2. Box plots of 15 best performing physicochemical properties using SVM (see Table S3a) of 97 highly effective peptides (IC_{50}<1µM) and an equal number of least effective peptides (IC_{50}>100µM)

Figure S3. AVP-IC_{50}Pred workflow for model development

Figure S4. AVP-IC_{50}Pred submit form

Table S1. Training/testing dataset used in the development of AVP-IC_{50}Pred prediction models.

Table S2. Validation dataset used in the development of AVP-IC_{50}Pred prediction models.

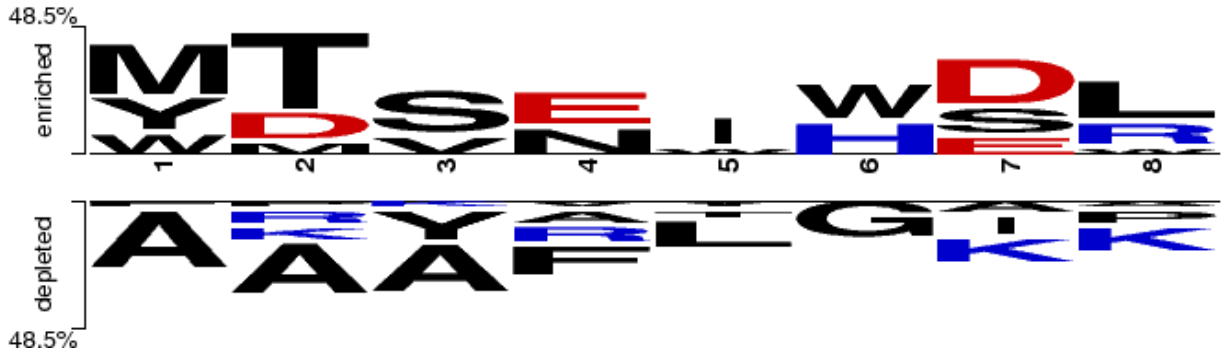
Table S3. 15 best performing physicochemical properties (a-d)

Table S4. SVM and RF parameters used to develop prediction models.

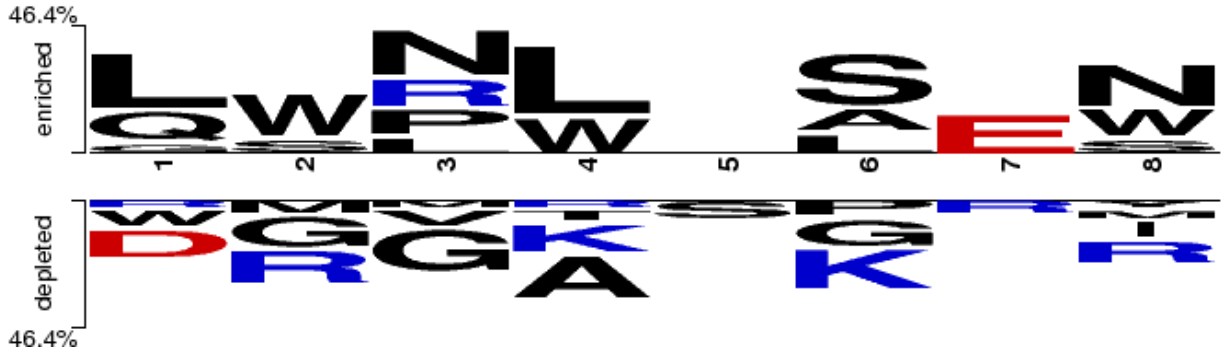
Table S5. Performance of SVM models based on training and validation data composed on randomized instances.

(i) **Charge:** Positively charged residues (K, R, H) are colored blue, and negatively charged residues (D, E) are colored red; all neutral residues are colored black.

a) N-terminal



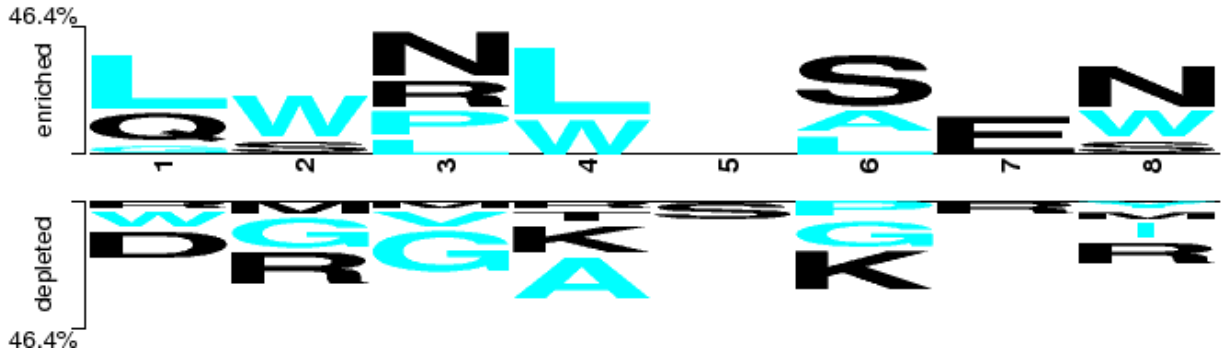
b) C-terminal



(ii) **Hydrophobicity:** Hydrophobic residues (A, F, G, I, L, P, V, W, Y) are cyan colored, while the remaining hydrophilic residues are colored black. This classification was based on (Eisenberg, 1984).

a) N-terminal

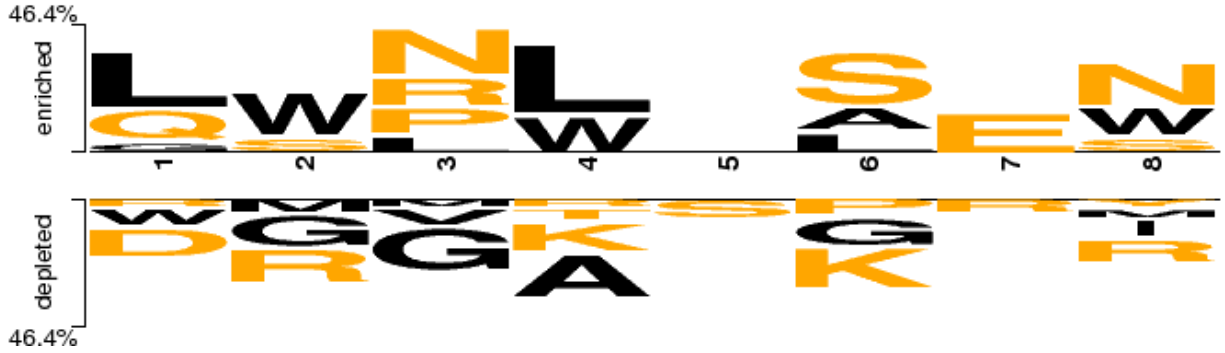
b) C-terminal



(iii) **Surface exposure:** Surface exposed residues (D, E, H, K, N, P, Q, R, S, T, Y) are colored orange, and buried residues (A, C, F, G, I, L, M, V, W) are colored black. This classification was based on (Janin, 1979).

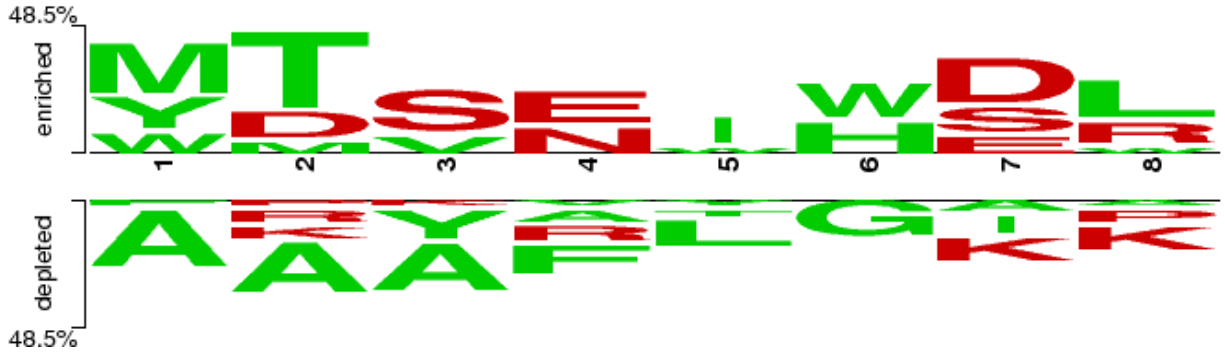
a) N-terminal

b) C-terminal

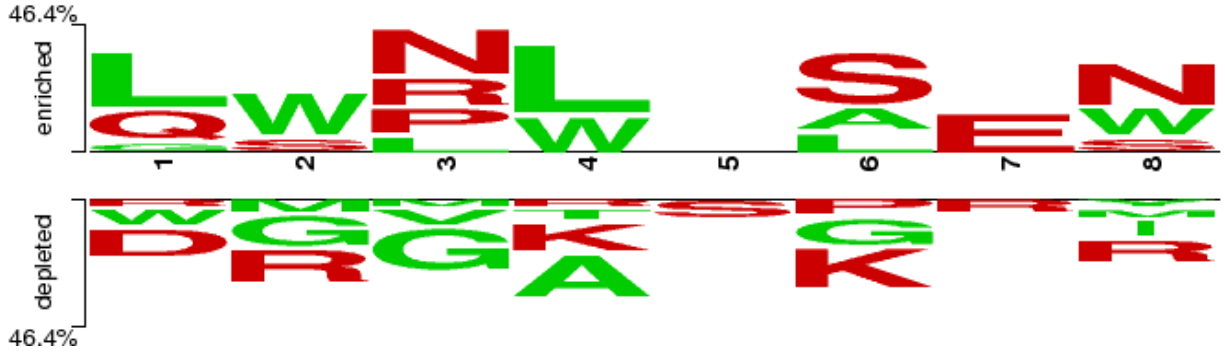


(iv**) Flexibility:** High flexibility residues (D, E, K, N, P, Q, R, S) are colored red, whereas low flexibility residues (A, C, F, G, H, I, L, M, T, V, W, Y) are colored green. This classification was based on (Vihinen et al., 1994).

a) N-terminal

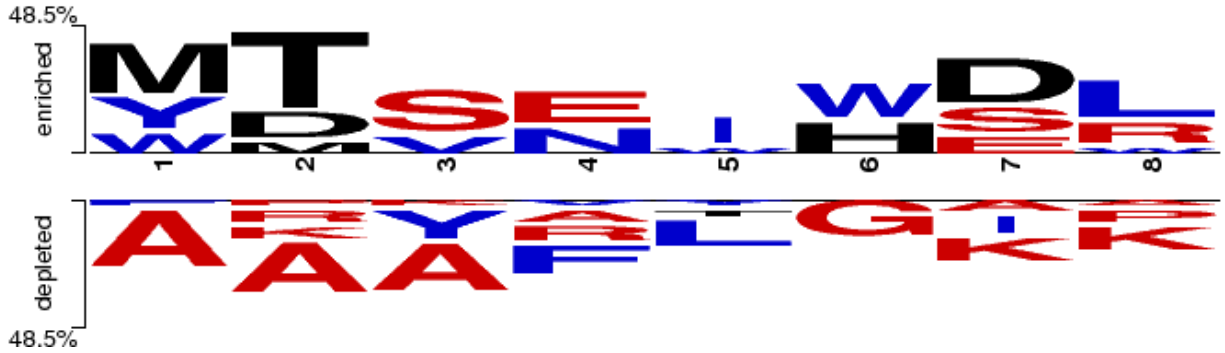


b) C-terminal



(v) **Disorder:** Disorder-promoting residues (A, R, S, Q, E, G, K, P) are colored red, order-promoting residues (N, C, I, L, F, W, Y, V) are colored blue, and disorder-order neutral residues (D, H, M, T) are colored black. This classification was based on (Dunker et al., 2001).

a) N-terminal



b) C-terminal

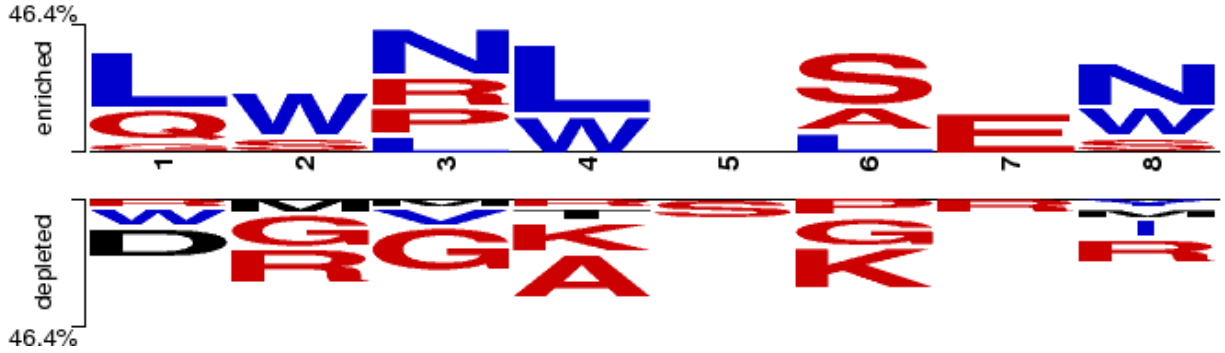
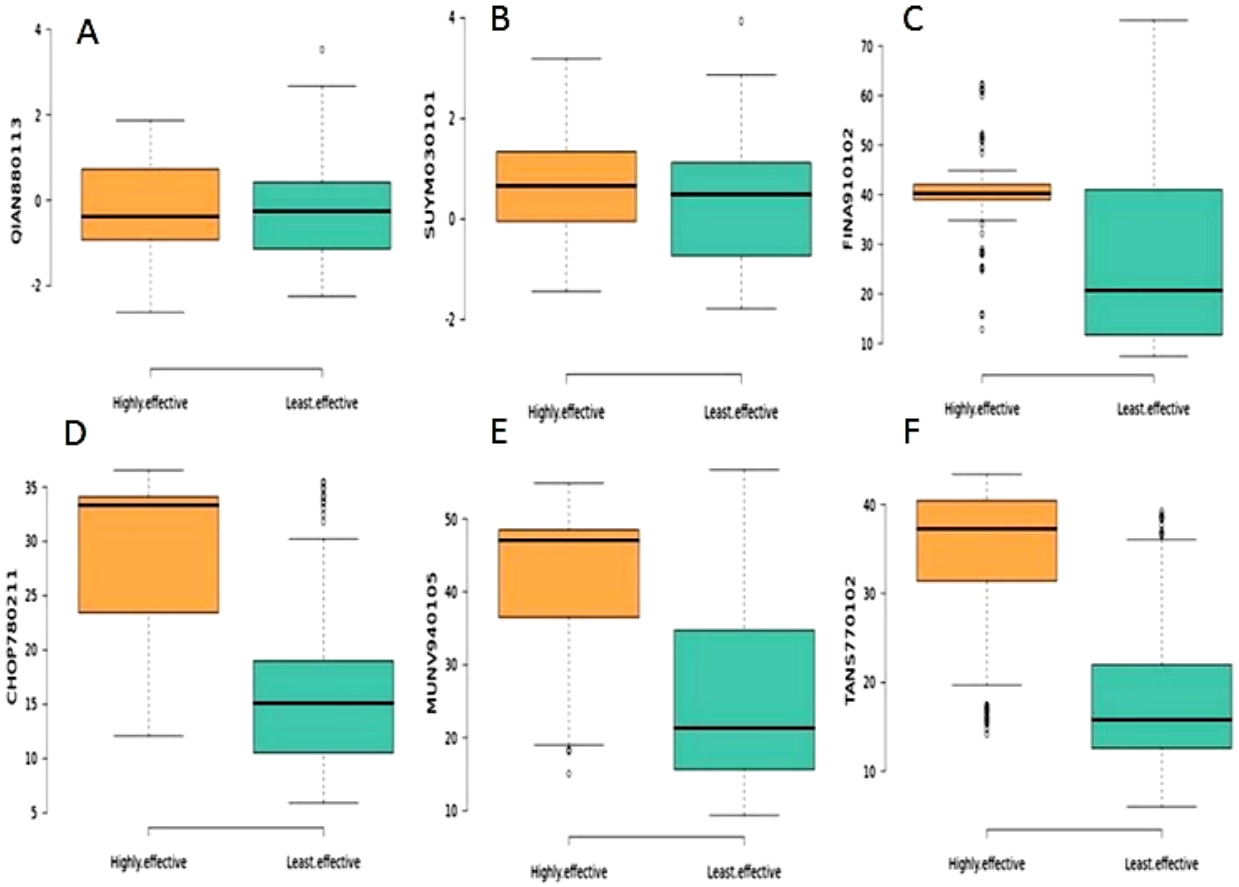


Figure S1. Two sample logos of eight N and C- terminal residues of 97 highly effective peptides (IC_{50}<1µM) and an equal number of least effective peptides (IC_{50}>100µM). Amino acids are colored as per their (i) Charge, (ii) Hydrophobicity, (iii) Surface exposure, (iv) Flexibility and (v) Disorder

(i) (A) QIAN880113, (B) SUYM030101, (C) FINA910102, (D) CHOP780211, (E) MUNV940105, (F) TANS770102



(ii) (G) MUNV940104, (H) PALJ810114, (I) PALJ810115, (J) QIAN880110, (K) AURR980103, (L) ISOY800104, (M) MUNV940102, (N) RACS820110, (O) AURR980105

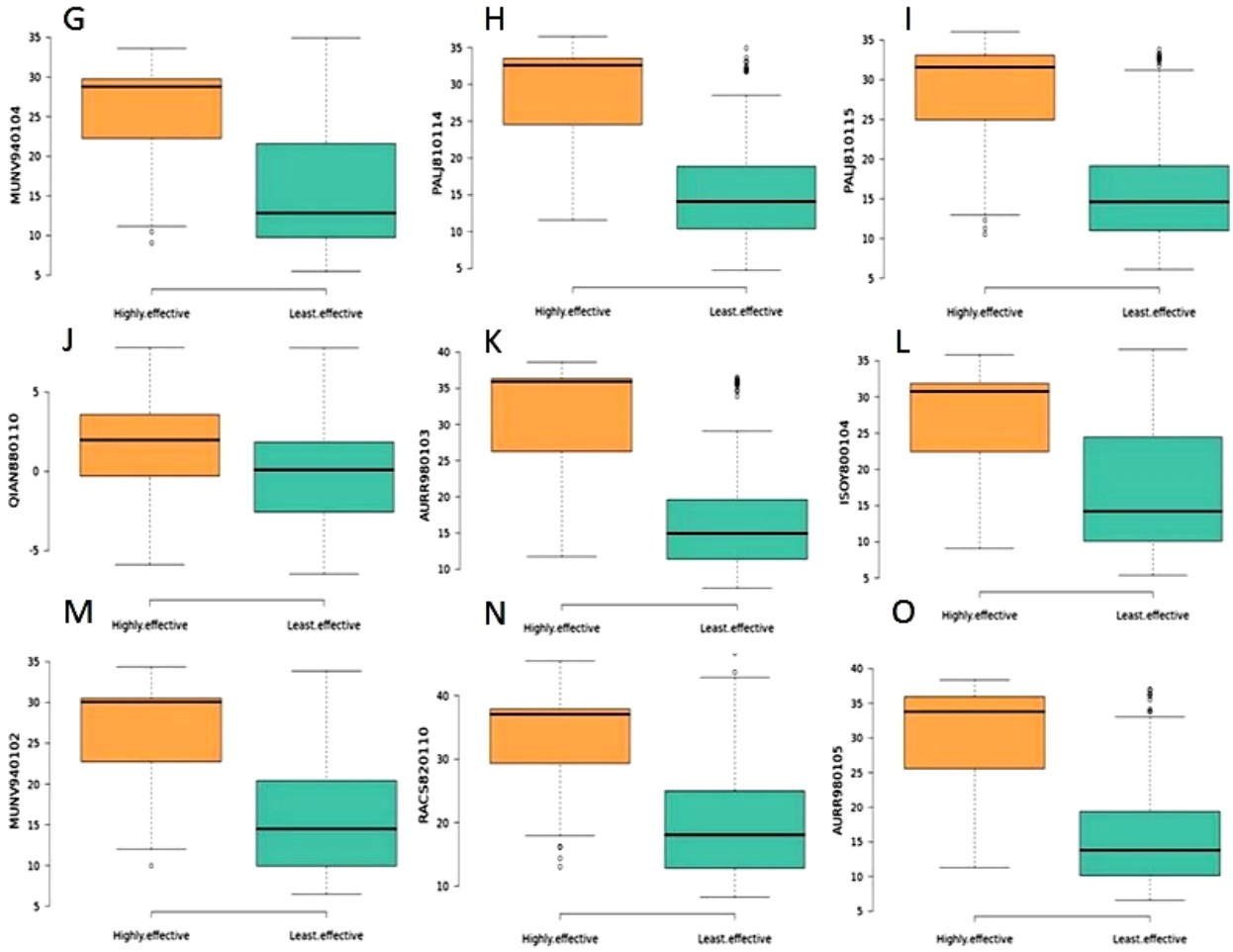


Figure S2. Box plots of 15 best performing physicochemical properties using SVM (see Table S3a) of 97 highly effective peptides (IC_{50}<1µM) and an equal number of least effective peptides (IC_{50}>100µM).

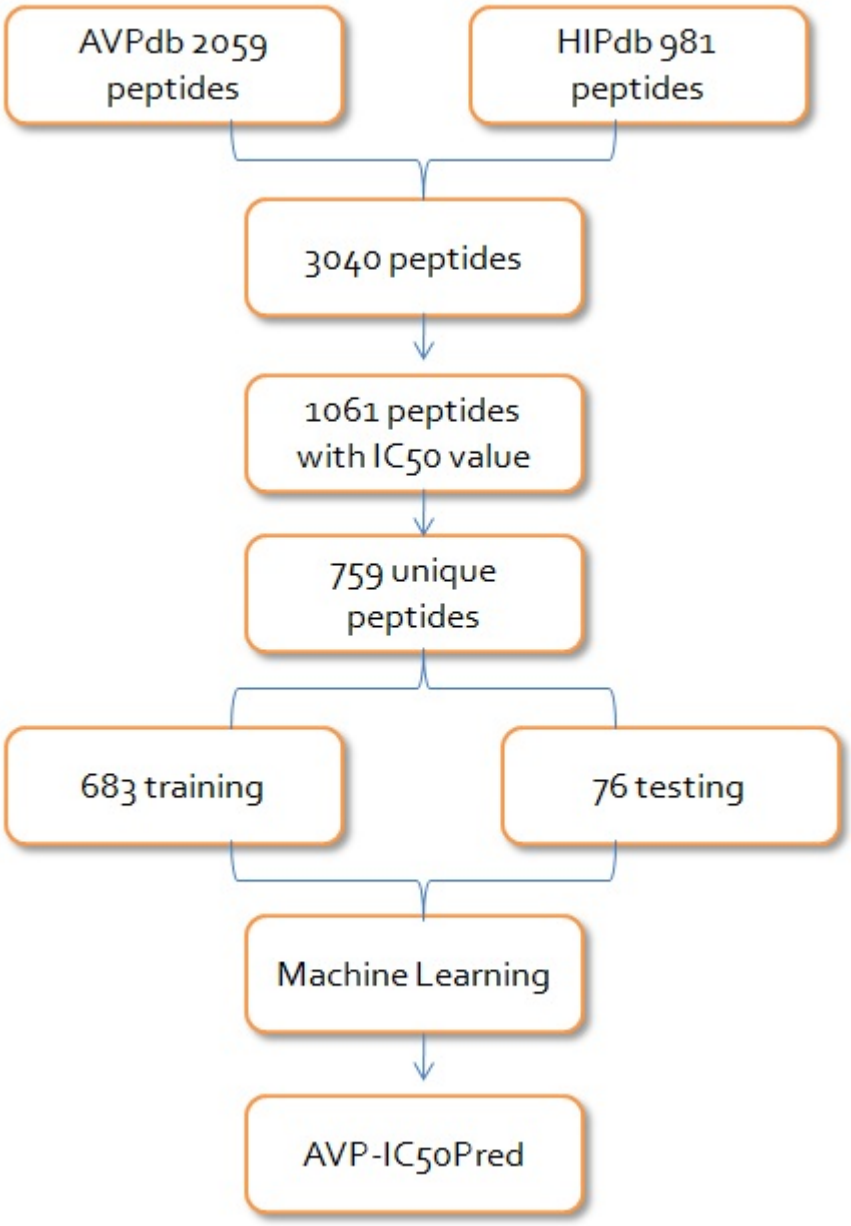


Figure S3. AVP-IC_{50}Pred workflow for model development

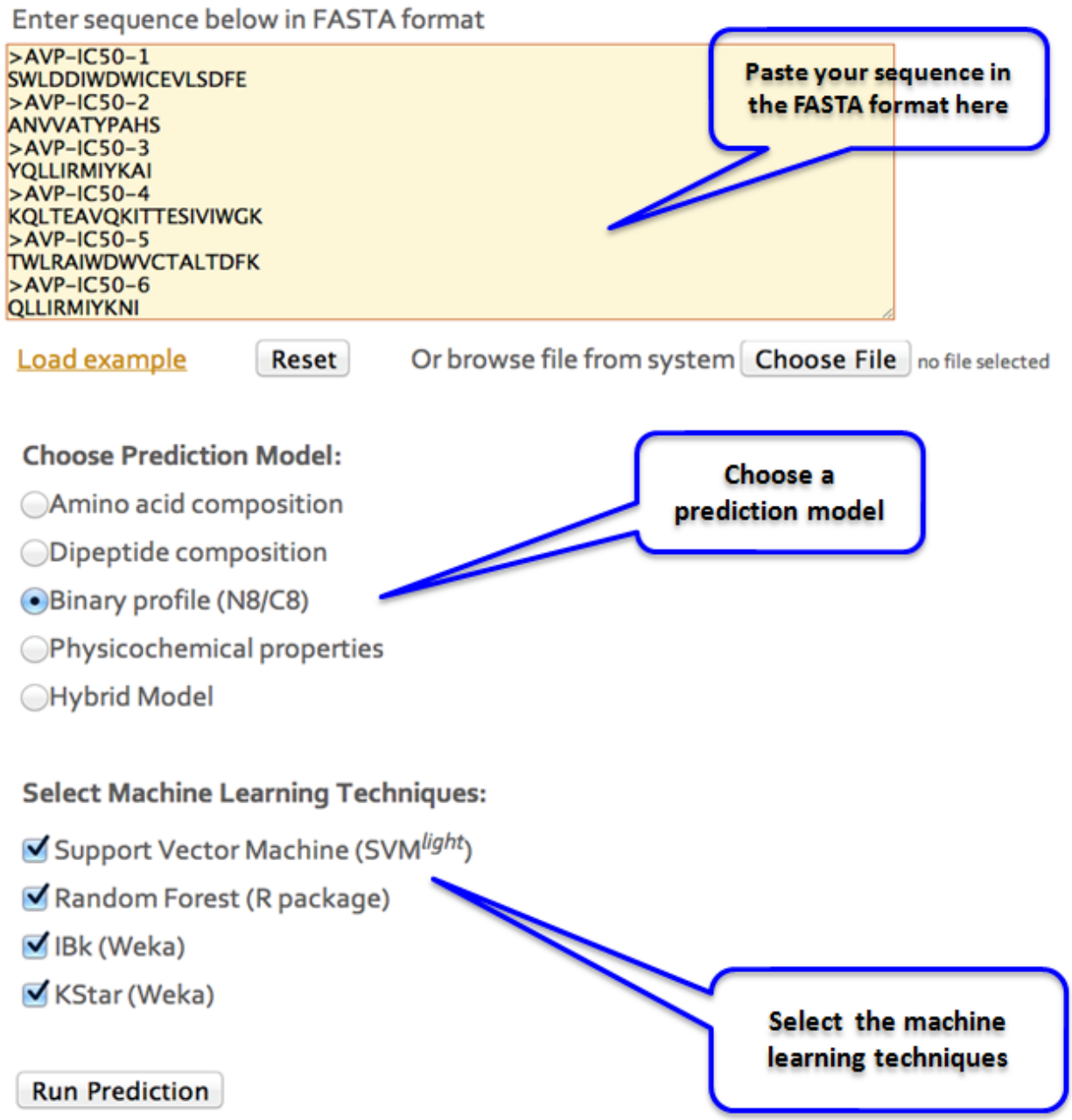


Figure S4. AVP-IC_{50}Pred submit form.

Table S1. Training/testing dataset used in the development of AVP-IC_{50}Pred prediction models.

AVPdb/HIPdb_ID	Sequence	Reference	IC50 (microM)	Virus
>**AVP0003**	ECRSTSYAGAVVNDL	3040743	42	HSV 1
>**AVP0004**	STSYAGAVVNDL	3040743	29	HSV 1
>**AVP0006**	AGAVVNDL	3040743	283	HSV 1
>**AVP0012**	AAGAVVNDL	3040743	280	HSV 1
>**AVP0014**	ALLGRMKG	7892246	10	HBV
>**AVP0016**	GQPEEGAPCQVVLQGA	8380075	3.5	HSV 1
>**AVP0017**	RGILIHNTIFGEQVF	8380075	7.5	HSV 1
>**AVP0018**	YRWRGPTAAFLSLV	8380075	9	HSV 1
>**AVP0019**	SSSTSTQVQILSNAL	8380075	0.8	HSV 1
>**AVP0020**	RRLQVGGGTLKFFLT	8380075	12	HSV 1
>**AVP0021**	FLDSKAELEKARKILSEVGRWY	8521809	1	SeV
>**AVP0088**	PDAVYLHRIDLGPPISLERLDVGTNLGNAIAKLED	8700906	100	MV
>**AVP0089**	DAVYLHRIDLGPPISLERLDVGTNLQNAIAKLEDA	8700906	100	MV
>**AVP0090**	AVYLHRIDLGPPISLERLDVGTNLQNAIAKLEDAK	8700906	100	MV
>**AVP0091**	VYLHRIDLGPPISLERLDVGTNLQNAIAKLEDAKE	8700906	85.3	MV
>**AVP0092**	YLHRIDLGPPISLERLDVGTNLGNAIAKLEDAKEL	8700906	90.7	MV
>**AVP0094**	HRIDLGPPISLERLDVGTNLGNAIAKLEDAKELLE	8700906	2.2	MV
>**AVP0095**	RIDLGPPISLERLDVGTNLGNAIAKLEDAKELLES	8700906	1.7	MV
>**AVP0096**	IDLGPPISLERLDVGTNLGNAIAKLEDAKELLESS	8700906	4.9	MV
>**AVP0097**	DLGPPISLERLDVGTNLGNAIAKLEDAKELLESSD	8700906	5.7	MV
>**AVP0098**	LGPPISLERLDVGTNLGNAIAKLEDAKELLESSDQ	8700906	6.5	MV
>**AVP0099**	GPPISLERLDVGTNLGNAIAKLEDAKELLESSDQI	8700906	10.1	MV
>**AVP0100**	PPISLERLDVGTNLGNAIAKLEDAKELLESSDQIL	8700906	1.1	MV
>**AVP0101**	PISLERLDVGTNLGNAIAKLEDAKELLESSDQILR	8700906	3.1	MV
>**AVP0102**	ISLERLDVGTNLGNAIAKLEDAKELLESSDQILRS	8700906	13	MV
>**AVP0103**	SLERLDVGTNLGNAIAKLEDAKELLESSDQILRSM	8700906	12.3	MV
>**AVP0105**	YTPNDITLNNSVALDPIDISIELNKAKSDLEESKE	8700906	100	HPIV 3

>**AVP0106**	TPNDITLNNSVALDPIDISIELNKAKSDLEESKEW	8700906	100	HPIV 3
>**AVP0107**	PNDITLNNSVALDPIDISIELNKAKSDLEESKEWI	8700906	100	HPIV 3
>**AVP0108**	NDITLNNSVALDPIDISIELNKAKSDLEESKEWIR	8700906	100	HPIV 3
>**AVP0109**	DITLNNSVALDPIDISIELNKAKSDLEESKEWIRR	8700906	100	HPIV 3
>**AVP0110**	ITLNNSVALDPIDISIELNKAKSDLEESKEWIRRS	8700906	62	HPIV 3
>**AVP0111**	TLNNSVALDPIDISIELNKAKSDLEESKEWIRRSN	8700906	72	HPIV 3
>**AVP0112**	LNNSVALDPIDISIELNKAKSDLEESKEWIRRSNQ	8700906	1	HPIV 3
>**AVP0113**	NNSVALDPIDISIELNKAKSDLEESKEWIRRSNQK	8700906	6	HPIV 3
>**AVP0114**	NSVALDPIDISIELNKAKSDLEESKEWIRRSNQKL	8700906	0.2	HPIV 3
>**AVP0115**	SVALDPIDISIELNKAKSDLEESKEWIRRSNQKLD	8700906	2	HPIV 3
>**AVP0116**	VALDPIDISIELNKAKSDLEESKEWIRRSNQKLDS	8700906	1	HPIV 3
>**AVP0117**	ALDPIDISIELNKAKSDLEESKEWIRRSNQKLDSI	8700906	0.1	HPIV 3
>**AVP0118**	LDPIDISIELNKAKSDLEESKEWIRRSNQKLDSIG	8700906	0.03	HPIV 3
>**AVP0119**	DPIDISIELNKAKSDLEESKEWIRRSNQKLDSIGN	8700906	0.2	HPIV 3
>**AVP0120**	PIDISIELNKAKSDLEESKEWIRRSNQKLDSIGNW	8700906	0.07	HPIV 3
>**AVP0122**	DISIELNKAKSDLEESKEWIRRSNQKLDSIGNWHQ	8700906	2	HPIV 3
>**AVP0123**	ISIELNKAKSDLEESKEWIRRSNQKLDSIGNWHQS	8700906	2	HPIV 3
>**AVP0124**	SIELNKAKSDLEESKEWIRRSNQKLDSIGNWHQSS	8700906	1	HPIV 3
>**AVP0125**	IELNKAKSDLEESKEWIRRSNQKLDSIGNWHQSST	8700906	2	HPIV 3
>**AVP0127**	IINFYDPLVFPSDEFDASISQVNEKINQSLAFIRK	8700906	91	RSV
>**AVP0128**	INFYDPLVFPSDEFDASISQVNEKINQSLAFIRKS	8700906	93	RSV
>**AVP0129**	NFYDPLVFPSDEFDASISQVNEKINQSLAFIRKSD	8700906	100	RSV
>**AVP0130**	FYDPLVFPSDEFDASISQVNEKINQSLAFIRKSDE	8700906	20	RSV
>**AVP0131**	YDPLVFPSDEFDASISQVNEKINQSLAFIRKSDEL	8700906	6	RSV
>**AVP0132**	DPLVFPSDEFDASISQVNEKINQSLAFIRKSDELL	8700906	8	RSV
>**AVP0133**	PLVFPSDEFDASISQVNEKINQSLAFIRKSDELLH	8700906	30	RSV
>**AVP0135**	VFPSDEFDASISQVNEKINQSLAFIRKSDELLHNV	8700906	19	RSV
>**AVP0136**	FPSDEFDASISQVNEKINQSLAFIRKSDELLHNVN	8700906	8	RSV
>**AVP0137**	PSDEFDASISQVNEKINQSLAFIRKSDELLHNVNA	8700906	6	RSV
>**AVP0138**	SDEFDASISQVNEKINQSLAFIRKSDELLHNVNAG	8700906	6	RSV

>**AVP0140**	EFDASISQVNEKINQSLAFIRKSDELLHNVNAGKS	8700906	13	RSV
>**AVP0142**	DASISQVNEKINQSLAFIRKSDELLHNVNAGKSTT	8700906	8	RSV
>**AVP0143**	FDASISQVNEKINQSLAFIRKSDELLHNVNAGKST	8700906	0.051	RSV
>**AVP0144**	IDISIELNKAKSDLEESKEWIRRSNQKLDSIGNWH	8700906	0.015	MV
>**AVP0145**	LHRIDLGPPISLERLDVGTNLGNAIAKLEDAKELL	8700906	0.068	HPIV 3
>**AVP0146**	DLSNQINSINKSLKSAEDWIADSNFFANQARTAK	8806544	5.1	HPIV 2
>**AVP0147**	ELNKAKSDLEESKEWIRRSNQKLDSIGNWHQSSTT	8806544	2.2	HPIV 3
>**AVP0159**	ISLERLDVGTNLGNAIAKLEDAKELLESSDQILRSM	9010292	0.02	MV
>**AVP0168**	SLLGRMKGA	9843489	2.4	HBV
>**AVP0169**	SLLGRMKG	9843489	6.4	HBV
>**AVP0172**	CRFPNITNSHVPILQERPPLENRVLTGWGL	10516085	2.5	HTLV 1
>**AVP0173**	KGSVVIVGRIILSGRK	10574908	5.7	HCV
>**AVP0174**	VRLGSISVIGIVRGKK	10574908	137	HCV
>**AVP0175**	RGGSVVIVGRIILSGRK	10574908	3.4	HCV
>**AVP0177**	RRKKAAVALLPAVLLALLAP	11222686	0.7	HSV 1
>**AVP0179**	CTLTTKLYC	12021868	1	NDV
>**AVP0182**	FKLRAKIKVRLRAKIKL	12208971	0.46	FIV
>**AVP0184**	FLAAARIAKRVAKKARKLAKRAARKRK	12208971	0.94	FIV
>**AVP0185**	FRFKIKFRLKFRFKARFKFRAKFRA	12208971	1.32	FIV
>**AVP0186**	FAVGLRAIKRALKKLRRGVRKVAKDL	12208971	5.47	FIV
>**AVP0187**	KRKRAVKRVGRRLKKLARKIARLGVAKLAGLRAVKLF	12208971	3.38	FIV
>**AVP0188**	GAKKGAKKGKKGAKKGAKGAGAKGAGAFKKKK	12208971	4.07	FIV
>**AVP0189**	FAKKFAKKFKKFAKKFAKFAFAF	12208971	3.05	FIV
>**AVP0191**	FLFAFRIFKRVFKKFRKLFKRAF	12208971	10.51	FIV
>**AVP0192**	KRKRAVKRVGRRLKKKLARKIARLGVAF	12208971	7.82	FIV
>**AVP0194**	FALALKALKKALKKLKKALKKAL	12208971	7.07	FIV
>**AVP0195**	FAIAIKAIKKAIKKIKKAIKKAI	12208971	8.71	FIV
>**AVP0196**	FKVKAKVKAKVKAKVKAKKKK	12208971	2.85	FIV
>**AVP0197**	AVKRVGRRLKKLARKIARLGVAF	12208971	3.37	FIV
>**AVP0200**	PLSPPLRNTHPQAMQWNSTTF	12469306	30	HBV

>**AVP0201**	PTSNHSPTSCPPTCPGYRWMCLRRF	12469306	35	HBV
>**AVP0202**	LPRRLHLEPAFLPYSVKAHECC	12857903	11	HCMV
>**AVP0204**	EQVLKAVTNVLSPVFPGGET	12857903	280	HCMV
>**AVP0207**	RRLHLEPAFLPYSVKAHECC	12857903	20	HCMV
>**AVP0208**	LPRRLHLEPAFLPYSVKAHEC	12857903	75	HCMV
>**AVP0210**	QLESLTDRELLLLIARKTCGSVE	12912982	1	EBoV
>**AVP0214**	ACFPWGNTWCGGK	12951030	13	HCV
>**AVP0216**	ACFPWGKEYCGGK	12951030	22	HCV
>**AVP0218**	ACFPWGNQWCGGK	12951030	6	HCV
>**AVP0421**	AALEAKICHQIEYYFGDF	15016896	6.688	HCV
>**AVP0422**	KWKVFKKIEKMGRNIRNGIVKAGPAIAVLGEAKAL	15081088	3.4	JV
>**AVP0423**	KVLTTGLPALISWIKRKRQQ	15081088	0.86	JV
>**AVP0424**	GIGKFLHSAGKFGKAFVGEIMKS	15081088	36.59	HSV 1
>**AVP0426**	ILPWKWPWWPWRR	15081088	28.68	HSV 1
>**AVP0427**	ILMCFSIDSPDSLEN	15113882	1.132	RSV
>**AVP0428**	ILMAFSIDSPDSLEN	15113882	72.368	RSV
>**AVP0429**	ILMCFSINSPNSLQN	15113882	70.875	RSV
>**AVP0430**	GSRVQIRCRFRNSTR	15130536	19	HSV 1
>**AVP0460**	YDHIQDHVNTMFSRLATSWCLLQNKERALWAEAA	15269351	5	BoHV 1
>**AVP0463**	FNLSDHSESIQKKFQLMKEHVNKIG	15564453	0.17	ASLV-A
>**AVP0464**	SDHSESIQKKFQLMKEHVNKIGVDS	15564453	5	ASLV-A
>**AVP0548**	GYHLMSFPQAAPHGVVFLHVTW	16616792	2	SARS-CoV
>**AVP0549**	GVFVFNGTSWFITQRNFFS	16616792	2	SARS-CoV
>**AVP0553**	GYFVQDDGEWKFTGSSYYY	16616792	4	MHV
>**AVP0583**	SWLRDIWDWICEVLSD	18287023	0.98	HCV
>**AVP0584**	SWLRDIWDWICEVLSDFK	18287023	0.79	HCV
>**AVP0585**	WICEVLSDFK	18287023	27	HCV
>**AVP0586**	CEVLSDFK	18287023	27	HCV
>**AVP0587**	SGSWLRDIWDWICEVLSDFK	18287023	1.7	HCV
>**AVP0588**	GSWLRDIWDWICEVLSDFK	18287023	0.51	HCV

>**AVP0589**	SWLRDIWDWICEVLSDFKT	18287023	1.7	HCV
>**AVP0590**	SWLRDIWDWICEVLSDFKTW	18287023	0.51	HCV
>**AVP0591**	SWRLIDWDWICEVLSDFK	18287023	4	HCV
>**AVP0595**	SWLRDIWDWICEVL	18287023	11.3	HCV
>**AVP0596**	VLDLIYSLHKQINRGLKKIVL	18287023	36	HCV
>**AVP0598**	KFDSLVECIWDWIDRLWS	18287023	0.48	HCV
>**AVP0600**	SIWRDWVDLICEFLSDWK	18287023	0.4	HCV
>**AVP0602**	SWLRDVWDWVCTILTDFK	18287023	2.1	HCV
>**AVP0603**	DWLRIIWDWVCSVVSDFK	18287023	0.55	HCV
>**AVP0606**	SWLRDIWDWICEV	18287023	27	HCV
>**AVP0608**	SWLRDIWDWISEVLSDFK	18287023	13.5	HCV
>**AVP0609**	SWLRDIWDWIREVLSDFK	18287023	12.5	HCV
>**AVP0610**	SWLRDIWDWIEEVLSDFK	18287023	13	HCV
>**AVP0613**	SWLRDIWDWICKVLSDFK	18287023	6.8	HCV
>**AVP0614**	SWLDRIWRWICKVLSRFE	18287023	1.7	HCV
>**AVP0615**	SWLRDIWRWICKVLSRFK	18287023	0.84	HCV
>**AVP0616**	SWLRRIWRWICKVLSRFK	18287023	0.89	HCV
>**AVP0617**	SWLRDIWDWICE	18287023	27	HCV
>**AVP0618**	SWLRDIWDWI	18287023	27	HCV
>**AVP0619**	SWLRDIWD	18287023	27	HCV
>**AVP0620**	LRDIWDWICEVLSDFK	18287023	27	HCV
>**AVP0621**	DIWDWICEVLSDFK	18287023	27	HCV
>**AVP0622**	WDWICEVLSDFK	18287023	27	HCV
>**AVP0623**	RTQRRGRTGRGKPGIYR	18479669	27.1	HCV
>**AVP0624**	STQRRGRTGRGRRGIYR	18479669	24.3	HCV
>**AVP0625**	RRGRTGRGRRGIYR	18479669	0.2	HCV
>**AVP0626**	RTGRGRRGIYR	18479669	34.6	HCV
>**AVP0627**	RGRRGIYR	18479669	313	HCV
>**AVP0633**	AAQRRGRIGRNPSQVGD	18479669	358	HCV
>**AVP0636**	RVGRNPNQVGD	18479669	374	HCV

>**AVP0637**	RRGRVGRNPNQVGD	18479669	196	HCV
>**AVP0638**	AAQRRGRVGRNPNQVGD	18479669	442	HCV
>**AVP0652**	RNPSQVGD	18479669	383	WNV
>**AVP0653**	RIGRNPSQVGD	18479669	285	WNV
>**AVP0654**	RRGRIGRNPSQVGD	18479669	156	WNV
>**AVP0674**	AAHLIDALYAEFLGGRVLTT	18572274	250	HSV 1
>**AVP0675**	GLASTLTRWAHYNALIRAF	18572274	250	HSV 1
>**AVP0677**	CCFLRIQNDSIIRLGDLQPLSQRVSTDWQ	18680566	3.49	BKV
>**AVP0816**	CCFLNITNSHVSILQERPPLENRVLTGWGL	19114713	0.18	HTLV 1
>**AVP0817**	LNITNSHVSILQERPPLENRVL	19114713	5.8	HTLV 1
>**AVP0818**	CFLNITNSHVSILQERPPLENRV	19114713	0.19	HTLV 1
>**AVP0820**	CFLNITNSHVSILQEAPPLENAV	19114713	8.5	HTLV 1
>**AVP0961**	GWWYKGRARPVSAVA	19558186	3.2	INFV A
>**AVP0962**	RAVWRHSVATPSHSV	19558186	6.5	INFV A
>**AVP0963**	GAWYKGRARPVSAVA	19558186	53	INFV A
>**AVP0964**	GWWYKGRARAVSAVA	19558186	89	INFV A
>**AVP0965**	AVASVPRARGKYWWG	19558186	44	INFV A
>**AVP0966**	DFRRLPGAFWQLRQP	19558186	52	INFV A
>**AVP0967**	AETVESCLAKPHTEN	19558186	66	INFV A
>**AVP0978**	VNPTLLFLKVPAQNAISTTFPYT	19841738	0.662	INFV A
>**AVP0979**	PTLLFLKVPAQNAISTTFPYT	19841738	0.483	INFV A
>**AVP0980**	LLFLKVPAQNAISTTFPYT	19841738	3	INFV A
>**AVP0981**	FLKVPAQNAISTTFPYT	19841738	3	INFV A
>**AVP0982**	KVPAQNAISTTFPYT	19841738	3	INFV A
>**AVP0983**	MDVNPTLLFLKVPAQNAIST	19841738	0.034	INFV A
>**AVP0984**	MDVNPTLLFLKVPAQNAI	19841738	0.029	INFV A
>**AVP0986**	MDVNPTLLFLKVPAQ	19841738	0.043	INFV A
>**AVP0987**	MDVNPTLLFLKVPA	19841738	0.035	INFV A
>**AVP0988**	MDVNPTLLFLKVP	19841738	0.138	INFV A
>**AVP0989**	MDVNPTLLFLKV	19841738	0.644	INFV A

>**AVP0990**	MDVNPTLLFLK	19841738	0.899	INFV A
>**AVP0992**	MDVNPTLLF	19841738	3	INFV A
>**AVP0993**	MDVNPTLL	19841738	3	INFV A
>**AVP0997**	MNINPTLLFLKVPIQ	19841738	0.007	INFV A
>**AVP0998**	MDVNPTLLFIDVPAQ	19841738	3	INFV A
>**AVP0999**	MNINPTLLFLKVPAQ	19841738	0.013	INFV A
>**AVP1001**	MDVNPTFLFLKVPAQ	19841738	3	INFV A
>**AVP1002**	MDVNPYLLFLKVPAQ	19841738	0.022	INFV A
>**AVP1003**	MDVNPFLLFLKVPAQ	19841738	0.003	INFV A
>**AVP1004**	MDVNPWLLFLKVPAQ	19841738	0.003	INFV A
>**AVP1005**	MDVNPHLLFLKVPAQ	19841738	0.292	INFV A
>**AVP1006**	MDVNPCLLFLKVPAQ	19841738	0.044	INFV A
>**AVP1008**	MNINPYPLFIDVPIQ	19841738	0.045	INFV B
>**AVP1056**	RWMVWRHWFHRLRLPYNPGKNKQNQQWP	20582308	8	DENV 2
>**AVP1057**	RQMRAWGQDYQHGGMGYSC	20582308	36	DENV 2
>**AVP1059**	RRRRRRRRHPAEPGSTVTTQNTASQTMS	20686048	20	ASFV
>**AVP1118**	ASLRVRIKKQ	21576348	0.21	HSV 1
>**AVP1160**	CEELRVRLASHLRKLRKRLLRDADDLQKRLAVY	22334503	0.5	HCV
>**AVP1161**	EELRVRLASHLRKLRKRLLRDADDLQKRLAVY	22334503	10	HCV
>**AVP1162**	CEEIRARLSTHLRKMRKRLMRDADDLQKRLAVY	22334503	10	HCV
>**AVP1163**	CEEQAQQIRLQAEAFQARLKSWFEPLVEDM	22334503	10	HCV
>**AVP1164**	CVRLASHLRKLRKRLLRDADDL	22334503	10	HCV
>**AVP1165**	CIRLQAEAFQARLKSWFEPLV	22334503	10	HCV
>**AVP1167**	LRVRLASHLRKLRKRLLRDADDLQKRLAVY	22334503	10	HCV
>**AVP1168**	EELRVRLASHLRKLRKRLLRDADDL	22334503	10	HCV
>**AVP1169**	VRLASHLRKLRKRLLRDADDLQKRLAVY	22334503	10	HCV
>**AVP1170**	CVRLASHLRKLRKRLLRDADDLQKRLAVY	22334503	0.8	HCV
>**AVP1171**	CLRVRLASHLRKLRKRLLRDADDL	22334503	4	HCV
>**AVP1172**	CLRKLRKRLLRC	22334503	10	HCV
>**AVP1175**	LLGFILAFLGWIGAIVST	22378192	4.3	HCV

>AVP1176	FILAFLGWIGAIVSTALP	22378192	8.9	HCV
>AVP1177	AFLGWIGAIVSTALPQWR	22378192	12.5	HCV
>AVP1178	GWIGAIVSTALPQWRIYS	22378192	21.5	HCV
>AVP1180	VSTALPQWRIYSYAGDNI	22378192	25	HCV
>AVP1181	ALPQWRIYSYAGDNIVTA	22378192	25	HCV
>AVP1183	MANAGLQLLGFILAFL	22378192	7.6	HCV
>AVP1185	MANAGLQLLGFILAFLGWIGAI	22378192	4	HCV
>AVP1186	MANAGLQLLGFILAFLGWIGAIVS	22378192	5.1	HCV
>AVP1188	AGALMFAWLLLGLQGIFN	22378192	25	HCV
>AVP1189	MASAGMQILGVVLTLLGW	22378192	25	HCV
>AVP1190	MANSGLQLLGFSMALLGW	22378192	25	HCV
>AVP1191	MASTGLELLGMTLAVLGW	22378192	25	HCV
>AVP1192	ATSSANSKA	22465300	10	JEV
>AVP1208	CAGKRKSG	22780881	6.7	DENV 2
>AVP1210	ASLRVRIKK	22850525	0.25	RSV
>AVP1211	GELGRLVYLLDGPGYDPIHCSLAYGDASTLVVF	22965230	0.021	HCV
>AVP1212	GELGRLVYLLDGPGYDPI	22965230	0.125	HCV
>AVP1213	HCSLAYGDASTLVVF	22965230	0.001	HCV
>AVP1214	GELGRPVYVLGDPGYYATHCIYATTNDALIFSV	22965230	0.026	HCV
>AVP1216	HCIYATTNDALIFSV	22965230	0.001	HCV
>AVP1217	GELGRIPSDTYDLAVGALHCPFYLVSGLVYLDG	22965230	0.001	HCV
>AVP1218	GELGRLVYLLDGPGYDPIHCDVVTRGGSHLFNF	22965230	0.011	HCV
>AVP1219	GELDELVYLLDGPGYDPIHCDVVTRGGSRLFNF	22965230	0.001	HCV
>AVP1220	GELGRLVYLLDGPGYDPIHCD	22965230	0.124	HCV
>AVP1221	GELDELVYLLDGPGYDPIHS	22965230	0.023	HCV
>AVP1222	RQIKINFQNRRMKNKKGELDELVYLLDGPGYDPIHS	22965230	0.024	HCV
>AVP1223	LLDCWVRLGRYLLRRLKT	23175359	10	HCV
>AVP1224	LLDCWVRLGRYLLRRLKTPFTRL	23175359	15	HCV
>AVP1225	LLDCWVRLGRYLLRRLKTPFT	23175359	20	HCV
>AVP1226	LLDCWVRLGRYLLRRLKTP	23175359	3	HCV

>AVP1227	LLDCWVRLGRYLLRRLK	23175359	5	HCV
>AVP1228	LDCWVRLGRYLLRRLKTPFTRL	23175359	20	HCV
>AVP1230	LDCWVRLGRYLLRRLKTP	23175359	4	HCV
>AVP1232	CWVRLGRYLLRRLKTPFT	23175359	2	HCV
>AVP1233	WVRLGRYLLRRLKTPFTR	23175359	12	HCV
>AVP1234	VRLGRYLLRRLKTPFTRL	23175359	20	HCV
>AVP1235	MAILGDTAWDFGSLGGVFTSIGKALHQVFGAIY	23226444	3.5	DENV 2
>AVP1250	HGLASTLTRWAHYNALIRAF	23429490	0.1	HSV 1
>AVP1365	GSLLGRMKGA	US6544520	0.79	HBV
>AVP1367	ADGSLLGRMKGAAG	US6544520	4.5	HBV
>AVP1369	RSLLGRMKGA	US6544520	0.29	HBV
>AVP1370	HRSLLGRMKGA	US6544520	0.5	HBV
>AVP1371	MHRSLLGRMKGA	US6544520	0.8	HBV
>AVP1477	FKLPLGINITNFRAILTAFS	US7491489	266.434	SARS-CoV
>AVP1478	PTTFMLKYDENGTITDAVDC	US7491489	60.018	SARS-CoV
>AVP1479	VLYNSTFFSTFKCYGVSATK	US7491489	262.667	SARS-CoV
>AVP1480	PALNCYWPLNDYGFYTTSGI	US7491489	258.33	SARS-CoV
>AVP1482	YQDVNCTDVSTAIHADQLTP	US7491489	61.704	SARS-CoV
>AVP1483	SNNTIAIPTNFSISITTEVM	US7491489	278.801	SARS-CoV
>AVP1484	QYGSFCTQLNRALSGIAAEQ	US7491489	13.854	SARS-CoV
>AVP1485	GIGVTQNVLYENQKQIANQF	US7491489	262.539	SARS-CoV
>AVP1487	IPESSELTLQELLGEERR	15182185	10.5	HPV
>AVP1490	YKFACPECPKRFMRSDHLSKHITLHELLGEERR	15182185	19.3	HPV
>AVP1492	ALQELLGQWLKDGGPSSGRPPPS	15182185	36.8	HPV
>AVP1493	ALQELLGEYIQWLKDGGPSSGRPPPS	15182185	26.2	HPV
>AVP1494	YLQELLGE	15182185	74.3	HPV
>AVP1505	SWLRDLWDWICEVLSDFK	21801309	1.4	HCV
>AVP1506	SWLRDIWDWLCEVLSDFK	21801309	1.2	HCV
>AVP1507	SWLRDIWDWICELLSDFK	21801309	1.2	HCV
>AVP1508	SWLRDLWDWLCEVLSDFK	21801309	1.4	HCV

>**AVP1509**	SWLRDLWDWICELLSDFK	21801309	0.78	HCV
>**AVP1511**	SWLRDLWDWLCELLSDFK	21801309	1.2	HCV
>**AVP1513**	SWLRDIWDWVCEVLSDFK	21801309	1.4	HCV
>**AVP1514**	SWLRDIWDWACEVLSDFK	21801309	4	HCV
>**AVP1515**	SWLRDIWDWGCEVLSDFK	21801309	5	HCV
>**AVP1516**	SWLRDIWDWSCEVLSDFK	21801309	5	HCV
>**AVP1517**	SWLRDIWDWECEVLSDFK	21801309	5	HCV
>**AVP1518**	SWLRDIWDWKCEVLSDFK	21801309	5	HCV
>**AVP1527**	KNGRKLCLDLQAALY	20347875	27.855	VACV
>**AVP1528**	AALYKKKIIKKLLES	20347875	33.871	VACV
>**AVP1558**	VYTDKVDISSQISSMNQSLQQSKDYIKEAQKILDTV	16973588	0.04	HeV
>**AVP1560**	VALDPIDISIELNKAKSDLEESKEWIRRSNQKLDSI	16973588	0.008	HeV
>**AVP1565**	VALDPIDISIELNKAKSDLEESKEWIRR	16973588	10	HeV
>**AVP1572**	VANDPIDISIELNKAKSDLEESKEWIRRSNQKLDSI	16973588	0.75	HPIV 3
>**AVP1573**	VALDPIDISIELNKAKSDLEESKEWIRRSNQKLDSD	16973588	0.35	HPIV 3
>**AVP1574**	VANDPIDISIELNKAKSDLEESKEWIRRSNQKLDSD	16973588	0.75	HPIV 3
>**AVP1578**	KVDISSQISSMNQSLQQSKDYIKEAQRLLDTVNPSL	16026621	0.013	NiV
>**AVP1579**	FWFTLIKTQAKQPARYRRFC	21093488	3	DENV 2
>**AVP1583**	GINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYE	18983873	0.62	SARS-CoV
>**AVP1584**	GINASVVNIQKEIDRLNEVAKNLNESLIDL	18983873	0.8	SARS-CoV
>**AVP1585**	GINASVVNIQKEIDRLNEVAKNL	18983873	1.04	SARS-CoV
>**AVP1641**	KKKKYRNIRRPG	18383098	33.889	SARS-CoV
>**AVP1695**	CPFVKTQLC	17441904	263	SNV
>**AVP1700**	FLHFLHHLF	23415044	0.811	HCV
>**AVP1701**	FLGFLHHLF	23415044	1.115	HCV
>**AVP1702**	FLGFLKNLF	23415044	1.965	HCV
>**AVP1705**	ISGINASVVNIQEEIKKLNEEAKKLNESLIDLQEL	17942557	0.004	SARS-CoV
>**AVP1706**	ISGINASVVNIQKEIDRLNEVAKNLNESLIDLQEL	17942557	0.005	SARS-CoV
>**AVP1707**	IEEINKKVEEIQKKIEELNKKAEELNKKLEELQKK	17942557	100	SARS-CoV
>**AVP1708**	PRPISYLKGSSGGPL	22910295	5	HCV

>**AVP1716**	TLKPIFKLPLGINITNFR	19853613	11	SARS-CoV
>**AVP1718**	YENQKQIANQFNKAISQIQESLTTTSTA	18442051	1.16	SARS-CoV
>**AVP1722**	FGGASCCLYCRCHIDHPNPKGFCDLKGKY	22659295	160	SARS-CoV
>**AVP1723**	GGASCCLYCRCH	22659295	160	SARS-CoV
>**AVP1727**	HRILMRIRAMMT	22743126	34.136	PRRSV
>**AVP1728**	HRALMRIRQMMT	22743126	34.181	PRRSV
>**AVP1729**	HRILMRIR	22743126	43.502	PRRSV
>**AVP1730**	HRIAMRIRQMMT	22743126	54.587	PRRSV
>**AVP1731**	HRILMRIRQMMT	22743126	56	PRRSV
>**AVP1732**	ARILMRIRQMMT	22743126	56.123	PRRSV
>**AVP1733**	HRILMRIRQAMT	22743126	57.981	PRRSV
>**AVP1737**	HRILMRIRQMMA	22743126	89.412	PRRSV
>**AVP1738**	HRILMRIRQMAT	22743126	120.524	PRRSV
>**AVP1739**	HRILMRARQMMT	22743126	150.141	PRRSV
>**AVP1740**	LMRIRQMMT	22743126	263.81	PRRSV
>**AVP1741**	HRILARIRQMMT	22743126	435.5	PRRSV
>**AVP1756**	WLVFFVIFYFFR	22258859	0.094	INFV A
>**AVP1757**	AWDFGSLGGVFTSIGKALHQVFGAIYGAA	20881042	0.1	DENV 1
>**AVP1758**	AWDFGSVGGVLNSLGKMVHQIFGSAYTAL	20881042	0.1	DENV 1
>**AVP1759**	RRKKIFYFFR	22258859	0.155	INFV A
>**AVP1762**	HVTTTFAPPPPR	21176936	9.054	TGEV
>**AVP1764**	FKPSSPPSITLW	21176936	17.489	TGEV
>**AVP1765**	ATCYCRTGRCATRESLSGVCRISGRLYRLCCR	23269800	8.226	HSV 2
>**AVP1766**	ATCYCRTGRCATRESLSGVCEISGRLYRLCCR	23269800	11.62	HSV 2
>**AVP1767**	ATCYCRTGRCATRESLSGVCEIRGRLYRLCCR	23269800	11.683	HSV 2
>**AVP1768**	ATCYCRRGRCATRESLSGVCEISGRLYRLCCR	23269800	13.041	HSV 2
>**AVP1770**	ALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFEL	16153058	41.6	SARS-CoV
>**AVP1772**	ATCYCRTGRCATRESRSGVCEISGRLYRLCCR	23269800	21.93	HSV 2
>**AVP1790**	AWDFGSIGGVFTSVGKLVHQVFGTAYGVL	20881042	1.5	DENV 1
>**AVP1797**	AWDFGSVGGLFTSLGKAVHQVFGSVYTIM	20881042	6	DENV 4

>**AVP1801**	NADIIKSLIRKTIINASKNTASLSILQHLYVLRS	21518442	12	MDV
>**AVP1802**	HMNASDMEIKSYINMIESVEESSNYDF	21518442	4	MDV
>**AVP1804**	CTEEHIVATELVIQEMYIKINVKNSP	21518442	8	MDV
>**AVP1805**	RIILGQCIKREAEAAIEQIFRTKYNDSH	21518442	9	MDV
>**AVP1808**	FFVIFYRRKK	22258859	1.482	INFV A
>**AVP1811**	GICRCICGRGICRCICGRIGGRVPGVGVPGVGHHHHHH	23171075	21.4	DENV 2
>**AVP1814**	LESEVTAIKNALKKTNEAVSTLGNGVRVLATAVRE	17967906	3.24	hMPV
>**AVP1816**	FNVALDQVFESIENSQALVDQSNRILSSAE	17967906	9	hMPV
>**AVP1817**	FNVALDQVFESIENSQALVDQSNRILSSAEKGN	17967906	9	hMPV
>**AVP1818**	NVALDQVFESIENSQALVDQSNRILSSA	17967906	9	hMPV
>**AVP1820**	LFRLIKSLIKRLVSAFK	22791717	11	HBV
>**AVP1821**	RGGRLCYCRRRFCVCVGR	23093838	11.7	DENV 2
>**AVP1831**	SISNALNKLEESNRNLDKVNVKLT	12127571	3.27	NDV
>**AVP1832**	KQNAANILRLKESIAATNEAVHEV	12127571	0.023	NDV
>**AVP1849**	FKCRRWQWRMKKLGAPSITCVRRAFA	17481742	9.7	HPV
>**AVP1855**	CNDFRSKTC	19497129	48	AIV
>**AVP1857**	NGIGVTQNVLYENQKQIANQFNKAISQIQESLTTTSTA	15184046	0.14	SARS-CoV
>**AVP1858**	IQKEIDRLNEVAKNLNESLIDLQELGK	15184046	1.19	SARS-CoV
>**AVP1924**	TDVILMCFSIDSPDSLENI	14576104	7.6	RSV
>**AVP1925**	CSIELSDIPLSVDFNTMID	14576104	50	RSV
>**AVP1926**	ADVILMCFSIDSPDSLENI	14576104	2.56	RSV
>**AVP1927**	TAVILMCFSIDSPDSLENI	14576104	1.37	RSV
>**AVP1928**	TDAILMCFSIDSPDSLENI	14576104	6.6	RSV
>**AVP1929**	TDVALMCFSIDSPDSLENI	14576104	11.6	RSV
>**AVP1930**	TDVIAMCFSIDSPDSLENI	14576104	5.42	RSV
>**AVP1931**	TDVILACFSIDSPDSLENI	14576104	1.43	RSV
>**AVP1932**	TDVILMAFSIDSPDSLENI	14576104	50	RSV
>**AVP1933**	TDVILMCASIDSPDSLENI	14576104	6.29	RSV
>**AVP1934**	TDVILMCFAIDSPDSLENI	14576104	6.82	RSV
>**AVP1935**	TDVILMCFSADSPDSLENI	14576104	3.52	RSV

>**AVP1936**	TDVILMCFSIASPDSLENI	14576104	4.36	RSV
>**AVP1937**	TDVILMCFSIDAPDSLENI	14576104	2.26	RSV
>**AVP1938**	TDVILMCFSIDSADSLENI	14576104	15.32	RSV
>**AVP1939**	TDVILMCFSIDSPASLENI	14576104	2.61	RSV
>**AVP1941**	TDVILMCFSIDSPDSAENI	14576104	2.27	RSV
>**AVP1942**	TDVILMCFSIDSPDSLANI	14576104	9.83	RSV
>**AVP1943**	TDVILMCFSIDSPDSLEAI	14576104	18.47	RSV
>**AVP1944**	TDVILMCFSIDSPDSLENA	14576104	4.89	RSV
>**AVP1945**	TDVILMCFSI	14576104	50	RSV
>**AVP1946**	TDVILMCFSIDSP	14576104	50	RSV
>**AVP1948**	DVILMCFSIDSPDSLENI	14576104	1.23	RSV
>**AVP1949**	VILMCFSIDSPDSLENI	14576104	16.95	RSV
>**AVP1950**	ILMCFSIDSPDSLENI	14576104	7.17	RSV
>**AVP1951**	CFSIDSPDSLENI	14576104	50	RSV
>**AVP1953**	ILMCFSIDSPDSLE	14576104	3.5	RSV
>**AVP1954**	ILMCFSIDSPDSL	14576104	12.4	RSV
>**AVP1955**	ILMCFSIDSPDS	14576104	6.36	RSV
>**AVP1956**	ILMCFSIDSPD	14576104	4.61	RSV
>**AVP1957**	ILMCFSIDSP	14576104	35.77	RSV
>**AVP1958**	ILMCFSIDS	14576104	50	RSV
>**AVP1959**	ILMCFSID	14576104	50	RSV
>**AVP1973**	QLQKWEDWVRWIGNIPQYLKG	12610147	0.013	FIV
>**AVP1974**	QKWEDWVRWIGN	12610147	0.178	FIV
>**AVP1975**	WEDWVRWIGNIP	12610147	0.03	FIV
>**AVP1976**	QLQKWEDWVRWI	12610147	0.02	FIV
>**AVP1977**	WEDWVRWI	12610147	0.057	FIV
>**AVP1983**	AEDWVRWI	12610147	4.623	FIV
>**AVP1984**	WADWVRWI	12610147	0.162	FIV
>**AVP1985**	WEAWVRWI	12610147	0.1	FIV
>**AVP1986**	WEDAVRWI	12610147	53.753	FIV

>**AVP1987**	WEDWARWI	12610147	0.069	FIV
>**AVP1989**	WEDWVRAI	12610147	53.753	FIV
>**AVP1990**	WEDWVRWA	12610147	0.15	FIV
>**AVP1993**	PPATHTIADRNHTPFSDV	11932408	0.33	MV
>**AVP1994**	ATHTICDRNHTW	11932408	60	MV
>**AVP2001**	FHFEVFNFVPCSICSNNPTCWAICKRIPNKKPGKK	11487583	80	RSV
>**AVP2002**	VPCSICSNNPTCWAICKRIPNKKPGKK	11487583	165	RSV
>**AVP2003**	CSICSNNPTCWAICKRIPNKKPGKK	11487583	177	RSV
>**AVP2005**	KPPSKPNNDFHFEVFNFVPCSICSNNPTCWAICKRI	11487583	12	RSV
>**AVP2006**	KPNNDFHFEVFNFVPCSICSNNPTCWAICKRI	11487583	25	RSV
>**AVP2007**	KQRQNKPPSKPNNDFHFEVFN	11487583	190	RSV
>**AVP2008**	KQRQNKPPSKPNNDFHF	11487583	240	RSV
>**AVP2009**	KPPSKPNNDFHFEVFNFVP	11487583	220	RSV
>**AVP2011**	KPPSKPNNDFHFEVFNF	11487583	7	RSV
>**AVP2020**	NILRLKESITATIEAVHEVTDGLS	10364347	2	NDV
>**AVP2021**	SMEKLAGFGAVGAGATAEETRRMLHRAFDTLA	10220447	15	HSV 1
>**AVP2023**	ALDKLEESNSKLDKVNVKLT	9527912	2	NDV
>**AVP2024**	ALDKAEESNSKLDKVNVKLT	9527912	6	NDV
>**AVP2025**	ALDKAEESNSKADKVNVKLT	9527912	35	NDV
>**AVP2032**	YGAVVNDL	3012359	29.749	HSV 1
>**AVP2033**	YAGAVVNDL	3012359	5.929	HSV 1
>**AVP2034**	YAGAVVNDLL	3012360	22	HSV 1
>**AVP2039**	FLQDSKAELEKARKILSEVG	12586338	17	SeV
>**AVP2040**	KAAKKAAKAAKKAAKWAKKAA	15498607	117	HSV 1
>**AVP2041**	AKKAAKKAKKAAKKAKKAAKK	15498607	41	HSV 1
>**AVP2042**	AKKAAKKAKKAAKKAKKWAKK	15498607	47.3	HSV 1
>**AVP2044**	ARRAWRRARRAARRARRAARR	15498607	18.2	HSV 1
>**AVP2045**	ARRAKRRARRAARRARRKARR	15498607	13	HSV 1
>**AVP2046**	ARRAKRRARRAKRRARRKKRR	15498607	24.6	HSV 1
>**AVP2047**	IFKAIWSGIKSLF	20950663	5	HCV

>AVP2057	LFGLIPSLIGGLVSAFK	21620914	1.03	INFV A
>HIP1013	DWLKAFYDKVAEKLKEAF	2170446	20	HIV
>HIP1014	KWLDAFYKDVAKELEKAF	2170446	100	HIV
>HIP1015	KAFEEVLAKKFYDKALWD	2170446	100	HIV
>HIP1113	DQAEHLKTAVQMAVFIHNYKA	22742518	0.085	HIV
>HIP1114	WIHAEIKNSLKIDNLDVNRCIEALD	19801648	2.02	HIV
>HIP1115	WNSLKIDNLDV	19801648	11.9	HIV
>HIP1125	AVGIGAMFLGFLGAAGSTMGAAS	22520838	0.023	HIV
>HIP1126	IRKAHCNISRAKWND	21264298	0.05	HIV
>HIP1127	IRKAHCNISRADWND	21264298	0.05	HIV
>HIP1129	RQLLSGIVQQQNNLLRAIEAQQHLLQK	20605950	0.338	HIV
>HIP1133	QETAYFLLKLAGRWP	12643937	3.5	HIV
>HIP1135	AGERIVDIIATDIQ	12643937	2	HIV
>HIP1137	AGERIVDIIA	12643937	30	HIV
>HIP1138	ATGQETAYFLLKLAGKA	11705373	250	HIV
>HIP1140	PDIVIYQYMDDLYVGSDLEI	16879966	6	HIV
>HIP1142	ETWETWWTEYWQATWIPEWE	16879966	6	HIV
>HIP1143	LQDSGLEVNIVTDSQYALGI	16879966	11	HIV
>HIP1146	WQCLTLTHRGFVLLTITVLR	18201721	12	HIV
>HIP1147	ILPWKWPWWPWPP	15482931	16	HIV
>HIP1148	RSQKEGLHYTCSSHFPYSQYQFWK	22403408	136	HIV
>HIP1149	CSSHFPYSQYQFWK	22403408	28	HIV
>HIP1150	QKEGLHYTCSSHFPYSQYQF	22403408	237	HIV
>HIP1151	SGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARIL	22235115	7	HIV
>HIP1156	NKPFVFLM	22406118	100	HIV
>HIP1162	RINNIPWSEAMM	14967033	1.6	HIV
>HIP1164	YGGIKKEIEAIKKEQEAIKKKIEAIEKEIEA	11572974	0.026	HIV
>HIP1167	TRQARRNRRRWRERQR	20580677	1.7	HIV
>HIP1168	TRQARRNRRRWRERQRAAAAC	20580677	0.35	HIV
>HIP1170	IYWNVSGW	19053244	7	HIV

>HIP1171	FWNWLSAWIKKTYEEIKKTYEEIKKTYEEIERDWEMV	18662985	0.58	HIV
>HIP1173	IKKTYEEIKKTYEEIKKTYEEIKKTYEEIKKTYEE	18662985	200	HIV
>HIP1174	IKKTYEEIKKTYEEIKKTYEEIKKTYEEIERDWEMV	18662985	117	HIV
>HIP1175	ITFEDLLDYYGP	18374356	4	HIV
>HIP138	YQLLIRMI	12054767	120	HIV
>HIP140	AAAMSQVTN	15113844	100	HIV
>HIP144	AEAMSQVTN	15113844	5	HIV
>HIP148	YQLLIRMIY	12054767	5	HIV
>HIP154	LFYLVPGPGH	12054767	200	HIV
>HIP157	AAWWAGIKQEF	16854053	277	HIV
>HIP158	ACAWAGIKQEF	16854053	33	HIV
>HIP159	ACGWAGIKQEF	16854053	46	HIV
>HIP160	ACWAAGIKQEF	16854053	333	HIV
>HIP161	ACWGAGIKQEF	16854053	333	HIV
>HIP162	ACWWAAIKQEF	16854053	90	HIV
>HIP164	ACWWAGIAQEF	16854053	62	HIV
>HIP165	ACWWAGIKAEF	16854053	333	HIV
>HIP166	ACWWAGIKQAF	16854053	333	HIV
>HIP167	ACWWAGIKQEA	16854053	245	HIV
>HIP171	ACWWAGIRQEF	16854053	333	HIV
>HIP173	AEPERRNIKYL	12054767	50	HIV
>HIP179	AAYFLLKLAGRW	16854053	100	HIV
>HIP185	KRIVQRIKDFLR	18591279	63.5	HIV
>HIP189	RPRLSHKGPMPF	10802050	52.209	HIV
>HIP197	TAAFLLKLAGRW	16854053	193	HIV
>HIP198	TASFLLKLAGRW	16854053	186	HIV
>HIP199	TAYALLKLAGRW	16854053	333	HIV
>HIP200	TAYFALKLAGRW	16854053	115	HIV
>HIP201	TAYFLAKLAGRW	16854053	333	HIV
>HIP202	TAYFLLALAGRW	16854053	113	HIV

>HIP203	TAYFLLILAGRW	16854053	4.1	HIV
>HIP204	TAYFLLKAAGRW	16854053	333	HIV
>HIP205	TAYFLLKLAARW	16854053	118	HIV
>HIP207	TAYFLLKLAGRA	16854053	333	HIV
>HIP208	TAYFLLKLAGRL	16854053	315	HIV
>HIP209	TAYFLLKLAGRW	16854053	21	HIV
>HIP218	YALLIRMIYKNI	12054767	8	HIV
>HIP221	YQALIRMIYKNI	12054767	165	HIV
>HIP222	YQLAIRMIYKNI	12054767	14	HIV
>HIP223	YQLLARMIYKNI	12054767	45	HIV
>HIP224	YQLLIAMIYKNI	12054767	34	HIV
>HIP225	YQLLIRAIYKNI	12054767	70	HIV
>HIP226	YQLLIRMAYKNI	12054767	35	HIV
>HIP227	YQLLIRMIAKNI	12054767	40	HIV
>HIP228	YQLLIRMIYANI	12054767	11	HIV
>HIP230	YQLLIRMIYKNA	12054767	11	HIV
>HIP231	YQLLIRMIYKNI	12054767	5	HIV
>HIP232	YQLLIRPIYKNI	12054767	200	HIV
>HIP233	AEASQVTNTATIM	15113844	126	HIV
>HIP235	FKRIVQRIKDFLR	18591279	3.4	HIV
>HIP236	GFLDIIEKIAKSW	20086159	10.5	HIV
>HIP237	GIFDKLAKEISIW	20086159	65.8	HIV
>HIP238	GIIDIAKKLFESW	20086159	20.1	HIV
>HIP239	GIWSDLAEIIKKF	20086159	11.4	HIV
>HIP240	GLFDIIKKIAESW	20086159	11.7	HIV
>HIP241	GLWEKIDKFASII	20086159	65.8	HIV
>HIP244	GWFDIIKKIASEL	20086159	10.7	HIV
>HIP245	GWLKKIESIIDAF	20086159	29.5	HIV
>HIP260	QRPRLSHKGPMPF	10802050	19.745	HIV
>HIP263	RLFDKIRQVIRKF	18591279	58.1	HIV

>HIP264	AEAASQVTNTATIM	15113844	142	HIV
>HIP265	AEAMSQVANTATIM	15113844	110	HIV
>HIP266	AEAMSQVTNTATIM	15113844	10	HIV
>HIP268	FLFPLITSFLSKVL	20086159	16.7	HIV
>HIP269	FRPALIVRTKGTRL	20086159	61.4	HIV
>HIP270	INLKAIAALAKKLL	20086159	67.6	HIV
>HIP271	NQIIEQLIKKEKVY	15790559	240	HIV
>HIP288	ESELVSQIIEQLIKK	15790559	120	HIV
>HIP292	FIHFRIGCQHSRIGI	17490682	200	HIV
>HIP297	GRFKRFRKKFKKLFK	18591279	49.6	HIV
>HIP300	HFPRIWLHSLGQHIY	17490682	187	HIV
>HIP344	QQLLFIHFRIGCQHS	17490682	33	HIV
>HIP361	TWAGVEAIIRILQQL	17490682	0.88	HIV
>HIP363	TYGDTWAGVEAIIRI	17490682	150	HIV
>HIP365	VDKPDYRPRPRPPNM	20086159	54.4	HIV
>HIP367	VEAIIRILQQLLFIH	17490682	0.22	HIV
>HIP388	GFKRIVQRIKDFLRNLV	18591279	0.98	HIV
>HIP450	GNNRPVYIPQPRPPHPRI	20086159	47.4	HIV
>HIP452	GRFKRFRKKFKKLFKKIS	18591279	0.35	HIV
>HIP453	GRFKRFRKPFKKLFKKIS	18591279	3.2	HIV
>HIP549	GKPRPYSPRPTSHPRPIRV	20086159	45.5	HIV
>HIP550	GLKKLLGKLLKKLGKLLLK	20086159	47.5	HIV
>HIP551	GLRRLLGRLLRRLGRLLLR	20086159	4.4	HIV
>HIP552	IKKEKVYLAWVPAHKGIGN	15790559	120	HIV
>HIP553	KGRGKQGGKVRAKAKTRSS	20086159	50	HIV
>HIP557	AEAIPMSIPPEVKFNKPFVF	17448989	100	HIV
>HIP581	ELVNQIIEQLIKKEKVYLAW	15790559	6.9	HIV
>HIP594	GIKEFKREFQRIKDFLRNLV	18591279	1.6	HIV
>HIP595	GIKEFKRIVQRIKDFLRNLV	18591279	1.08	HIV
>HIP596	GIKEWKRIVQRIKDFLRNLV	18591279	7.4	HIV

>HIP597	GIKQFKRIVQRIKDFLRNLV	18591279	0.91	HIV
>HIP610	IQAQPDQSESELVNQIIEQL	15790559	120	HIV
>HIP620	KILEPFRKQNPDIVIYQYMD	15790559	4.8	HIV
>HIP625	KRIVQRIKDFLRNLVPRTES	18591279	40.5	HIV
>HIP630	LAAIPMSIPPEVKFNKPFVF	17448989	100	HIV
>HIP633	LEAIAMSIPPEVKFNKPFVF	17448989	23.5	HIV
>HIP634	LEAIPASIPPEVKFNKPFVF	17448989	13	HIV
>HIP635	LEAIPCSIPPCFAFNKPFVF	17448989	0.27	HIV
>HIP637	LEAIPCSIPPCFLFGKPFVF	17448989	0.39	HIV
>HIP638	LEAIPCSIPPCVAFNKPFVF	17448989	0.18	HIV
>HIP639	LEAIPCSIPPCVFFGKPFVF	17448989	0.28	HIV
>HIP640	LEAIPCSIPPCVFFNKPFVF	17448989	0.93	HIV
>HIP646	LEAIPMCIPPECAFNKPFVF	17448989	1	HIV
>HIP647	LEAIPMSAPPEVKFNKPFVF	17448989	23.46	HIV
>HIP648	LEAIPMSIAPEVKFNKPFVF	17448989	16.33	HIV
>HIP650	LEAIPMSIPPAVKFNKPFVF	17448989	11	HIV
>HIP651	LEAIPMSIPPEAKFNKPFVF	17448989	10.64	HIV
>HIP652	LEAIPMSIPPEFLFGKPFVF	17448989	1.34	HIV
>HIP654	LEAIPMSIPPEVAFAKPFVF	17448989	3.45	HIV
>HIP655	LEAIPMSIPPEVAFNKPFVF	17448989	4.73	HIV
>HIP656	LEAIPMSIPPEVFFNKPFVF	17448989	0.66	HIV
>HIP657	LEAIPMSIPPEVKANKPFVF	17448989	4.62	HIV
>HIP658	LEAIPMSIPPEVKFAKPFVF	17448989	17.41	HIV
>HIP659	LEAIPMSIPPEVKFNAPFVF	17448989	10.81	HIV
>HIP660	LEAIPMSIPPEVKFNKAFVF	17448989	12.72	HIV
>HIP661	LEAIPMSIPPEVKFNKPAVF	17448989	100	HIV
>HIP662	LEAIPMSIPPEVKFNKPFAF	17448989	100	HIV
>HIP663	LEAIPMSIPPEVKFNKPFVA	17448989	100	HIV
>HIP664	LEAIPMSIPPEVKFNKPFVF	17448989	14.79	HIV
>HIP700	SPAIFQSSMTKILEPFRKQN	15790559	35	HIV

>HIP703	SWKSMAKKLKEYMEKLKQRA	20086159	32.7	HIV
>HIP736	ILGPVLGLVSDTLDDVLGIL	20086159	49.4	HIV
>HIP737	ILGPVLGLVSRTLRRVLGIL	20086159	2.2	HIV
>HIP740	RQRVEELSKFSKKGAAARRRK	20086159	40	HIV
>HIP741	SKEKIGKEFKRIVQRIKDFLR	18591279	10.8	HIV
>HIP742	TRSSRAGLQFPVGRVHRLLRK	20086159	41.1	HIV
>HIP743	FFHHIFRGIVHVGKTIHRLVTG	20086159	2.1	HIV
>HIP746	LLGDLLRKSKEKIGKEFKRIVQR	18591279	35.4	HIV
>HIP748	RPKHPIKHQGLPQEVLNENLLRF	20086159	36.2	HIV
>HIP749	VFQFLGRIIHHVGNFVHGFSHVF	20086159	7.1	HIV
>HIP750	GLNTLKKVFQGLHEAIKLINNHVQ	20086159	35.7	HIV
>HIP751	GLRSKIWLWVLLMIWQESNKFKKM	20086159	31.6	HIV
>HIP752	GLRSRIWLWVLLMIWQESNRFKRM	20086159	1.25	HIV
>HIP755	GAWKNFWSSLRKGFYDGEAGRAIRR	20086159	34.1	HIV
>HIP756	HVDKKVADKVLLLKQLRIMRLLTRL	20086159	3.05	HIV
>HIP757	IWLTALKFLGKHAAKHLAKQQLSKL	20086159	35.2	HIV
>HIP758	NLVSGLIEARKYLEQLHRKLKNRKV	20086159	33.3	HIV
>HIP761	INNYTSLIGSLIEESQNQQEKNEQELLE	US6861059	36.226	HIV
>HIP762	INNYTSLIHSLIEESQNQQEKNEQELLE	US6861059	1.267	HIV
>HIP763	MTLTVQARQLLSQIVQQQNNLLRAIEAQ	US6861059	0.028	HIV
>HIP764	QARQLLSQIVQQQNNLLRAIEAQQHLLQ	US6861059	0.03	HIV
>HIP765	QHLLQLTVWGIKQLQARILAVERYLKDQ	US6861059	0.035	HIV
>HIP766	RQLLSQIVQQQNNLLRAIEAQQHLLQLT	US6861059	0.016	HIV
>HIP767	VQQQNNLLRAIEAQQHLLQLTVWGIKQL	US6861059	0.036	HIV
>HIP768	VWGIKQLQARILAVERYLKDQQLLGIWG	US6861059	0.02	HIV
>HIP770	RMIYKNILFYLVPGPGHGAEPERRNIKYL	12054767	85	HIV
>HIP776	LSELDDRADALQAGASQFETSAAKLKRKYWWKN	12054767	200	HIV
>HIP777	QLLIRMIYKNILFYLVPGPGHGAEPERRNIKYL	12054767	9	HIV
>HIP782	WMEWDREINNYTSLIHSLIEEAQNQQEKNEQELL	19073606	0.002	HIV
>HIP783	WMEWDREINNYTSLIHSLIEELQNQQEKNEQELL	19073606	0.002	HIV

>HIP784	WMEWDREINNYTSLIHSLIEEPQNQQEKNEQELL	19073606	0.046	HIV
>HIP785	WMEWDREINNYTSLIHSLIEETQNQQEKNEQELL	19073606	0.003	HIV
>HIP786	WMEWDREINNYTSLIHSLIEEWQNQQEKNEQELL	19073606	1	HIV
>HIP793	INNYTSLIHSLIEESQNQQEKNEQELLELDKWASL	US6861059	0.065	HIV
>HIP804	DREINNYTSLIHSLIEESQNQQEKNEQELLELDKWA	US6861059	0.023	HIV
>HIP808	EESQNQQEKNEQELLELDKWASLWNWFNITNWLWLI	US6861059	0.255	HIV
>HIP812	ESQNQQEKNEQELLELDKWASLWNWFNITNWLWLIK	US6861059	0.255	HIV
>HIP813	EWDREINNYTSLIHSLIEESQNQQEKNEQELLELDK	US6861059	0.012	HIV
>HIP814	HSLIEESQNQQEKNEQELLELDKWASLWNWFNITNW	US6861059	0.036	HIV
>HIP816	IEESQNQQEKNEQELLELDKWASLWNWFNITNWLWL	US6861059	25.91	HIV
>HIP817	IHSLIEESQNQQEKNEQELLELDKWASLWNWFNITN	US6861059	0.029	HIV
>HIP818	INNYTSLIHSLIEESQNQQEKNEQELLELDKWASLW	US6861059	0.392	HIV
>HIP819	IWNNMTWMEWDREINNYTSLIHSLIEESQNQQEKNE	US6861059	0.593	HIV
>HIP820	KSLEQIWNNMTWMEWDREINNYTSLIHSLIEESQNQ	US6861059	0.067	HIV
>HIP821	LEQIWNNMTWMEWDREINNYTSLIHSLIEESQNQQE	US6861059	0.049	HIV
>HIP822	LIEESQNQQEKNEQELLELDKWASLWNWFNITNWLW	US6861059	0.047	HIV
>HIP823	LIHSLIEESQNQQEKNEQELLELDKWASLWNWFNIT	US6861059	0.003	HIV
>HIP824	LVQPRGPRSGPGPWQGGRRKFRRQRPRLSHKGPMPF	10802050	0.085	HIV
>HIP825	MEWDREINNYTSLIHSLIEESQNQQEKNEQELLELD	US6861059	0.01	HIV
>HIP827	MTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELL	17640899	0.009	HIV
>HIP828	NKSLEQIWNNMTWMEWDREINNYTSLIHSLIEESQN	US6861059	0.078	HIV
>HIP829	NMTWMEWDREINNYTSLIHSLIEESQNQQEKNEQEL	US6861059	0.001	HIV
>HIP830	NNMTWMEWDREINNYTSLIHSLIEESQNQQEKNEQE	US6861059	0.016	HIV
>HIP831	NNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWN	US6861059	0.027	HIV
>HIP832	NQQEKNEQELLELDKWASLWNWFNITNWLWLIKIFI	US6861059	0.253	HIV
>HIP833	NYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNW	US6861059	0.004	HIV
>HIP834	QIWNNMTWMEWDREINNYTSLIHSLIEESQNQQEKN	US6861059	0.001	HIV
>HIP835	QNQQEKNEQELLELDKWASLWNWFNITNWLWLIKIF	US6861059	0.252	HIV
>HIP836	REINNYTSLIHSLIEESQNQQEKNEQELLELDKWAS	US6861059	1.347	HIV
>HIP838	SLEQIWNNMTWMEWDREINNYTSLIHSLIEESQNQQ	US6861059	0.597	HIV

>HIP839	SLIEESQNQQEKNEQELLELDKWASLWNWFNITNWL	US6861059	0.145	HIV
>HIP840	SLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNI	US6861059	0.001	HIV
>HIP841	SQNQQEKNEQELLELDKWASLWNWFNITNWLWLIKI	US6861059	0.256	HIV
>HIP844	TWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLE	US6861059	0.011	HIV
>HIP845	WDREINNYTSLIHSLIEESQNQQEKNEQELLELDKW	US6861059	0.022	HIV
>HIP850	WMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEL	US6861059	0.002	HIV
>HIP854	YTSLIHSLIEEDQNQQEKNEQELLELDKWASLWNWF	19073606	0.21	HIV
>HIP855	YTSLIHSLIEEEQNQQEKNEQELLELDKWASLWNWF	19073606	0.283	HIV
>HIP856	YTSLIHSLIEEFQNQQEKNEQELLELDKWASLWNWF	19073606	0.009	HIV
>HIP857	YTSLIHSLIEEGQNQQEKNEQELLELDKWASLWNWF	19073606	0.001	HIV
>HIP858	YTSLIHSLIEEHQNQQEKNEQELLELDKWASLWNWF	19073606	0.21	HIV
>HIP860	YTSLIHSLIEEKQNQQEKNEQELLELDKWASLWNWF	19073606	0.708	HIV
>HIP863	YTSLIHSLIEENQNQQEKNEQELLELDKWASLWNWF	19073606	0.019	HIV
>HIP864	YTSLIHSLIEEPQNQQEKNEQELLELDKWASLWNWF	19073606	0.446	HIV
>HIP865	YTSLIHSLIEEQQNQQEKNEQELLELDKWASLWNWF	19073606	0.034	HIV
>HIP866	YTSLIHSLIEERQNQQEKNEQELLELDKWASLWNWF	19073606	0.362	HIV
>HIP868	YTSLIHSLIEESQNQQEKLEQELLELDKWASLWNWF	US6861059	0.009	HIV
>HIP870	YTSLIHSLIEESQNQQEKNEQELLELDKPASLWNWF	US6861059	0.137	HIV
>HIP871	YTSLIHSLIEESQNQQEKNEQELLELDKWASLANAA	US6861059	28.6	HIV
>HIP874	YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNAF	US6861059	0.012	HIV
>HIP875	YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNSF	US6861059	0.018	HIV
>HIP880	YTSLIHSLIEESQNQQEKNEQELLELDKWASPWNWF	US6861059	0.004	HIV
>HIP881	YTSLIHSLIEESQNQQEKNEQELLELNKWASLWNWF	US6861059	0.007	HIV
>HIP882	YTSLIHSLIEESQNQQEKNEQELLQLDKWASLWNWF	US6861059	0.003	HIV
>HIP885	YTSLIHSLIEESQNQQEKNQQELLQLDKWASLWNWF	US6861059	0.015	HIV
>HIP889	YTSLIHSLIEEWQNQQEKNEQELLELDKWASLWNWF	19073606	0.029	HIV
>HIP890	YTSLIHSLIEEYQNQQEKNEQELLELDKWASLWNWF	19073606	0.025	HIV
>HIP891	YTSLIHSLIEQSQNQQEKNEQELLELDKWASLWNWF	US6861059	0.005	HIV
>HIP892	YTSLIHSLIQESQNQQEKNEQELLELDKWASLWNWF	US6861059	0.006	HIV
>HIP893	YTSLIHSLIQQSQNQQQKNQQQLLQLDKWASLWNWF	US6861059	0.263	HIV

>HIP895	YTSLIQSLIEESQNQQEKNEQQLLELDKWASLWNWF	US6861059	0.001	HIV
>HIP899	LLGDLLRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES	18591279	1.6	HIV
>HIP904	MTWEAWDRAIAEYAARIEALIRAAQEQQEKNEAALREL	17640899	0.007	HIV
>HIP905	MTWMAWDRAIANYAALIHALIEAAQNQQEKNEAALLEL	17640899	0.022	HIV
>HIP906	MTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEL	17640899	0.008	HIV
>HIP909	TTWEAWDRAIAEYAARIEALIRAAQELQEKLEAALREL	17640899	0.161	HIV
>HIP910	TTWEAWDRAIAEYAARIEALIRAAQELQEKNEAALREL	17640899	0.006	HIV
>HIP912	TTWEAWDRAIAEYAARIEALIRAAQEQQEKLEAALREL	17640899	0.012	HIV
>HIP913	TTWEAWDRAIAEYAARIEALIRAAQEQQEKLEAVLREL	17640899	0.013	HIV
>HIP915	TTWEAWDRAIAEYAARIEALIRAAQEQQEKNEAALREL	17640899	0.005	HIV
>HIP916	TTWEAWDRAIAEYAARIEALIRAAQEQQEKNEAILREL	17640899	0.006	HIV
>HIP917	TTWEAWDRAIAEYAARIEALIRALQELQEKLEAILREL	17640899	5.263	HIV
>HIP918	TTWEAWDRAIAEYAARIEALIRALQELQEKNEAALREL	17640899	0.014	HIV
>HIP919	TTWEAWDRAIAEYAARIEALIRALQELQEKNEAILREL	17640899	0.031	HIV
>HIP920	TTWEAWDRAIAEYAARIEALIRALQEQQEKNEAALREL	17640899	0.009	HIV
>HIP921	TTWEAWDRAIAEYAARIEALIRALQEQQEKNEAILREL	17640899	0.011	HIV
>HIP922	TTWEAWDRAIAEYAARIEALIRASQEQQEKNEAELREL	17640899	0.003	HIV
>HIP924	TTWEEWDREINEYTSRIESLIRESQEQQEKNEQELREL	17640899	0.005	HIV
>HIP945	WEEWDKKIEEYTKKIEELIKKSEEQQKN	19114674	0.002	HIV
>HIP947	GCKKYRRFRWKFKGKFWFWG	22457281	2.9	HIV
>HIP948	GKKYRRFRWKFKFGKWFWFG	22457281	8	HIV
>HIP955	APKEWMEWDREINNYTSLIHSLIKQGI	11118065	5.8	HIV
>HIP956	APKEWMAWAREIAAYAKLIAALI	11118065	1.7	HIV
>HIP957	APKEWMAWAREIAAYAKLIAALIKQGI	11118065	1	HIV
>HIP960	GIGAVLKVLTTGLPALISWIKRKRQQ	9568968	1.2	HIV
>HIP961	KWKLFKKIEKVGQNIRDGIIKAGPAVAVVGQATQIAK	9568968	2.5	HIV

Table S2. Validation dataset used in the development of AVP-IC_{50}Pred prediction models.

AVPdb/HIPdb_ID	Sequence	Reference	IC50 (microM)	Virus
>**AVP0001**	PYVGSGLYRR	1331099	50	AMV
>**AVP0002**	SMIENLEYM	1383569	50	LCMV
>**AVP0104**	LERLDVGTNLGNAIAKLEDAKELLESSDQILRSMK	8700906	7.3	MV
>**AVP0134**	LVFPSDEFDASISQVNEKINQSLAFIRKSDELLHN	8700906	9	RSV
>**AVP0139**	DEFDASISQVNEKINQSLAFIRKSDELLHNVNAGK	8700906	12	RSV
>**AVP0183**	FAVAVKAVAVKAVAVKAVKKAVKKVKKAVKKAVKKKK	12208971	0.75	FIV
>**AVP0190**	KKKKFVKKVAKKVKKVAKKVAKVAVAV	12208971	3	FIV
>**AVP0193**	KRKRFAKKFLRFLRKVIRFLKRFIRRF	12208971	3.69	FIV
>**AVP0425**	GIGKFLHSAKKFGKAFVGEIMNS	15081088	22.16	HSV 1
>**AVP0592**	SWRLDIWDWICESVLDFK	18287023	30	HCV
>**AVP0599**	KWLCRIWSWISDVLDDFE	18287023	0.5	HCV
>**AVP0601**	SWLRDVWDWICTVLTDFK	18287023	3.9	HCV
>**AVP0604**	SWLWEVWDWVLHVLSDFK	18287023	7	HCV
>**AVP0605**	TWLRAIWDWVCTALTDFK	18287023	7.1	HCV
>**AVP0607**	SWLRDVWDWVCTVLSDFK	18287023	3.5	HCV
>**AVP0611**	SWLDDIWDWICEVLSDFE	18287023	4.7	HCV
>**AVP0819**	CFLNITNSHVSILQEAPPLENRV	19114713	1.55	HTLV 1
>**AVP0977**	MDVNPTLLFLKVPAQNAISTTFPYT	19841738	0.002	INFV A
>**AVP0985**	MDVNPTLLFLKVPAQN	19841738	0.046	INFV A
>**AVP0991**	MDVNPTLLFL	19841738	3	INFV A
>**AVP1000**	MDVNPYFLFLKVPAQ	19841738	0.008	INFV A
>**AVP1173**	MANAGLQLLGFILAFLGW	22378192	2.1	HCV
>**AVP1174**	GLQLLGFILAFLGWIGAI	22378192	25	HCV

>AVP1179	GAIVSTALPQWRIYSYAG	22378192	23.8	HCV
>AVP1182	MANAGLQLLGFILA	22378192	25	HCV
>AVP1184	MANAGLQLLGFILAFLGWIG	22378192	17.8	HCV
>AVP1215	GELGRPVYVLGDPGYYAT	22965230	0.228	HCV
>AVP1229	CWVRLGRYLLRRLKTPFTRL	23175359	12	HCV
>AVP1231	DCWVRLGRYLLRRLKTPF	23175359	12	HCV
>AVP1366	DGSLLGRMKGAA	US6544520	3	HBV
>AVP1481	RDVSDFTDSVRDPKTSEILD	US7491489	258.344	SARS-CoV
>AVP1486	IQKEIDRLNEVAKNLNESLI	US7491489	37.128	SARS-CoV
>AVP1510	SWLRDIWDWLCELLSDFK	21801309	0.82	HCV
>AVP1566	SIELNKAKSDLEESKEWIRRSNQKLDSI	16973588	10	HeV
>AVP1734	HRILMRIAQMMT	22743126	60.608	PRRSV
>AVP1763	SVVPSKATWGFA	21176936	14.515	TGEV
>AVP1803	VVTTRLFMSLVASVRNAFQSGYISFDEIIKTE	21518442	8	MDV
>AVP1806	WLVFFVRRKK	22258859	0.638	INFV A
>AVP1813	FNVALDQVFESIENSQALVDQSNRILSSAEKGNTG	17967906	1.52	hMPV
>AVP1940	TDVILMCFSIDSPDALENI	14576104	1.19	RSV
>AVP1947	TDVILMCFSIDSPDSL	14576104	10.86	RSV
>AVP1988	WEDWVAWI	12610147	0.094	FIV
>AVP2010	KPPSKPNNDFHFEVFNFV	11487583	14	RSV
>AVP2036	CNIAPASIVSRNIVYTRAQPNQDIA	9682337	88.299	BRV
>AVP2037	ANVVATYPAHS	10517309	5	HSV 1
>AVP2043	AKKAWKKAKKAAKKAKKWAKK	15498607	40.8	HSV 1
>HIP1130	SGIDQEQNNLTRLIEAQIHELQLTQWKIKQLLARILK	19593361	0.488	HIV
>HIP1131	SGIVQQLNNQLRAEEANQHLEQLSVWGSKQNQARRLK	19593361	0.531	HIV
>HIP1134	HLKTAVQMAVFIHNFKR	12643937	3	HIV
>HIP1136	QETAYFLLKLAGR	12643937	150	HIV
>HIP1139	IHAEIKNSLKIDNLDVNRCIEAL	18331842	25	HIV
>HIP1141	KQLTEAVQKITTESIVIWGK	16879966	7	HIV

>HIP1163	RMKQIEDKIEEIESKQKKIENEIARIKKLIGERY	11572974	25	HIV
>HIP142	AEAMAQVTN	15113844	124	HIV
>HIP163	ACWWAGAKQEF	16854053	333	HIV
>HIP168	ACWWAGIKQEF	16854053	95	HIV
>HIP174	ASWWAGIKQEF	16854053	294	HIV
>HIP178	QLLIRMIYKNI	12054767	21	HIV
>HIP206	TAYFLLKLAGAW	16854053	83	HIV
>HIP229	YQLLIRMIYKAI	12054767	7	HIV
>HIP293	FRKQNPDIVIYQYMD	15790559	119	HIV
>HIP307	IRILQQLLFIHFRIG	17490682	0.7	HIV
>HIP405	KFRRQRPRLSHKGPMPF	10802050	2.619	HIV
>HIP454	GRFKRIRKKLKKLFKKIS	18591279	44	HIV
>HIP628	KVINPEPIVEPFMSKPFALF	17448989	100	HIV
>HIP632	LEAAPMSIPPEVKFNKPFVF	17448989	100	HIV
>HIP642	LEAIPCSIPPCVGFGKPFVF	17448989	0.73	HIV
>HIP643	LEAIPCSIPPCVLFNKPFVF	17448989	0.84	HIV
>HIP649	LEAIPMSIPAEVKFNKPFVF	17448989	9.72	HIV
>HIP753	LLKELWTKIKGAGKAVLGKIKGLL	20086159	1.4	HIV
>HIP801	WMEWDREINNYTSLIGSLIEESQNQQEKNEQELLE	US6861059	0.002	HIV
>HIP810	EINNYTSLIHSLIEESQNQQEKNEQELLELDKWASL	US6861059	0.464	HIV
>HIP811	EQIWNNMTWMEWDREINNYTSLIHSLIEESQNQQEK	US6861059	0.033	HIV
>HIP843	TSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFN	US6861059	0.006	HIV
>HIP944	GRKKRRQRRR	22319541	50	HIV
>HIP951	PKSSWSDHEASSGV	21198428	27	HIV

Table S3. List of 15 best performing physicochemical properties of AVPs using different a) Support Vector Machine, b) Random Forest, c) Instance Based Classifier and d) K-Star

a) List of 15 best performing physicochemical properties of AVPs using Support Vector Machine (SVM)

Property	**Reference**
QIAN880113	Weights for alpha-helix at the window position of 6 (Qian-Sejnowski, 1988)
SUYM030101	Linker propensity index (Suyama-Ohara, 2003)
FINA910102	Helix initiation parameter at position i,i+1,i+2 (Finkelstein et al., 1991)
CHOP780211	Normalized frequency of C-terminal non beta region (Chou-Fasman, 1978b)
MUNV940105	Free energy in beta-strand region (Munoz-Serrano, 1994)
TANS770102	Normalized frequency of isolated helix (Tanaka-Scheraga, 1977)
MUNV940104	Free energy in beta-strand region (Munoz-Serrano, 1994)
PALJ810114	Normalized frequency of turn in all-beta class (Palau et al., 1981)
PALJ810115	Normalized frequency of turn in alpha+beta class (Palau et al., 1981)
QIAN880110	Weights for alpha-helix at the window position of 3 (Qian-Sejnowski, 1988)
AURR980103	Normalized positional residue frequency at helix termini N" (Aurora-Rose, 1988)
ISOY800104	Normalized relative frequency of bend R (Isogai et al., 1980)
MUNV940102	Free energy in alpha-helical region (Munoz-Serrano, 1994)
RACS820110	Average relative fractional occurrence in EL(i-1) (Rackovsky-Scheraga, 1982)
AURR980105	Normalized positional residue frequency at helix termini Nc (Aurora-Rose, 1998)

b) List of 15 best performing physicochemical properties of AVPs using Random Forest (RF)

Property	**Reference**
PALJ810113	Normalized frequency of turn in all-alpha class (Palau et al, 1981)
COSI940101	Electron-ion interaction potential values (Cosic: 1994)
VELV850101	Electron-ion interaction potential (Veljkovic et al, 1985)
GRAR740101	Composition (Grantham: 1974)
RACS820103	Average relative fractional occurrence in AL(i) (Rackovsky-Scheraga: 1982)
YUTK870101	Unfolding Gibbs energy in water: pH7.0 (Yutani et al, 1987)
YUTK870103	Activation Gibbs energy of unfolding: pH7.0 (Yutani et al, 1987)
OLSK800101	Average internal preferences (Olsen: 1980)
YUTK870102	Unfolding Gibbs energy in water: pH9.0 (Yutani et al, 1987)
QIAN880129	Weights for coil at the window position of -4 (Qian-Sejnowski: 1988)
YUTK870104	Activation Gibbs energy of unfolding: pH9.0 (Yutani et al, 1987)
GUOD860101	Retention coefficient at pH 2 (Guo et al, 1986)
BLAS910101	Scaled side chain hydrophobicity values (Black-Mould: 1991)
JOND750101	Hydrophobicity (Jones: 1975)
ARGP820101	Hydrophobicity index (Argos et al, 1982)

c) List of 15 best performing physicochemical properties of AVPs using Instance Based Classifier (IBk)

Property	**Reference**
FINA770101	Helix-coil equilibrium constant (Finkelstein-Ptitsyn, 1977)
FINA910101	Helix initiation parameter at posision i-1 (Finkelstein et al., 1991)
GEIM800101	Alpha-helix indices (Geisow-Roberts, 1980)
GEIM800102	Alpha-helix indices for alpha-proteins (Geisow-Roberts, 1980)
GEIM800104	Alpha-helix indices for alpha/beta-proteins (Geisow-Roberts, 1980)
GEIM800109	Aperiodic indices for alpha-proteins (Geisow-Roberts: 1980)
RACS820110	Average relative fractional occurrence in EL(i-1) (Rackovsky-Scheraga, 1982)
QIAN880113	Weights for alpha-helix at the window position of 6 (Qian-Sejnowski, 1988)
YUTK870101	Unfolding Gibbs energy in water: pH7.0 (Yutani et al, 1987)
QIAN880126	Weights for beta-sheet at the window position of 6 (Qian-Sejnowski: 1988)
QIAN880132	Weights for coil at the window position of -1 (Qian-Sejnowski: 1988)
RICJ880107	Relative preference value at N4 (Richardson-Richardson: 1988)
ROBB760109	Information measure for N-terminal turn (Robson-Suzuki: 1976)
SUYM030101	Linker propensity index (Suyama-Ohara: 2003)
ZIMJ680101	Hydrophobicity (Zimmerman et al, 1968)

d) List of 15 best performing physicochemical properties of AVPs using K-Star (K*)

Property	**Reference**
YUTK870104	Activation Gibbs energy of unfolding: pH9.0 (Yutani et al, 1987)
SUYM030101	Linker propensity index (Suyama-Ohara, 2003)
RACS820110	Average relative fractional occurrence in EL(i-1) (Rackovsky-Scheraga, 1982)
QIAN880129	Weights for coil at the window position of -4 (Qian-Sejnowski: 1988)
QIAN880108	Weights for alpha-helix at the window position of 1 (Qian-Sejnowski: 1988)
MUNV940102	Free energy in alpha-helical region (Munoz-Serrano, 1994)
ISOY800104	Normalized relative frequency of bend R (Isogai et al., 1980)
GEIM800101	Alpha-helix indices (Geisow-Roberts, 1980)
FINA910101	Helix initiation parameter at posision i-1 (Finkelstein et al., 1991)
MUNV940105	Free energy in beta-strand region (Munoz-Serrano, 1994)
FINA770101	Helix-coil equilibrium constant (Finkelstein-Ptitsyn, 1977)
GEIM800102	Alpha-helix indices for alpha-proteins (Geisow-Roberts, 1980)
BULH740102	Apparent partial specific volume (Bull-Breese, 1974)
AURR980105	Normalized positional residue frequency at helix termini Nc (Aurora-Rose, 1998)
CASG920101	Hydrophobicity scale from native protein structures (Casari-Sippl, 1992)

Table S4. SVM and RF parameters used to develop prediction models.

S. No.	Feature	No. of features	Machine learning technique				
			SVM	**Parameters**		**RF**	**Parameters**
			PCC	***g***	***c***	**PCC**	***mtry***
1	Amino acid composition (Mono)	20	0.59	*0.001*	*20*	0.64	*20*
2	Di-peptide composition (Di)	400	0.61	*0.0003*	*30*	0.62	*144*
3	C8 Binary profile (C8 Bin)	160	0.56	*0.1*	*10*	0.60	*14*
4	N8 Binary profile (N8 Bin)	160	0.51	*0.5*	*70*	0.60	*16*
5	Physicochemical properties (Physico)	315	0.59	*0.001*	*50*	0.68	*209*
6	Solvent accessibility (SA)	21	0.22	*0.06*	*10*	0.18	*6*
7	Secondary structure (SS)	3	0.18	*5*	*5*	0.16	*1*
8	**1+2**	420	0.60	*0.001*	*60*	0.62	*210*
9	**3+4**	320	0.59	*0.1*	*150*	0.65	*22*
10	**1+2+5**	735	0.63	*0.001*	*10*	0.64	*212*
11	**3+4+5**	635	0.63	*0.001*	*50*	0.67	*208*
12	**1+2+3+4**	740	0.61	*0.0001*	*500*	0.63	*400*

13	**1+2+3+4+5**	1055	0.62	*0.0003*	*100*	0.64	*430*
14	**6+7**	23	0.22	*0.3*	*1*	0.19	*6*
15	**1+2+5+6+7**	**758**	**0.66**	*0.005*	*100*	**0.68**	*240*
16	**3+4+5+6+7**	**658**	**0.65**	*0.001*	*200*	**0.70**	*280*

Abbreviations SVM: Support vector machine, RF: Random forest, PCC :Pearson Correlation Coefficient

Table S5. Performance of SVM models based on training and validation data composed on randomized instances.

S. No.	Feature	No. of features	PCC obtained using SVM					
			Dataset 1		Dataset 2		Dataset 3	
			Train	**Val**	**Train**	**Val**	**Train**	**Val**
1	Mono aa composition	735	0.59	0.64	0.60	0.58	0.59	0.53
2	Di aa composition	635	0.61	0.62	0.57	0.55	0.60	0.58
3	Mono + Di + C8 Bin + N8 Bin	740	0.61	0.63	0.58	0.53	0.61	0.56
4.	Binary(N8/C8) + Physico + SS + SA	658	0.65	0.70	0.63	0.66	0.61	0.68

Abbreviations SVM: Support vector machine, RF: Random forest, PCC :Pearson Correlation Coefficient , SA: Solvent accessibility, SS :Secondary structure