Bioinformation
www.bioinformation.net

open access
Hypothesis

Identification and characterization of merozoite surface protein 1 epitope

Satarudra Prakash Singh[1,2], Bhartendu Nath Mishra[2*]

[1]Amity Institute of Biotechnology, Amity University, Uttar Pradesh, Gomti Nagar, Lucknow, 226010, India; [2]Department of Biotechnology, Institute of Engineering and Technology, U. P. Technical University, Sitapur Road, Lucknow, 226021, India; Bhartendu Nath Mishra - E-mail: profbnmishra@gmail.com; *Corresponding author

Received January 29, 2009; Revised March 27, 2009; Accepted May 22, 2009; Published August 17, 2009

Abstract:
Malaria is an important tropical infection which urgently requires intervention of an effective vaccine. Antigenic variations of the parasite and allelic diversity of the host are main problems in the development of an effective malaria vaccine. Cytotoxic T lymphocytes (CTL) directed against *Plasmodium falciparum*-derived antigens are shown to play an important role for the protection against malaria. The merozoite surface protein 1 (MSP1) is expressed in all the four life-cycle stages of *Plasmodium falciparum* and did not find any sequence similarity to human and mouse reference proteins. MSP1 is a known target of the immune response and a single CTL epitope binding to the HLA-A*0201 is available for merozoite form. Here, we report the results from the computational characterization of MSP1, precursor (1720 residue) and screening of highest scoring potential CTL epitopes for 1712 overlapping peptides binding to thirty four HLA class-I alleles and twelve HLA class-I supertypes (5 HLA-A and 7 HLA-B) using bioinformatics tools. Supertypes are the clustered groups of HLA class-I molecules, representing a sets of molecules that share largely overlapping peptide binding specificity. The prediction results for MSP1 as adhesin and adhesin-like in terms of probability is 1.0. Results also show that MSP1 has orthologs to other related species as well as having non allergenicity and single transmembrane properties demonstrating its suitability as a vaccine candidate. The predicted peptides are expected to be useful in the design of multi-epitope vaccines without compromising the human population coverage.

Key words: epitope; supertype; vaccine; malaria; bioinformatics

Background:
Malaria infects 500 million people and kills an estimated 2.7 million annually, representing one of the most significant diseases in the world **[1]**. It is caused by four species of the genus *Plasmodium*, in humans, of which *Plasmodium falciparum* is the most virulent that infect and destroy red blood cells **[2]**. *Plasmodium falciparum* strains have evolved resistance against known anti-malarial drugs **[3]** and yet, no effective vaccine is available against malaria for mass production. Malarial vaccine development is probably hampered by the several factors such as multiple stages of the life-cycle, multiple antigens per stage, multiple epitopes per antigen, multiple arms of the immune system, multiple immune responses in different hosts and multiple strains of the parasite **[4]**.

Malarial immunity is a stage specific and the previous strategies are being used to develop a vaccine against each stage–sporozoite, asexual blood stage and sexual stages. A vaccine against a single stage in the life-cycle need to be 100% effective, because parasite that progress to the next stage may express a new set of antigens, which may be unaffected by the vaccine induced immune responses **[5]**. Although, a good number of candidates for such vaccines exist **[6]**, however, it is an accepted view that an effective malaria vaccine need to target all the life-cycle stages of the parasite as well as allelic diversity of the human host.

The recent approaches such as genome sequence analysis, microarray, proteomics and computational vaccinology can be effectively applied for vaccine development of several diseases including malaria **[7]**. In a large-scale, high-accuracy mass spectrometric proteome analysis, 152 proteins were found common to the four stages of *Plasmodium falciparum* life-cycles **[8]**. One of these common proteins, merozoite surface protien 1 (MSP1), normally found in a number of *Plasmodium* species is synthesized as a high-molecular-weight precursor (190 kDa) and then processed into several fragments. At the time of red blood cell invasion by the merozoite, only the 19-kDa C-terminal fragment remains on the cell surface. The full-length MSP1 as well as processing fragments expressed in *Escherichia coli* have been evaluated as vaccines **[9]**. Antibodies against MSP1 also inhibit merozoite entry into red blood cells and immunization with MSP1 protects monkeys from infections **[10]**. Therefore, MSP1 may have importance as a protective immunogen in novel vaccine formulation **[11]**.

Although several $CD8^+$ cytotoxic T lymphocyte (CTL) epitopes have been identified from *Plasmodium falciparum* sporozoite-derived antigens, recently a single epitope has also been described for the merozoite form **[12]**. However, the full range of CTL epitopes present in MSP1 still remains unknown. The precise characterization of the immunogenic sequences from MSP1, involved in the cytotoxic immune response, i.e. the MHC class-I epitopes, is critical for in vitro monitoring of the therapeutically induced anti- *Plasmodium falciparum* CTL responses in patients for peptide-based vaccine development. In this article, we report the characterization of MSP1, precursor antigen as a suitable vaccine candidate followed by identification of the HLA class-I binding epitope using bioinformatics tools. The predicted CTL epitopes may be evaluated as potential components of malaria vaccine formulation that circumvent the problems associated with the host genetic restriction and antigenic variability.

Methodology:
Target protein sequence retrival:
A comprehensive set of the protein complements has been isolated from sporozoites, merozoites, trophozoites and gametocytes of the *Plasmodium falciparum* 3D7. About 46% of all the gene products (2,415 parasite proteins) were detected in all the four stages of *Plasmodium falciparum* life-cycle and only 6% (152 proteins) were common to four stages **[13]**. One of these common proteins, MSP1 was targated for the characterarization as a vaccine candidate and their immunogenic regions analysis. The amino acid sequence of the MSP1 (Orfid: PFI1475w; chromosome: 9) was retrieved from the PlasmoDB database **[14]**.

Vaccine candidate characterization tool:
The vaccine candidate characterization of MSP1 was done using various computational tools such as Vaxijen, MAPP, SPAAN,

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(1): 1-5 (2009)

TargetP, SignalP and TMHMM. The conserved domain and similarity to the human and mouse reference proteins of MSP1 were searched using NCBI's conserved domain database **[15]**. The orthologs of the MSP1 related to the other *Plasmodium* species were also searched using OrthoMCL database **[16]**.

Immunogenic region analysis tool:
Here, we used Immune Epitope Database (IEDB) analysis resources such as ARB, SMM and ANN available at http://tools-int-01.liai.org/analyze/html/mhc_processing.html to predict CTL epitopes for thirty four HLA alleles. The software tool combines predictions of proteasomal cleavage processing, TAP transport and MHC binding to produce an overall score for each peptide as well as intrinsic potential value for CTL epitope. The default setting parameters were used for the immunoproteasome cleavage and TAP transporter predictions. The results of computational analysis included peptide sequences, their start positions and corresponding IC_{50} values. In order to, predict supertype CTL epitopes in MSP1, a web based software tool (NetCTL-2.1) available at http://www.cbs.dtu.dk/services/NetCTL was used. The NetCTL-1.2 server demonstrated a better predictive performance than other available web servers like EpiJen, MAPPP, MHC-pathway and WAPP on all performance measures **[17]**. In order to, classify 1712 overlapping peptides into binders and non-binders to twelve HLA class I supertypes (5 HLA-A and 7 HLA-B), the transformed binding affinity value of ≥ 0.426 (equivalent to affinity ≤ 500 nM) was used. To get, an average optimal predictive performance, the default weight on C terminal cleavage and TAP transport efficiency was used as 0.1 and 0.05 respectively. The peptides with a combined processing score value ≥ default threshold value (0.75) were predicted as potential supertype CTL epitopes.

Discussion:
The characterization of *Plasmodium falciparum* proteome by multidimensional protein identification technology showed that 46% of all gene products (2,415 parasite proteins) were detected in all the four stages of the life-cycle. Almost half (49%) of the sporozoite proteins were unique to this stage and shared on an average 25% of its protein to other stages. On the other hand, trophozoites, merozoites and gametocytes have 20-33% unique proteins and shared their 39-56% proteins. Consequently, only 152 proteins (6%) were common to all four stages of the parasite life-cycle **(Figure 1).** These common proteins were mostly housekeeping proteins such as ribosomal proteins, transcription factors, histones and cytoskeletal proteins **[13]**. One of these common proteins, MSP1 has the potential to become a suitable vaccine candidate which also contained domains that indicate a role in cell-cell interactions **[18]**.

The characterization of the targeted MSP1 antigen as a vaccine candidate was performed using a number of computational tools available in the public domain. The MSP1 was predicted as a subunit vaccine candidate using Vaxijen server with a score of 0.57. The prediction of MSP1 as adhesin and adhesin-like was done using MAAP and SPANN servers based on their prediction probabilities of 1.00 and 0.65, respectively **(Table 1 in supplementary material).** The prediction of localization of MSP1 in a secretory pathway with signal peptide probability of 0.94 and 1.0 was done using TargetP1.1 and SignalP 3.0 servers. MSP1 having single transmembrane spanning region were predicted using TMHMM server. The exhaustive search for orthologs and conserved domains based on homology prediction was performed using OrthoMCL and NCBI's CDD databases. The MSP1 has two orthologs namely *Plasmodium vivax*-Pv099980 and *Plasmodium yoelii*-PY05748. The similarity to the human and mouse reference proteins were also searched using the BlastP server **(Table 1 in supplementary material).** MSP1 did not find any sequence similarity to human and mouse reference database. A vaccine candidate with similar sequences to the host i.e. human or mouse is likely to cause autoimmunity in the host. Non allergenicity of the MSP1 protein was also predicted using Algpred and Allermatch tools. Here, MSP1 demonstrated all the desirable properties of a suitable vaccine candidate **(Table 1 in supplementary material).**

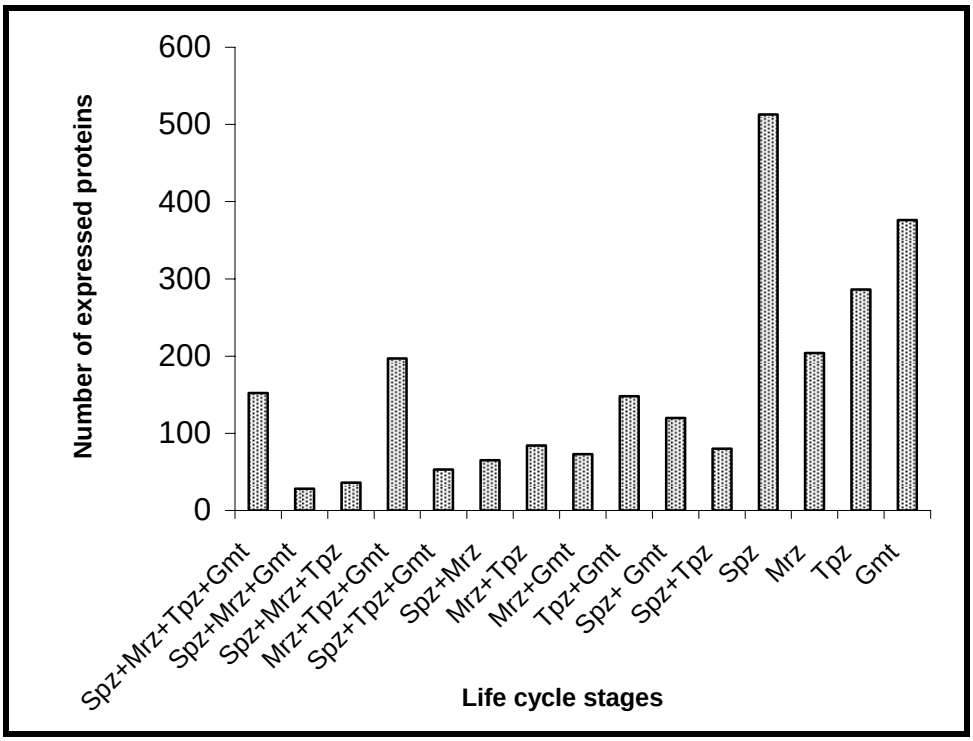

Figure 1: Clusters of proteins expressed in different stages of the *P.falciparum* 3D7 life-cycle (Spz- Sporozoites, Mrz-Merozoites, Tpz-Trophozoites and Gmt-Gametocytes)

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(1): 1-5 (2009)

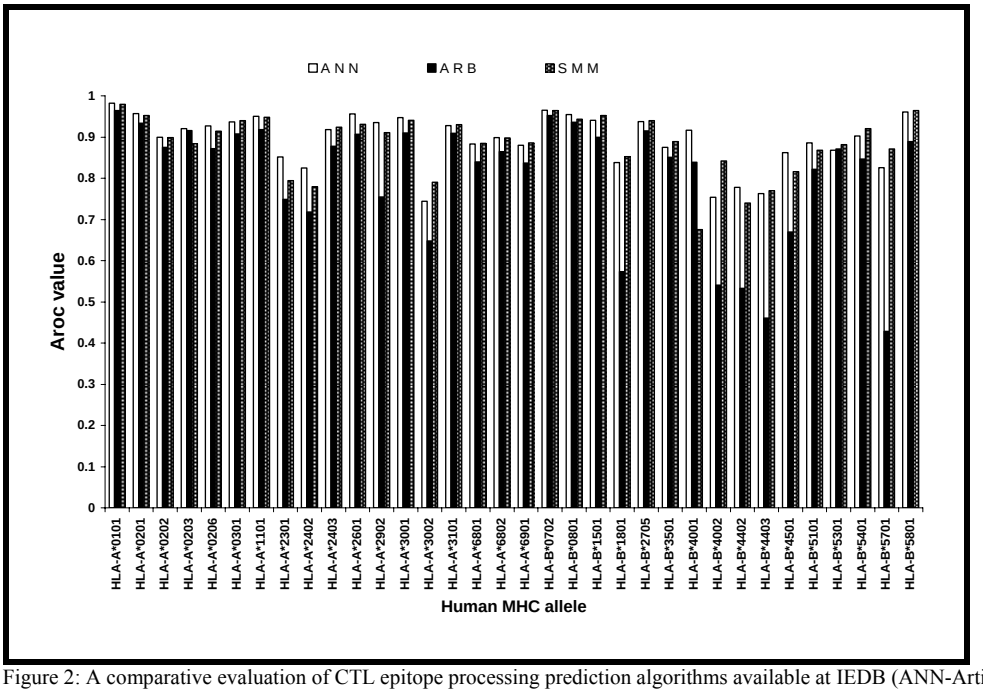

Figure 2: A comparative evaluation of CTL epitope processing prediction algorithms available at IEDB (ANN-Artificial Neural Network, ARB-Average Relative Binding and SMM-Stabilized Matrix Method) for thirty four HLA class-I alleles

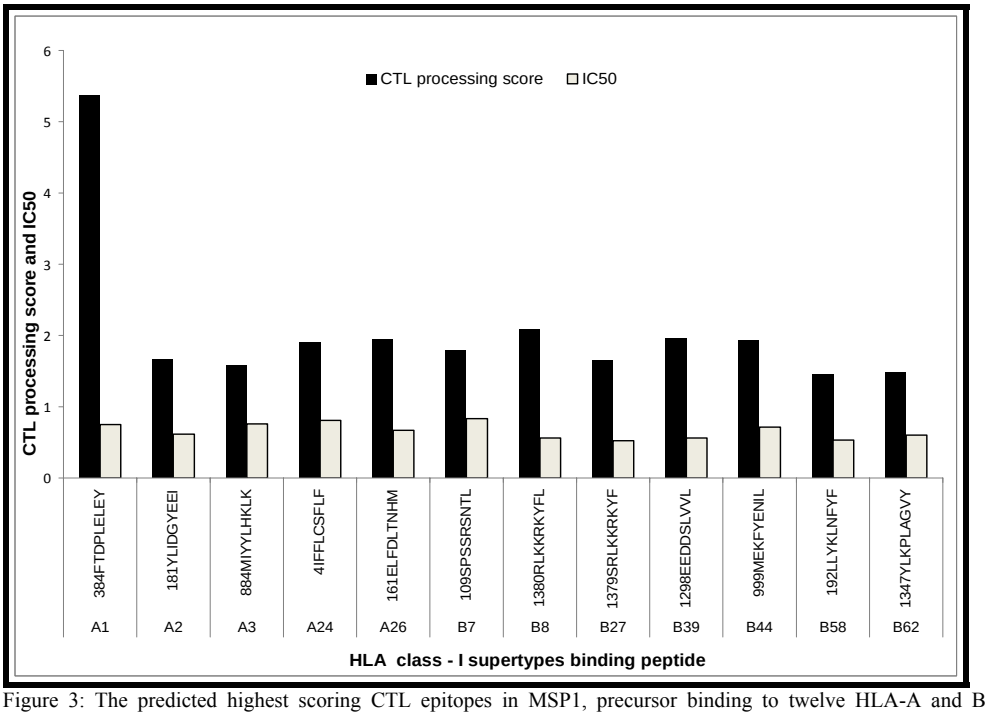

Figure 3: The predicted highest scoring CTL epitopes in MSP1, precursor binding to twelve HLA-A and B supertypes using NetCTL2.1 algorithm**.** Peptides are shown with their start position in the native protein.

A total seventeen CD8+ epitopes have been identified from *P. falciparum* sporozoite-derived antigens and available at IEDB as on 15th March, 2009. The latest one, a single CTL epitope binding to the HLA-A*0201 for merozoite form is additional **[12]**. However, the full range of CTL epitopes binding to other HLA class-I molecules present in MSP1 still remains unknown. Therefore, the binding information regarding thirty four HLA class-I epitopes was predicted using IEDB analysis algorithms like ARB, SMM and ANN, whereas, supertype CTL epitopes were predicted using NetCTL2.1 algorithm. The best performing method for the individual HLA class-I molecules was used for epitope mapping of MSP1 **(Figure 2).** According to the analysis, peptides with the best predicted binding affinities for each HLA class-I molecules are presented in **(Table 2 in supplementary material).** The predicted output is given in units of IC_{50} nM. A lower value of peptide

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(1): 1-5 (2009)

IC_{50} indicates higher affinity towards MHC molecule. As a rough guideline, peptides with IC_{50} values <50 nM, <500 nM and <5000 nM are considered having high affinity, intermediate affinity and low affinity towards MHC molecule respectively. Among all HLA alleles and peptides, the promiscuous peptide FLSSGLHHL (position: 1195-1204) binds with the two HLA alleles A*0202 and A*0203, peptide KIFSARYTY (position: 532-541) binds with four alleles A*0301, A*1101, A*3002 and A*3101 and peptide VPYPNGIVY (position: 419-427) binds with two allele B*3501 and B*5101 **[19]**. In order to, reach for a consensus prediction by other alternative software tools, the predicted highest scoring CTL epitopes by most accurate method were also ranked by two other alternative methods **[20] (Table 1 in supplementary material).**. Most of the highest scoring CTL epitopes, predicted by one method were also ranked within five by other alternative tools. In addition to, HLA class-I allelic CTL epitope prediction, MSP1 was also searched for putative HLA-A and B supertype CTL epitopes. The highest scoring predicted supertype CTL epitopes are shown in the **Figure 3.** In this study, none of the highest scoring predicted CTL epitopes shared sequence identity to human proteins, therefore, they could be included in malaria vaccine design except the peptide sequence IFFLCSFLFFI (position:4-15), which corresponds to the signal peptide.

Conclusion:

Identification of epitopes capable of binding to multiple HLA alleles and supertypes shall significantly rationalize the development of epitope-based vaccines. The present study can be considered as a good example of application of advanced bioinformatics techniques in CTL epitope prediction. The promiscuous and supertype peptide binders allow reducing the time and minimizing the total number of predicted epitopes required for wet lab tests without compromising the population coverage. As per our knowledge, the highest scoring predicted epitopes using IEDB analysis resource and NetCTL are reported first time and could eventually be proposed as a component of a peptide-based anti-malaria vaccine. However, this study has some limitations, as it reports only predicted results and further validation is required through in vitro synthesis of the determined peptides. Besides, in vivo experimental study has to be done, in order to, finally test the efficacy of malaria vaccine.

Acknowledgments:

We are thankful to U.P Technical University, Lucknow and Amity University Uttar Pradesh, Lucknow for their laboratory support to research work.

References:

[1] J Sachs, P Malaney, *Nature* **415:**680 (2002) [PMID: 11832956].
[2] SM Todryk, AV Hill, *Nat. Rev. Microbiol.* **5:**487 (2007) [PMID: 17571459].
[3] R Arav-Boger, TA Shapiro, *Annu. Rev. Pharmacol. Toxicol.* **45:**565 (2005) [PMID: 15822189].
[4] P Brost *et al., Cell* 82:1 (1995) [PMID: 7606774].
[5] T L Richie, A Saul, *Nature* **415:** 694 (2002) [PMID: 11832958].
[6] MP Girard *et al., Vaccine* **25:** 1567 (2007) [PMID: 17045367].
[7] R Rappuoli, *Vaccine* **19:**2688 (2001) [PMID: 11257410].
[8] L Florens *et al., Nature* **419:**520 (2002) [PMID: 12368866].
[9] WR Ballou *et al., Am. J. Trop. Med. Hyg.* **71:**239 (2004) [PMID: 15331843].
[10] MJ Blackman *et al., Exp. Med.* **172:**379 (1990) [PMID: 1694225].
[11] C Putaporntip *et al., Proc. Natl. Acad. Sci. U S A.,* **99:**16348 (2002) [PMID: 12466500].
[12] JP Carralot et al., *Int. Immunol.* **20:**1451 (2008) [PMID: 18794202].
[13] L Florens *et al., Nature* **419:**520 (2002) [PMID: 12368866].
[14] MJ Blackman *et al., Exp. Med.* **172:**379 (1990) [PMID: 1694225].
[15] A Marchler-Bauer *et al., Nucleic. Acids. Res.* **33:** D192 (2005) [PMID: 12520028].
[16] F Chen *et al., Nucleic Acid Res.* **34:** D363 (2006) [PMID: 16381887].
[17] MV Larsen *et al. BMC Bioinformatics* **8:**424 (2007) [PMID: 17973982].
[18] TM Wizemann *Emerg. Infect. Dis.* **5:**395 (1999) [PMID: 10341176].
[19] SP Singh *et al., Online J. Bioinformatics* **7:**69 (2006).
[20] V Wiwanitkit, *Eur. J. Immunol.* **32:**169 (2007).

Edited by P. Kangueane
Citation: Singh & Mishra, Bioinformation 4(1): 1-5 (2009)

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(1): 1-5 (2009)

Supplementary Material

Table 1: Computational characterization of MSP1, precursor as a suitable vaccine candidate

Prediction Characteristics		Status
Vaxijen prediction as a antigen		Yes (0.5706)
MAAP prediction as a adhesin and adhesin-like		Yes (1.00)
SPAAN prediction as a adhesin and adhesin-like		Yes (0.615)
TargetP prediction for subcellular localization	Localization	Secretory Pathway
	Signal peptide probability	0.94
TMHMM prediction for transmembrane regions	Number of TM Helices	1
	Topology	outside1696-1718inside
SignalP prediction for signal peptide	Signal peptide probability	1.00
	Is signal peptide	Yes
Orthologs prediction to other *Plasmodium* species	*P.Vivax*	Yes (Pv099980)
	P. Yoelii	Yes (PY05748)
Allergenicity prediction	Using Algpred	Non Allergen
	Using Allermatch	Non Allergen
Similarity to human and mouse proteins using BlastP		No
Conserved Domain Database hits using BlastP		pfam07462, MSP1_C, CD Length: 574

Table 2: The predicted highest scoring CTL epitopes in MSP1, precursor using IEDB algorithms (ARB, SMM and ANN)

Human MHC Class-I allele	Peptide sequence	Start position	ARB		SMM		ANN	
			MHC affinity (IC_{50})	Total processing Score (Rank)	MHC affinity (IC_{50})	Total processing Score (Rank)	MHC affinity (IC_{50})	Total processing Score (Rank)
A*0101	FTDPLELEY	384	4.5	1.94(1)	15.9	1.40(1)	5.4	1.86(1)
A*0201	FLCSFLFFI	6	0.0	2.60(1)	1.0	1.17(1)	2.5	0.76(6)
A*0202	FLSSGLHHL	1195	0.8	1.98(3)	1.6	1.66(1)	2.7	2.72(1)
A*0203	FLSSGLHHL	1195	0.7	2.00(1)	3.8	1.28(1)	1.39	3.0(1)
A*0206	MQIKKLTLL	1080	10.3	0.82(4)	7.5	0.96(4)	5.2	1.12(1)
A*0301	KIFSARYTY	532	1.0	1.27(1)	304.2	0.43(1)	69.3	1.07(1)
A*1101	KIFSARYTY	532	130.6	0.80(5)	100.7	0.91(1)	1207.4	-0.17(14)
A*2301	FFLCSFLFF	5	3.3	2..08(2)	90.3	0.64(1)	33.1	1.08(1)
A*2402	IFFLCSFLF	4	18.9	1.33(1)	268.0	0.18(1)	na	na
A*2403	VYLKKLDEF	798	0.6	2.6(1)	1.7	2.13(1)	3.3	1.85(1)
A*2601	LLILMLILY	1709	7.2	1.85(1)	205.2	0.4(1)	23.4	3.4(1)
A*2902	YYEKVLAKY	1447	7.6	1.86(4)	34.4	1.20(5)	20.7	1.42(1)
A*3001	KLKKALSYL	563	9.6	0.88(2)	28.8	0.40(5)	4.9	1.17(1)
A*3002	KIFSARYTY	532	2.2	2.57(3)	35.3	1.37(1)	12.2	1.82(1)
A*3101	KIFSARYTY	532	158.6	0.71(4)	251.9	0.51(1)	89.4	0.96(1)
A*3301	DYCQIPFNL	214	1.5	1.80(3)	64.3	0.16(1)	16122	-2.24(116)
A*6801	YTYNVEKQR	538	2.76	0.2(1)	1.23	6.6(1)	na	na
A*6802	ETVGHTTTV	753	0.9	1.29(2)	9.9	0.23(1)	na	na
A*6901	EMIYYLHKL	883	11.9	0.86(3)	51.8	0.22(1)	na	na
B*0702	SPSSRSNTL	109	25.4	0.27(1)	31.7	0.17(1)	3.4	1.15(1)
B*0801	RLKKRKYFL	1380	2.1	1.70(1)	1.8	1.76(1)	3.6	1.46(1)
B*1501	YLKPLAGVY	1347	75.2	0.93(15)	32.0	1.30(1)	42.8	1.17(3)
B*1503	FKHISSNEY	1398	0.5	2.89(1)	0.7	2.78(1)	na	na
B*1517	LSFDLYNKY	1053	156.9	0.56(32)	3.5	2.20(1)	na	na
B*1801	NEYIIEDSF	1404	28.1	1.10(4)	24.8	1.16(1)	8.2	1.64(3)
B*2705	SRLKKRKYF	1379	94.3	0.62(1)	279.9	0.14(1)	135.5	0.46(1)
B*3501	VPYPNGIVY	419	93.5	0.81(10)	64.2	0.98(1)	23.2	1.42(1)
B*4001	KEIAKTIKF	370	86.8	0.76(2)	539.4	-0.04(2)	34.6	1.15(1)
B*4002	KDFNHYYTL	583	27.3	0.63(5)	19.0	0.79(1)	11.6	1.01(1)
B*4402	KEIAKTIKF	370	532.7	-0.03(3)	na	na	60.4	0.9(1)
B*4403	NELLYKLNF	190	57.1	0.73(1)	33.0	0.97(1)	83.6	0.56(4)
B*4501	SEKDFNHYY	581	635.5	-0.28(5)	595.8	-0.25(2)	108.0	0.49(1)
B*5101	VPYPNGIVY	419	687.4	-0.05(6)	2578.6	-0.63(2)	742.1	-0.09(1)
B*5301	NPHNVLQNF	1101	5.6	1.68(1)	83.5	0.51(2)	39.6	0.83(1)
B*5401	IPFNLKIRA	218	15.7	-0.20(3)	14.2	-0.15(1)	25.2	-0.16(2)
B*5701	KALSYLEDY	566	107.0	0.53(1)	361.4	0.01(1)	228.8	0.21(1)
B*5801	ISTTEMEKF	994	139.3	0.44(5)	154.2	0.39(1)	127.5	0.47(2)

na - not available for prediction

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(1): 1-5 (2009)