Interdiscip Sci Comput Life Sci (2011) 3: 36–42
DOI: 10.1007/s12539-011-0055-z

Codon Optimization of the Major Antigen Encoding Genes of Diverse Strains of Influenza A Virus

Indra MANI[1,2†], Vijai SINGH[3†], Dharmendra Kumar CHAUDHARY[2], Pallavi SOMVANSHI[4*], M.P.S. NEGI[5]
[1](Department of Biochemistry, Faculty of Science, Banaras Hindu University, Varanasi-221005, India)
[2](National Bureau of Fish Genetic Resources, Canal ring road, PO-Dilkusha, Lucknow-226002, India)
[3](Epigenomics Project, Universite d'Evry Val d'Essonne-Genopole-CNRS UPS3201 5, rue Henri Desbruères, 91030 Evry, France)
[4](Bioinformatics Centre, Biotech Park, Lucknow-226021, India)
[5](Biometrics division, Central Drug Research Institute, Lucknow-226001, India)

Received 16 January 2010 / Revised 6 May 2010 / Accepted 14 May 2010

Abstract: A large number of influenza A virus outbreaks and mortality occurred in the world recently, an urgent attention to develop effective and sufficient quantity of vaccines are needed. Vaccines are generally protein with immunogenic properties and are not expressed in sufficient quantity because of the codon bias, so it is necessary to optimize its codon in the expression host. Codon optimization was used to improve the protein expression in living organisms by increasing the translational efficiency of gene of interest. Two surface antigenic glycoproteins, hemagglutinin (HA) and neuraminidase (NA) are present in influenza A viruses. We have used HA and NA genes from 19 strains of influenza A viruses for codon optimization in *E. coli*. Both genes of the influenza virus show that the codon adaptation index (CAI) and GC content of the genes in optimized DNA were enhanced significantly (p <0.01) as compared to wild type. CAI and GC of HA in optimized DNA was enhanced by 3.2 (68.5%) and 1.2 (16.2%) fold respectively, while in NA it was increased by 3.3 (69.7%) and 1.2 (15.8%) fold respectively. Our finding demonstrates that the optimized genes could be useful for better expression in host without any truncated proteins and also helpful for protein folding and function. This work provides new insight in the synthetic biology research.
Key words: influenza A virus, codon optimization, CAI, vaccines.

1 Introduction

Influenza viruses are the major public health concern and resulted disease outbreaks occur globally. The virus belongs to the family *Orthomyxoviridae*. Influenza causes epidemics around the globe resulting in 250,000 - 500,000 deaths annually. The most efficient method to prevent these annual outbreaks which resulted in morbidity and mortality is the use of pre-exposure immunization (de Jong and Hein, 2006). The genome of influenza virus is segmented and consists of single stranded negative sense RNA which is pleomorphic in nature; and encodes eight structural proteins and non structural gene (NS1). It was reported that the NS1 gene exhibits evolutionary stability and helps in the replication of influenza viruses (Wan *et al.*, 2007; Somvanshi *et al.*, 2008a). Hemagglutinin (HA) and neuraminidase (NA) are the two surface proteins (Fouchier *et al.*, 2005) and three RNA polymerase (PA, PB1 and PB2). Nucleoprotein (NP) and matrix protein (M1 and M2) play role in the cell cycle, development and pathogenesis. Recently, the NP has been reported as a drug target for controlling the strains of influenza virus spreading in newer vicinity (Somvanshi and Singh, 2010). Sometimes, mutations are caused at the active sites region of NP which leads to influenza resistance against the drugs. Therefore, development of effective and sufficient quantity is required to use for vaccination purpose.

The pathogenicity of influenza virus is multifactorial and is dependent on various surface encoding proteins. HA and NA determine how to recognize the host cell receptor which plays role in host immune response, other proteins are also used to enhance the virulence of highly pathogenic H5N1 strains (Obenauer *et al.*, 2006). The strains of influenza viruses are phylogenetically distinct on the basis of HA and NA genes (Somvanshi *et al.*,

*Corresponding author.
E-mail: psomvanshi@gmail.com
†Authors equally contributed.

2008c). Influenza viruses are categorized into type A and the subtype of influenza A viruses based on antigenic difference between the two surface protein HA and NA. Several subtypes of HA (H1–H16) and NA subtypes (N1–N9) of influenza A viruses have been identified till now (Fouchier *et al.*, 2005). The envelope proteins (HA, NA and M2) of influenza A virus are targeted either by the host antibody or antiviral drugs. These proteins act as a suitable target for the development of anti influenza drugs. HA glycoprotein forms spikes at virions mediating attachment to host cell receptors and subsequently takes entry by membrane fusion whereas NA forms a knob-like structure on the surface of virus particle and catalyzes their releases from infected cells further, allowing the virus to spread. It is an antigenic determination for neutralizing antibodies or antiviral drugs. By catalyzing cleavage of the glycosidic linkage to sialic acid on host cell and virion surfaces, this glycoprotein prevents the aggregation of virions, thus facilitating the release of progeny virus from the infected cells (de Jong and Hein, 2006).

There is always a constant need for suitable diagnosis and vaccination of strain specific influenza A virus to prevent its spreading. The host specific epitopes have been earlier identified in HA and NA of influenza A viruses (Somvanshi *et al.*, 2008b). A potentially important vaccine strategy against viruses, bacteria, and protozoa has been recently emerged in the form of DNA immunization. A successful DNA immunization requires high expression of genes derived from microorganisms in mammalian cells; but the main hindrance for gene expression may be due to the inter-specific difference of codon usage (Koide *et al.*, 1998). In fact, DNA vaccines are plasmids capable of expressing antigenic peptides within the host (Babiuk *et al.*, 2000; Babiuk *et al.*, 2003). These are used as an attractive alternative to conventional vaccines as they generate both a cellular and a humoral immune response; the combined response has proven an effective method in combating intracellular pathogens. There are different ways to optimize the DNA vaccine efficiency with the choice of antigen. It has been successfully enhanced by codon optimization (Uchijima *et al.*, 1998). There are many software tools and technologies which have been developed for gene expression studies and predicting the expression level of genes through computational methods, this is appealing as expensive and difficult experiments are not required (Roymondal *et al.*, 2009). The synonymous codon usage of any organism is almost calculated and varies not only among genomes but also among genes of a given genome (Ikemura, 1985).

In most synonymous cases codons are not used in the same frequencies; it is known as codon usage bias. Codon bias is generally controlled by a balance between the mutation, natural selection and genetic drift. If gene contains the codon is rarely used by host, its expression level will be maximized in heterologous system (Gustafsson *et al.*, 2004; Fuglsang, 2003; Basak *et al.*, 2008). Therefore, on the basis of codon this is used in the abundant amount in the host. It is selected for optimization of codon for the two surface target genes. The aim of this study is to optimize the codon for over expression of different strains of influenza A virus genes of hemaglutinin and neuraminidase in *E. coli* for adequate amount of vaccines. It was assumed that the studied genes may be useful for increasing the expression level of proteins to ensure efficient production for research and clinical trials without any bias. It was proposed that specifical codons were changed without any changes of amino acids sequences. Thus, the antigenicity and functional activity of each protein are exactly similar to its native type.

2 Materials and methods

2.1 Collection and alignment sequences

Nucleotide sequences of different strains of influenza virus were retrieved from http://www.ncbi.nlm.nih.gov/. The relatedness of sequences deposited in databases was evaluated using the BLAST (Altschul *et al.*, 1997).

2.2 Codon optimization and analysis

The given DNA sequences were arranged into triplets (codons), the triplets were replaced with new ones and generated with a given frequency distribution. In this process amino acid was same, but codon of low frequency of an amino acid was replaced with codon of high frequency, according to desired species frequency distribution. Gene designer (https://www.dna20.com / index.php?pageID = 220) was used for designing and simulation of genes in a given expression hosts, Optimizer was used for optimization and calculation of CAI, G+C and A+T (Puigbo *et al.*, 2007), CAIcal and MrGene were used for optimization of DNA sequences at maximum suitable threshold level. Codon optimization of genes were performed on 10–15% threshold level of host cellular codons. Codon adaptation index (CAI) was also calculated for each gene. It is widely acceptable as an effective measure of potential level of gene expression (Sharp and Li, 1987). Codon optimization is a technique to exploit the protein expression in living organism by increasing the translational efficiency of gene of interest by transforming DNA sequence of nucleotides of one species into DNA sequence of nucleotides of another species like plant sequence to human sequence, human sequence to bacteria or yeast sequences.

2.3 Statistical analysis

The CAI, GC and AT of HA and NA of influenza virus were compared using Wilcoxon matched pairs test. A two-tailed ($\alpha = 2$) probability $p < 0.05$ were considered to be statistically significant. STATISTICA

(version 7.0) was used for the analysis of genes of influenza virus.

2.4 Nucleotide sequence accession number

The nucleotide sequences of the codon optimized HA of influenza A virus have been submitted to NCBI GenBank under different accession numbers from GQ483315 to GQ483330.

3 Results and discussion

The remarkable quantity of sequence data that have become accessible in recent times opens a new research potential for new diagnostic techniques. In the current study, we have used the codon optimization method for adequate quantity of protein production in desire host. There is an urgent need to diagnose and control the infection of influenza globally using the require amount of vaccine doses. It is not available easily so we tried to investigate the new type of synthetic genes to maximize the expression level. There are several advantages of the codon optimization as genes takes complete machinery for expression of gene from host. Match codon frequencies in target and host organisms to ensure proper folding and bias GC content to increase mRNA stability or reduce secondary structures. Modify ribosome binding sites and mRNA degradation sites and also adjust translational rates to allow various domains of the protein to fold properly.

The codon usage for hemaglutinin and neuraminidase of nineteen influenza virus strains were summarized in Table 1 and Table 2 respectively. The CAI of HA and NA from wild type of DNA sequences were selected and its codon was optimized with reference to *E. coli* as it is a popular host for heterologous gene expression. In this study, we have observed the CAI values of HA and NA was more comparison to wild type of sequences. The revolution of molecular biology gains momentum, as different number and variety of natural genes have been re-designed at the nucleotide level and synthesized in attempts to improve protein yields (Gustafsson *et al.*, 2006). Codon distribution acts in respect to GC content of genome and the changes in codon usage are at least partly explained by mutation-selection equilibrium between the different synonymous codons in each organism (Knight *et al.*, 2001).

Table 1 Hemagluttinin (HA) gene of Influenza A virus expression level in *E. coli* of wild type and codon optimized sequences

Influenza strains	Wild type DNA			Optimized DNA		
	CAI	GC %	AT %	CAI	GC %	AT %
H1N1	0.277	40.7	59.3	0.799	49.6	50.4
H5N1	0.260	41.0	59.0	0.799	49.9	50.1
H3N8	0.239	39.5	60.5	0.813	49.3	50.7
H3N2	0.284	42.1	57.9	0.798	50.0	50.0
H9N2	0.243	41.8	58.2	0.835	50.6	49.4
H10N8	0.256	43.4	56.6	0.810	51.5	48.5
H10N7	0.254	43.4	56.6	0.814	50.9	49.1
H7N7	0.220	44.0	56.0	0.797	51.9	48.1
H7N2	0.222	43.7	56.3	0.825	52.6	47.4
H11N9	0.290	41.6	58.4	0.803	49.8	50.2
H7N3	0.230	42.2	57.8	0.794	50.8	49.2
H4N3	0.275	43.5	56.5	0.824	51.4	48.6
H1N2	0.267	41.2	58.8	0.811	49.9	50.1
H5N3	0.255	41.5	58.5	0.807	49.7	50.3
H13N2	0.256	42.3	57.7	0.830	51.9	48.1
H14N5	0.273	44.2	55.8	0.797	51.3	48.7
H6N8	0.253	42.4	57.6	0.833	50.1	49.9
H7N8	0.243	43.7	56.3	0.807	50.4	49.6
H11N2	0.255	42.2	57.8	0.811	48.4	51.6
N	19	19	19	19	19	19
Min	0.220	39.50	55.80	0.794	48.40	47.40
Max	0.290	44.20	60.50	0.835	52.60	51.60
Mean ± SD	0.255 ± 0.020	42.34 ± 1.27	57.66 ± 1.27	0.811 ± 0.013**	50.53 ± 1.05**	49.47 ±1.05**

**– p <0.01: in comparison with wild type.

Table 2 Neuraminidase (NA) gene of Influenza A virus expression level in *E. coli* of wild type and codon optimized sequences

Influenza strains	Wild type DNA			Optimized DNA		
	CAI	GC %	AT %	CAI	GC %	AT %
H1N1	0.254	42.0	58.0	0.798	50.1	49.9
H5N1	–	–	-	–	–	–
H3N8	0.240	44.6	55.4	0.757	53.3	46.7
H3N2	0.257	42.9	57.1	0.776	49.5	50.5
H9N2	0.247	43.5	56.5	0.775	52.6	47.4
H10N8	0.253	44.5	55.5	0.775	52.8	47.2
H10N7	–	–	–	–	–	–
H7N7	0.251	42.9	57.1	0.779	52.2	47.8
H7N2	–	–	–	–	–	–
H11N9	0.216	43.0	57.0	0.795	52.7	47.3
H7N3	0.239	43.0	57.0	0.780	49.9	50.1
H4N3	0.230	43.0	57.0	0.815	51.2	48.8
H1N2	–	–	–	–	–	–
H5N3	0.234	43.4	56.6	0.790	50.1	49.9
H13N2	–	–	–	–	–	–
H14N5	–	–	–	–	–	–
H6N8	0.197	44.9	55.1	0.802	52.8	47.2
H7N8	–	–	–	–	–	–
H11N2	–	–	–	–	–	–
N	19	19	19	19	19	19
Min	0.197	42.00	55.10	0.752	49.80	46.70
Max	0.257	44.90	58.00	0.815	53.30	50.50
Mean ± SD	0.238 ± 0.018	43.43 ± 0.89	56.57 ± 0.89	0.786 ± 0.016**	51.56 ± 1.42**	48.44 ± 1.42**

**– $p < 0.01$: in comparison with wild type.

In the present study, we have considered the CAI, GC and AT frequencies in all 19 strains of wild type HA ranged from 0.220 to 0.290, 39.50 to 44.20 and 55.80 to 60.50 respectively with an average (± SD) of 0.255 ± 0.020, 42.34 ± 1.27 and 57.66 ± 1.27 respectively. The respective frequencies of these in optimized DNA ranged from 0.794 to 0.835, 48.40 to 52.60 and 47.40 to 51.60 respectively with an average (± SD) of 0.811 ± 0.013, 50.53 ± 1.05 and 49.47 ± 1.05 respectively. On comparing the mean, CAI, GC and AT of all HA strains of optimized DNA was found to be significantly ($p < 0.01$) different and higher than the respective values of wild type. The mean CAI, GC and AT in optimized DNA were 3.2 (68.5%), 1.2 (16.2%) and 0.9 (−16.6%) fold higher than the respective mean values of wild type. In this study, we have observed that the CAI values of HA of the studied strains of influenza virus were more as compared to wild type sequences. The CAI value was shown at Y-axis, while the number of studied genes on X-axis (Fig. 1). The construction of an H5 HA gene, optiHA, containing chicken biased codons based on the HA amino acid sequence of the highly pathogenic H5N1 virus. The experiment was conducted to assess the dose of the pCAGGoptiHA vaccine effectiveness; result indicates that two doses of 10 microgram of pCAGGoptiHA could induce complete protection in chickens against H5 lethal virus challenge. Based on our results, we conclude that construction optimization could dramatically increase the H5 HA gene DNA vaccine efficacy in chickens, and therefore, greatly decrease the dose necessary for inducing complete protection in chickens (Jiang *et al.*, 2007). Improvement of hemagglutinin (HA) expression of influenza virus has been optimized and synthesized the whole length of HA gene of H5N1 in accordance with the human's codon and inserted it to the eukaryotic expression vector pDC315 to construct a eukaryotic expression vector shows the better expression than wild type of gene (Li *et al.*, 2008).

Similarly, the CAI, GC and AT frequencies in all 11 strains of wild type NA ranged from 0.197 to 0.257, 42.00 to 44.90 and 55.10 to 58.00 respectively with an average (± SD) of 0.238 ± 0.018, 43.43 ± 0.89 and 56.57 ± 0.89 respectively. The respective frequencies

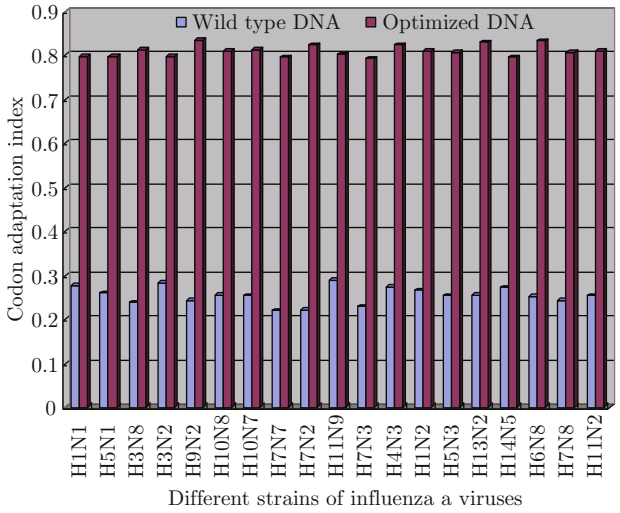

Fig. 1 Comparison of the wild and the optimized types of DNA sequences of hemagglutinin gene in different strains of influenza A viruses.

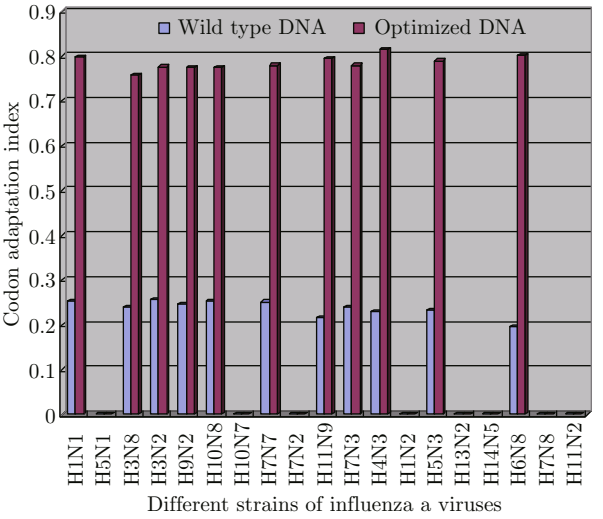

Fig. 2 Comparison of the wild and the optimized types of DNA sequences of neuraminidase gene in different strains of influenza A viruses.

of these in optimized DNA ranged from 0.757 to 0.815, 49.50 to 53.30 and 46.70 to 50.50 respectively with an average (± SD) of 0.786 ± 0.016, 51.56 ± 1.42 and 48.44 ± 1.42 respectively. On comparing the mean, the mean CAI, GC and AT of all NA strains of optimized DNA was found to be significantly ($p < 0.01$) different and higher than the respective values of wild type. The mean CAI, GC and AT in optimized DNA were 3.3 (69.7%), 1.2 (15.8%) and 0.9 (−16.8%) fold higher than the respective mean values of wild type. In this study, the CAI values of NA of the studied strains were more as compared to native type of sequences. The CAI value was shown at Y-axis, while the number of studied genes on X-axis (Fig. 2). Improving expression of antigen is critical to the immunogenicity of DNA vaccines. The modification of NDV F48E9 strain HN gene by optimizing the codon usage and inserting the secretary leader sequence. By optimizing codon usage in transiently transfected 293T cells, expression levels of HN gene were higher from the codon-optimized gene than the counterpart (He *et al.*, 2008). Recently, the expression of aerolysin (Singh *et al.*, 2010a) and hemolysin of *A. hydrophila* have been reported in *E. coli* (Singh *et al.*, 2009). Furthermore, another study demonstrates that the codon optimization of virulence proteins such as hemolysin, aerolysin and lipase of *A. hydrophila* for over expression in *E. coli* without codon bias. The CAI values of these genes were 2-fold greater than the wild type of genes (Singh *et al.*, 2010b). The level of gene expression of eukaryotic genes introduced into mammalian cells depends on many factors viz. gene copy number, transcriptional control elements, site of chromosomal integration, mRNA stability and translational efficiency (Gross and Hauser, 1995).

In *E. coli* the heterologous protein production efficiency was increased with the use of production of synthetic genes. The expression patterns of 30 human short-chain dehydrogenase/reductase genes (SDRs) in *E. coli* have been analyzed and the native and synthetic (codon-optimized) version of each gene were compared (Burgess- Brown *et al.*, 2008). DNA immunization by the gene codon-optimized to mammals in the entire region is successful. As the interspecific difference of codon usage for effective induction of specific immune responses against several bacteria and protozoa is one of the main problems. It has been successfully done in the genes encoding MHC class I-restricted CTL epitopes of*Listeria monocytogenes* and *Plasmodium yoelii* (Nagata *et al.*, 1999).

4 Conclusion

The wild type frequencies of hemaglutinin and neuraminidase DNA of influenza virus strains showed significantly (p <0.01) different codons and a good correlation between the strains based on statistical analysis was observed. These optimized DNA will be chemically synthesized and over expressed in *E. coli* as compare to its wild type counterparts. The expression of both the gene could be more due to the high CAI value as compare to wild type DNA. These have numerous applications like to remove the stop codons, to clone, in custom design of synthetic genes, to improve the functionality of genes, to increase protein expression level and for lower production costs, in drug development. Furthermore, studies will be needed for *in vitro* validation of codon optimized DNA as the studies provide more information of high level of expression and that could be

used in adequate amount for vaccine production.

References

[1] Altschul, S.F., Thomas, L.M., Alejandro, A.S., Jinghui, Z., Zheng, Z., Miller, W., Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402.

[2] Babiuk, L.A., Babiuk, S.L., Loehr, B.I., van Drunnen Littel-van den Hurk S. 2000. Nucleic acid vaccines: Research tool or commercial reality. Vet Immunol Immunopathol 76, 1–23.

[3] Babiuk, L.A., Pontarollo, R., Babiuk, S., Loehr, B., van Drunen Littel-van den Hurk S. 2003. Induction of immune responses by DNA vaccines in large animals. Vaccine 21, 649–658.

[4] Basak, S., Mukherjee, I., Choudhury, M., Das, S. 2008. Unusual codon usage bias in low expression genes of *Vibrio cholerae*. Bioinformation 3, 213–217.

[5] Burgess-Brown, N.A., Sharma, S., Sobott, F., Oppermann, C.L.U., Gileadi, O. 2008. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. Protein Expr Purif 59, 94–102.

[6] de Jong, M.D., Hein, T.T. 2006. Avian influenza A (H5N1). J Clin Virol 35, 2–13.

[7] Fouchier, R.A., Munster, V., Wallensten, A., Bestebroer, T.M., Herfst, S., Smith, D., Rimmelzwaan G.F., Olsen B., Osterhaus A.D. 2005. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. J Virol 79, 2814–2822.

[8] Fuglsang, A. 2003. Codon optimizer: A freeware tool for codon optimization. Protein Expr Purif 31, 247–249.

[9] Gross, G., Hauser, H. 1995. Heterologous expression as a tool for gene identification and analysis. J Biotechnol 41, 91–110.

[10] Gustafsson, C., Govindarajan, S., Minshull, J. 2004. Codon bias and heterologous protein expression. Trends Biotechnol 22, 346–353.

[11] He, S., Shi, X., Wang, Y., Wang, M., Ran, D., Tong, G. 2008. Effect of modified NDV F48E9 strain HN gene and *in vitro* expression of its DNA vaccine. Sheng Wu Gong Cheng Xue Bao 24, 226–231.

[12] Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol 2, 13–34.

[13] Jiang, Y., Yu, K., Zhang, H., Zhang, P., Li, C., Tian, G., Li, Y., Wang, X., Ge, J., Bu, Z., Chen, H. 2007. Enhanced protective efficacy of H5 subtype avian influenza DNA vaccine with codon optimized HA gene in a pCAGGS plasmid vector. Antiviral Res 75, 234–241.

[14] Knight, R.D., Freeland, S.J., Landweber, L.F. 2001. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol 2, RESEARCH 0010.1-0010.13.

[15] Koide, Y., Nagata, T., Yoshida, A., Uchijima, M. 1998. Optimization of codon usage of plasmid DNA vaccine is required for the effective MHC class I restricted T cell responses against an intracellular bacterium. J Immunol 161, 5594–5599.

[16] Li, K.B., Zhang, X.G., Ma, J., Jia, X.J., Wang, M., Dong, J., Zhang, X.M., Xu, H., Shu, Y.L. 2008. Codon optimization of the H5N1 influenza virus HA gene gets high expression in mammalian cells. Bing Du Xue Bao 24, 101–105.

[17] Nagata, T., Uchijima, M., Yoshida, A., Kawashima, M., Koide, Y. 1999. Codon optimization effect on translational efficiency of DNA vaccine in mammalian cells: Analysis of plasmid DNA encoding a CTL epitope derived from microorganisms. Biochem Biophys Res Commun 261, 445–451.

[18] Obenauer, J.C., Denson, J., Mehta, P.K., Sa, X., Mukatora, S., Finkelstein, D.B., Xu X., Wang, J., Ma, J., Fan, Y., Rakestraw, K.M., Webster, R.G., Hoffmann, E., Krauss, S., Zheng, J., Zhang, Z., Naeve. C.W. 2006. Large scale sequences analysis of avian influenza isolates. Science 311, 1576–1580.

[19] Puigbo, P., Guzman, E., Romeu, A., Garcia-Vallve, S. 2007. OPTIMIZER: A web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res 35, W126–W131.

[20] Roymondal, U., Das, S., Sahoo, S. 2009. Predicting gene expression level from relative codon usage bias: an application to *Escherichia coli* genome. DNA Res 16, 13–30.

[21] Sharp, P.M., Li, W.H. 1987. The codon Adaptation Index - a measure of directional synonymous codon usage bias and its potential applications. Nucleic Acids Res 15, 1281–1295.

[22] Singh, V., Somvanshi, P., Rathore, G., Kapoor, D., Mishra, B.N. 2009. Gene cloning, expression and homology modelling of hemolysin gene from *Aeromonas hydrophila*. Protein Expr Purif 65, 1–7.

[23] Singh, V., Somvanshi, P., Rathore, G., Kapoor, D., Mishra, B.N. 2010a. Gene cloning, expression and characterization of recombinant aerolysin from *Aeromonas hydrophila*. Appl Biochem Biotechnol 160, 1985–91.

[24] Singh, V., Chaudhary, D.K., Mani, I., Somvanshi, P., Rathore, G., Sood, N. 2010b. Molecular identification and codon optimization analysis of major virulence encoding genes of *Aeromonas hydrophila*. African J Microbiol Res 4, 952–957.

[25] Somvanshi, P., Singh, V. 2010. Homology modelling and identification of catalytic residue in the nucleoprotein of Influenza A virus. Int J Med Eng Informatics 2, 26–36.

[26] Somvanshi, P., Singh, V., Arshad, M. 2008a. Modeling of RNA secondary structure of non structural gene and evolutionary stability of the Influenza virus through*in silico* methods. J Proteom Bioinform 1, 219–226.

[27] Somvanshi, P., Singh, V., Seth, P.K. 2008b. Prediction of epitopes in hemagglutinin and neuraminidase proteins of influenza A virus H5N1 strain: A clue for diagnostic and vaccine development. J Integrative Biol 12, 61–69.

[28] Somvanshi, P., Singh, V., Seth, P.K. 2008c. Phylogenetic and computational proteome analysis of Influenza A virus subtype H5N1. Internet J Genom Proteom 3.

[29] Uchijima, M., Yoshida, A., Nagata, T., Koide, Y. 1998. Optimization of codon usage of plasmid DNA vaccine is required for the effective MHC class I-restricted T-cell responses against an intracellular bacterium. J Immunol 161, 5594–5599.

[30] Wan, X.F., Wu, X., Lin, G., Holton, S.B., Desmone, R.A., Shyu, C.R., Guan, Y., Emch, M.E. 2007. Computational identification of re-assortments in avian influenza viruses. Avian Dis 51, 434–439.