CODON OPTIMIZATION OF THE POTENTIAL ANTIGENS ENCODING GENES FROM *MYCOBACTERIUM TUBERCULOSIS*

Indra Mani[1, 2], Dharmendra Kumar Chaudhary[2], Pallavi Somvanshi[3], Vijai Singh[4*]

E-mail: vijaisingh15@gmail.com

[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]Bioinformatics Centre, Biotech Park, Sector-G, Jankipuram, Lucknow-226021, India.
[4]Epigenomics Project, Universite d'Evry Val d'Essonne-Genopole-CNRS UPS3201,5 rue Henri Desbruères, 91030 Evry, France.

ABSTRACT: Current approach for controlling of tuberculosis is going on by recommended doses of vaccines. Codon optimization and simulation techniques are used to improve the protein expression in living organism by increasing their translational efficiency of gene of interest. We have designed; optimized the codon and simulated in nineteen indigenous genes of *Mycobacterium tuberculosis* H37Rv in the *Escherichia coli*. We minimized the G+C content in optimized genes from 64.75% to 59.67% of the studied genes as the richness of G+C content is reflected in a strong bias. CAI and AT of optimized DNA were enhanced by 1.9 (47.8%) and 1.1 (12.5%) fold more with respect to its native type. Our finding indicates the optimized genes can be useful for over expression in host and the study provides a new insight for the emerging research in synthetic biology.

KEY WORDS: *Mycobacterium tuberculosis*, codon bias, optimization, CAI, Vaccines.

INTRODUCTION

Tuberculosis (TB) is a leading infectious disease that is responsible for more than one-quarter of the world's preventable deaths (Cole, et al., 19981; Kaufmann, 2001). The global incidence of TB varies from less than 10 per 100,000 in North America, to 100–300 per 100,000 in Asia and Western Russia, to over 300 per 100,000 in Southern and Central Africa. Total 8 million individuals in the world suffered with this disease annually due to negligence in proper treatment, about 60% of patients were died (Smith, 2003). However, childhood immunization with *Mycobacterium bovis* (BCG) has likely reduced several severe diseases i.e., tuberculosis and meningitis. Its effectiveness against adult pulmonary disease is highly variable with effective estimation ranging from 0% to 80%. But the adult pulmonary disease is the main threat posed to public health by tuberculosis (Fine, et al., 1999).

Therefore, the drugs and vaccines are needed for prevention and cure of tuberculosis. The drugs are mainly targeted at active site region in the proteins. The highly bacterial pathogen like *Mycobacterium tuberculosis* is developed the resistance against the drug because active site region are mutated and resultant changes in the amino acids. Earlier, several proteins have been targeted for searching of potent drugs against tuberculosis namely 3-oxoacyl-acyl carrier protein synthase II (Singh and Somvanshi, 2009a), NAD^+ dependent DNA ligase (Singh and Somvanshi, 2010), Methionine sadenosyl transferase (Khedkar, et al., 2005) and Peptide deformylase (Singh and Somvanshi, 2009b). These proteins are essential major component for cell viability and cell division of *M. tuberculosis*. This bacterium has been developed resistance against the drugs so it is difficult to target the proteins. Recently, new advancement in understanding the mechanism of drug-protein interaction of *M. tuberculosis* and stabilised functional genomic techniques like microarray and proteomics in combination with modern approach is available. But there is no new and potent drug has developed in the past three decades against tuberculosis (Chopra, et al., 2003). Thus, there is a need to develop antigen based effective vaccine for the prevention and cure of tuberculosis. An efficient vaccination and chemotherapeutic treatment usually takes a long time, more expensive and not available in many parts of the world. So it is an urgent need to produce the effective and an adequate amount of vaccine candidates against *M. tuberculosis*.

The cloning and expression of antigenic proteins coding genes are generally used for production of adequate quantity of proteins. Generally in to the prokaryotic system *E. coli* used for expression of desired genes but G+C % is high in the protein coding genes of *M. tuberculosis* comparatively to *E. coli* leads to codon bias (Andersson and Sharp, 1996). However, the culture, growth and maintenance of *M. tuberculosis* strains are difficult due to its high level of pathogenicity.
For this required high level of bio-safety facilities which is extremely expensive and requires well-trained manpower. Here it is an interesting and novel approach for the expression of same proteins without the need of *M. tuberculosis* strains. In this approach chemically synthesized the optimized genes and clones in to the expression vector then subsequently express in *E. coli*. Therefore, *in silico* tools are highly useful for the designing and optimization of gene sequences which encodes the same amino acids

Moreover recent advancement in the field of bioinformatics and computational biology provide a detail statistical approach used to analyze codon usage bias. The frequency of optimal codon and CAI (Codon adaptation index) are used to predict levels of gene expression (Comeron and Aguade, 1998). Human peptide deformylase (hPDF) situated in the mitochondria has been become a promising target for anti-cancer therapies whereas expression of the hPDF gene in *E. coli* was not more efficient because of high levels of GC content and also the presence of rare codons (Han, et al., 2010). The native and codon optimized lipase gene (lipJ08) from *Candida rugosa* ATCC14830 genes have been expressed in *Pichia pastoris*. The hydrolytic activity of the recombinant lipJ08 was observed higher than the wild-type (Xu, et al., 2010). So, codon bias occurs during the expression of genes.

Gene expression levels depend on many factors such as promoter sequences and regulatory elements. One of the most important factors is the adaptation of the codon usage of the transcript gene to the typical codon usage of the host (Lithwick and Margalit, 2003). Therefore, highly expressed genes in prokaryotic genomes under translational selection have a pronounced codon usage bias. This is because they use a small subset of codons that are recognized by the most abundant tRNA species (Ikemura, 1981). The force that modulates this codon adaptation is called translational selection and its strength is important in fast-growing bacteria (Rocha, 2004; Sharp, et al., 2005). If a gene contains codons that are rarely used by the host, its expression level will not be maximal. This may be one of the limitations of heterologous protein expression (Gustafsson, 2004) and the development of DNA vaccines (Ivory and Chadee 2004). CAI is widely used in biological research for characterizing gene expression in general and translation efficiency (Xia, 2007). Other than its primary use for measuring the efficiency of translation elongation, it has been used to study of functional conservation of gene expression across different microbial species (Lithwick and Margalit, 2005), to predict protein production (Futcher, 1999; Gygi, et al., 1999), and to optimize DNA vaccines (Ruiz, 2006).

Therefore, codon optimization is an important aspect to express an adequate amount of proteins without bias. In this study, we have designed, optimized and simulated the gene sequences of desired proteins of *M. tuberculosis.*

MATERIALS AND METHODS

Collection and Analysis of Sequences

The complete genome sequence of pathogenic *Mycobacterium tuberculosis* H37Rv strain (accession number: NC_000962) was retrieved from http://www.ncbi.nlm.nih.gov/. The open reading frame (ORF) of each gene was validated and changes to reverse complementary to ensure +1 frame (if required) using different softwares i.e., APEplasmid editor, Generunner and DNAstar (Lasergene7v). Homology of each gene sequence was evaluated by BLAST (Altschul, et al., 2007).

Gene Designing and Codon Optimization

The nucleotides of DNA sequence were separated into triplets (codon) and then replaced codon with a new one (degenerate codon), generated with a given frequency distribution. After this amino acid will be same, but codon of low frequency of an amino acid will be replaced with a codon of high frequency, according to the desired species frequency distribution.

Gene designer (https://www.dna20.com/index.php?pageID=220) was used for designing of genes in a given expression hosts. Optimizer software was used to optimization and calculation of CAI, G+C and A+T (Puigbo, et al., 2007). CAIcal and MrGene were used to optimization of DNA sequences at maximum suitable threshold level. Codon optimization and simulation of genes were performed on 10-15 % threshold level of host cellular codons. Codon adaptation index (CAI) was calculated for each gene and it is widely acceptable as an effective measure of potential level of gene expression (Sharp and Li, 1987). Codon optimization is a technique used to exploit the protein expression in living organism by increasing the translational efficiency of particular gene of interest by transforming nucleotides of DNA sequence of one species into nucleotides of DNA sequence of another species like plant sequences to human sequences, human sequences to bacteria or yeast sequences.

Statistical Analysis

The CAI, GC and AT of potential genes of *M. tuberculosis* encoding antigenic proteins or enzymes were compared using the Wilcoxon matched pairs test. A two-tailed (α=2) probability $p<0.05$ were considered to be statistically significant. STATISTICA (version 7.0) was used for the analysis of native and optimized DNA sequences.

RESULTS AND DISCUSSION

In this study, the recent approach to codon optimize for production of adequate quantity of desire protein in specific host was used. The global perspective, is an urgent need to diagnose and control the infection and transmission of *mycobacterium tuberculosis* using the require dose of vaccines. It is not easily available to require people in adequate amount so we have tried to investigate the new type of synthetic genes for enhanced the expression level. Codon optimized genes have used complete physiological machinery of host for expression of desired genes. Degenerate codon frequencies in target and host organisms ensures proper folding and bias GC content to increase the mRNA stability or reduce secondary structures. Modify the ribosome binding sites, mRNA degradation sites and also adjust translational rates to allow the various domains of the protein to fold properly.

In the present study, we selected the genes based on previously identified the antigenic epitopes in the several proteins of *M. tuberculosis*. These antigenic epitopes for MHC alleles have been identified in heat shock protein, lipoprotein, secreted antigen 85-a mycolyl transferase (fbpa), mycolic acid synthase (pcaA) and outer membrane protein A (ompA) of *M. tuberculosis*. It indicates that these proteins encoding genes could be used as a good candidate for diagnostic antigen and vaccines development (Somvanshi, et al., 2008). There is limited information available on the codon optimization of *M. tuberculosis* in other host.

Codon optimization of the mycobacterial antigen Ag85B gene investigated for enhancement of the expression and antigenicity of the Ag85B DNA vaccine. The synthetic humanized Ag85B (hAg85B) gene in which codon usage was optimized for expression in human cells. DNA plasmids with codon-optimized hAg85B increased the level of protein expression *in vitro* and *in vivo* (Ko, et al., 2005).

Aeromonas hydrophila is pathogenic bacteria affecting the aquatic animals because of highly virulence proteins (hemolysin, aerolysin and lipase), that has been codon optimized for over expression in *E. coli* without codon bias (Singh, et al., 2010). The level of gene expression of eukaryotic genes introduced into mammalian cells depends on many factors i.e., gene copy number, transcriptional control elements, site of chromosomal integration, mRNA stability and translational efficiency (Gross and Hauser, 1995).

In this study, 19 potential protein encoding antigenic genes from *M. tuberculosis* H37Rv were used for codon optimization and simulation in *E. coli*. Heat shock protein, lipoprotein, secreted antigen 85-a mycolyl transferase (fbpa), mycolic acid synthase (pcaA), outer membrane protein A (ompA), transmembrane protein, serine protease PepD, lipoprotein LpqT, 3 hypothetical protein, proline-rich antigen, secreted protein, 9.5 kDa culture filtrate antigen CFP10A, PE-PGRS family protein, immunogenic protein MPT64, arsenic-transport integral membrane protein ArsA and secreted L-alanine dehydrogenase ALD encoding genes present in genome of *M. tuberculosis*. The characteristics of these genes i.e., G+C, A+T content and CAI of *M. tuberculosis* in references to *E. coli* and statistical analysis were given (Table 1).

Table 1: *Mycobacterium tuberculosis* antigens encoding genes used in this study for codon optimization with reference to *E. coli* host.

Sl No	Antigen Coding Genes	Location within genome (bp)	Native type DNA			Optimized type DNA		
			CAI	GC %	AT %	CAI	GC %	AT %
1	Heat shock protein	302173-302652	0.475	66.7	33.3	0.775	56.7	43.3
2	Lipoprotein	2562599-2563114	0.395	63.6	36.4	0.776	58.9	41.1
3	Secreted antigen 85-a mycolyl transferase (fbpa)	4264563-4265462	0.389	67.1	32.9	0.802	62.9	37.1
4	Mycolic acid synthase (pcaA)	559888-560748	0.445	52.8	47.2	0.817	49.9	50.1
5	Outer membrane protein A (ompA)	1002812-1003792	0.412	60.8	39.2	0.826	57.8	42.2
6	Transmembrane protein	1065127..1066038	0.419	66.9	33.1	0.788	60.1	39.9
7	Serine protease PepD	1099066..1100460	0.449	66.7	33.3	0.797	60.3	39.7
8	Lipoprotein LpqT	1134785..1135465	0.315	61.8	38.2	0.816	60.4	39.6
9	Hypothetical protein	1179396..1180577	0.498	66.9	33.1	0.837	55.6	44.4
10	Proline-rich antigen	1203313..1204035	0.439	66.3	33.7	0.801	60.3	39.7
11	Secreted protein	1305669..1306001	0.307	64.0	36.0	0.823	60.4	39.6
12	Hypothetical protein	1441348..1442718	0.429	65.0	35.0	0.833	58.0	42.0
13	9.5 kDa culture filtrate antigen CFP10A	1503103..1503384	0.492	63.8	36.2	0.786	54.6	45.4
14	Hypothetical protein	1901748..1902296	0.498	63.6	36.4	0.784	60.3	39.7
15	PE-PGRS family protein	1989833..1992577	0.396	77.3	22.7	0.788	68.5	31.5
16	Hypothetical protein	2107736..2108713	0.404	69.3	30.7	0.770	64.1	35.9
17	Immunogenic protein MPT64	2223343..2224029	0.314	64.2	35.8	0.811	64.2	35.8
18	Arsenic-transport integral membrane protein ArsA	3000614..3001903	0.270	59.4	40.6	0.437	62.9	37.1
19	Secreted L-alanine dehydrogenase ALD	3086820..3087935	0.437	64.2	35.8	0.831	57.9	42.1
		N	19	19	19	19	19	190
		Min	0.27	52.8	22.7	0.437	49.9	31.5
		Max	0.498	77.3	47.2	0.837	68.5	50.1
		Mean ± SD	0.409±0.06	64.757±4.73	35.242±4.73	0.784±0.08**	59.673±4.06**	40.326±4.06**

** $p<0.01$: in comparison with native type of gene sequences.

The improvement of CAI can predict better protein expression than the CAI from native codon used by host cells (Xia, 2007). PE-PGRS family protein encoding gene showed the 77.3% of G+C content and 0.0396 CAI but after the codon optimization we minimized the 68.5% G+C content and increased the 0.788 CAI value. So it indicate this gene can express better than native gene.

The location of each gene was localized in the genome of *M. tuberculosis*. The CAI of any gene is an important aspect to predict the expression of genes in host. The native type of genes was directly validated in the *E. coli* showed the less CAI value than the codon optimized genes. In the present study, we have evaluated the CAI, GC and AT frequencies in all nineteen genes of wild type strain ranges from 0.27 to 0.498, 52.8 to 77.3 and 22.7 to 47.2 with an average (±Standard Deviation: SD) of 0.409±0.06, 64.757±4.73 and 35.242±4.73 respectively. The specific frequencies of CAI, GC and AT in the optimized DNA ranges from 0.437 to 0.837, 49.9 to 68.5 and 31.5 to 50.1 with an average (±SD) of 0.784±0.08, 59.673±4.06 and 40.326±4.06 respectively. CAI and AT of optimized genes of *M. tuberculosis* were enhanced by 1.9 (47.8%) and 1.1 (12.5%) fold more with respect to the native type.

The number of studied genes is shown on X-axis while the CAI value at Y-axis (Figure 1). The minimized G+C content and increase the A+T content in the genes. The minimum 0.27 CAI value was observed in the arsenic-transport integral membrane protein later on doing the codon optimization we found that 0.437 CAI about 2 times more than native type of gene. In the present study, the maximum CAI was (0.498) while the codon optimization 0.837 was observed and established the compatible gene relationship of *M. tuberculosis* in *E. coli*.

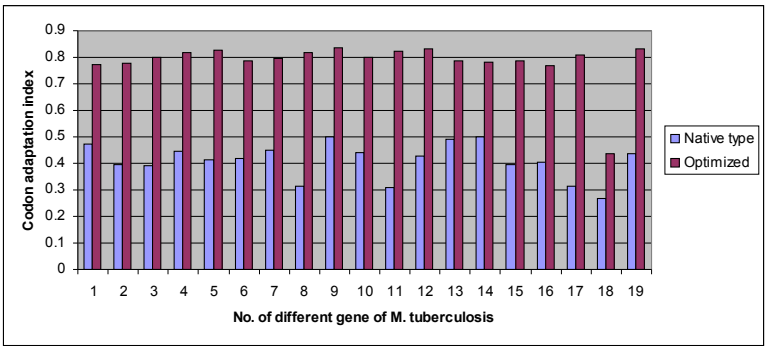

Figure 1. Comparison of optimized and native type of genes of *M. tuberculosis*.

The effects of codon optimization and chaperone coexpression in the heterologous expression of mammalian cytochrome P450s (P450) in *E. coli*, the expression of P450s 2B1, 2S1, 2U1, 2W1, and 27C1 have been reported. The codon optimization for N-terminus or the entire gene, the expression levels of P450 27C1, 2U1 and 2W1 increased 22-fold, 3.6-fold and 2.1-fold, respectively whereas P450s 2B1 and 2S1 remained unchanged (Wu, et al., 2009).

The effect of codon optimization on the expression levels of heterologous proteins in *Aspergillus oryzae,* using the mite allergen Derf7 as a model protein has been reported. A codon optimized Derf7 gene synthesized as per frequency of codon usage in *A. oryzae* by recursive PCR. Both native and optimized Derf7 genes have been expressed with their own signal peptides or in a fusion construct with *A. oryzae* glucoamylase (GlaA). Codon optimization markedly increased protein and mRNA production levels in both nonfused and GlaA-fused Derf7 constructs (Tokuoka, et al., 2008). It is thought that optimal codons help to achieve faster translation rates and high accuracy. As a result of these factors, translational selection is expected to be stronger in highly expressed genes, as is indeed the case for the above-mentioned organisms. In other organisms that do not show high growing rates or that present small genomes, codon usage optimization is normally absent, and codon preferences are determined by the characteristic mutational biases seen in that particular genome.

Codon optimization of the hPDF gene has been performed to reduce the GC content and to eliminate rare codons. Codon optimization increased the expression of hPDF protein by reducing the GC content. A large amount of soluble hPDF was obtained from its fusion with thioredoxin and also in the insoluble fraction (Han, et al., 2010). The usage of alternative synonymous codons in *M. tuberculosis* genes has been investigated. This species is a member of the high G + C Gram-positive bacteria with a genomic G+C content around 65 %. This G+C-richness is reflected in a strong bias towards C and G ending codons for every amino acid: overall, the G+C content at the third positions of codons is 83%. However, there is significant variation in codon usage patterns among genes, which appears to be associated with gene expression level. The set of selectively favoured codons seems to be highly conserved between *M. tuberculosis* and another high G+C Gram-positive bacterium, *Corynebacterium glutamicum,* even though the genome and overall codon usage of the later are much less G+C-rich (Andersson and Sharp 1996). Gene encoding the Sm14 antigen of *Schistosoma mansoni* was generated and codon-optimized in mycobacterial. This synthetic gene enhanced approximately 4 fold the protein expression level in recombinant *M. bovis Bacille* Calmette-Guérin (rBCG) when compared to the native gene (Varaldo, et al., 2006).

In this study, the high frequency mean 64.75% of G+C content in the genes of *M. tuberculosis* indicates the codon bias during the expression in *E. coli* was observed. However, the optimized genes of *M. tuberculosis* was reduced the 59.67% of G+C content that helps to prevent the codon bias.

For the enhancement of the expression levels the optimized the codon usage of HIV type 1 (HIV-1) p24 antigen gene of gag (p24 gag) and established a codon optimized recombinant BCG (rBCG)-p24 Gag which expressed a 40-fold higher level of p24 Gag than the non-optimized rBCG-p24 Gag in *M. bovis* bacille Calmette-Guérin (BCG). Inoculation of mice with the codon-optimized rBCG-p24 Gag elicited effective immunity, as evidenced by virus-specific lymphocyte proliferation, gamma interferon ELISPOT cell induction, and antibody production. In comparison to inoculation of animals with the non-optimized rBCG-p24 Gag induced only low levels of immune responses (Kanekiyo, et al., 2005).

The efficiency has been increased with the use of production of synthetic genes in *E. coli* the heterologous protein production. 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) of each gene had been reported (Burgess-Brown, et al., 2008).

As the interspecific difference of codon usage for effective induction of specific immune responses against several bacteria and protozoa is a main problem. 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). The expression of antimicrobial peptides human β-defensin-26 and human β-defensin-27 in *E. coli* investigated. HBD26 and HBD27 genes have been synthesized through codon optimization and each gene was then cloned into the expression vector pET32 and then transformed into *E. coli* BL21 (DE3), which gave yields of HBD26 and HBD27 fusion proteins up to 1.38 and 1.29 g/litre. Both variants showed salt-sensitive antimicrobial activity against gram-negative *E. coli* but not against gram-positive *Staphylococcus aureus* and *S. cerevisiae* (Huang, et al., 2009).

Conclusion

In conclusion, the recent approach to the antigen coding genes that produce antigenic proteins was investigated. These antigenic proteins are required for immunoassay based diagnosis of *Mycobacterium tuberculosis*. Furthermore, it is also needed for tuberculosis vaccination. Moreover it has 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. In this study we found the good correlation among the genes based on the statistical analysis, the native type frequencies of codon optimized DNA of *M. tuberculosis* showed significantly (p<0.01) different codons. These optimized DNA will be synthesized and over expressed in *E. coli* as compared to its native type counterparts. The expression of optimized gene may be more due to its high CAI value and available codon present in as compare to native type DNA in *E. coli*. However these finding also provides a new insight for other pathogens which is difficult to grow in laboratory condition. Subsequently, it can be used as a diagnostic antigen for immunoassays and also helpful to development of vaccines. Moreover, the *in vitro* studies will be needed for validation of codon optimized DNA for further high level of expression and their adequate amount for vaccine production.

ACKNOWLEDGMENTS

The authors are thankful to AK Singh, JK Singh and J Yadav (BHU) for providing the laboratory facilities and valuable suggestion in the manuscript.

REFERENCES

Altschul SF, Thomas LM, Alejandro AS, Jinghui Z, Zheng Z, Miller W and Lipman DJ. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389-3402.

Andersson SGE and Sharp PM. (1996). Codon usage in the *Mycobacterium tuberculosis* complex. Microbiol. **142:**915-925.

Burgess-Brown NA, Sharma S, Sobott F, Oppermann CLU and Gileadi O. (2008). Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. Protein Expr. Purif. **59:**94–102.

Chopra P, Meena LS and Singh Y. (2003). New drug targets for *Mycobacterium tuberculosis.* Indian J. Med. Res. **117**:1-9.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S_and Barrell BG. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. **393:**537–544.

Comeron JM and Aguade M. (1998). An evaluation of measures of synonymous codon usage bias. J. Mol. Evol. **47(3):**268-274.

Fine PEM, Carneiro IAM, Milstien JB and Clements CJ. **(**1999). Issues relating to the use of BCG in immunization programmes: a discussion document. WHO Department of Vaccines and Biologicals, document 99.23. WHO, Geneva, Switzerland.

Futcher B, Latter G I, Monardo P, McLaughlin C S and Garrels J I. (1999). A sampling of the yeast proteome. Mol. Cell Biol. **19:**7357–7368.

Gross G and Hauser H. (1995). Heterologous expression as a tool for gene identification and analysis. J. Biotechnol. **41(2):**91-110.

Gustafsson C, Govindarajan S and Minshull J. (2004). Codon bias and heterologous protein expression. Trends Biotechnol. **22**:346–353.

Gygi SP, Rochon Y, Franza BR and Aebersold R. (1999). Correlation between protein and mRNA abundance in yeast. Mol. Cell Biol. **19:**1720–1730.

Han JH, Choi YS, Kim WJ, Jeon YH, Lee SK, Lee BJ and Ryu KS. (2010). Codon optimization enhances protein expression of human peptide deformylase in *E. coli.* Protein Expr. Purif. **70(2):**224-230.

Huang L, Leong SS and Jiang R. (2009). Soluble fusion expression and characterization of bioactive human beta-defensin 26 and 27. Appl. Microbiol. Biotechnol. **84(2):**301-308.

Ikemura T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151:**389–409.

Ivory C and Chadee K. (2004). DNA vaccines: designing strategies against parasitic infections. Genet. Vaccines Ther. **2**:17.

Kanekiyo M, Matsuo K, Hamatake M, Hamano T, Ohsu T, Matsumoto S, Yamada T, Yamazaki S, Hasegawa A, Yamamoto N, Honda M. (2005). Mycobacterial codon optimization enhances antigen expression and virus-specific immune responses in recombinant *Mycobacterium bovis* bacille Calmette-Guérin expressing human immunodeficiency virus type 1 Gag. J. Virol. **79(14):**8716-8723.

Kaufmann S H. (2001). How can immunology contribute to the control of tuberculosis?. Nat. Rev. Immunol. **1:**20–30.

Khedkar SA, Malde AK and Coutinho EC. (2005). Comparative protein modeling of methionine S-adenosyltransferase (MAT) enzyme from *Mycobacterium tuberculosis*: a potential target for antituberculosis drug discovery. J. Mol. Graph. Model. **23:**355–366.

Ko HJ, Ko SY, Kim YJ, Lee EG, Cho SN and Kang CY. (2005). Optimization of codon usage enhances the immunogenicity of a DNA vaccine encoding mycobacterial antigen Ag85B. Infect. Immun. **73(9):**5666-5674.

Lithwick G and Margalit H. (2003). Hierarchy of sequencedependent features associated with prokaryotic translation. Genome Res. **132:**665–2673.

Lithwick G and Margalit H. (2005). Relative predicted protein levels of functionally associated proteins are conserved across organisms. Nucleic Acids Res. **33**:1051–1057.

Nagata T, Uchijima M, Yoshida A, Kawashima M and 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.

Puigbo P, Guzman E, Romeu A and Garcia-Vallve S. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. 1–6.

Rocha E P. (2004). Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. **14:** 2279–2286.

Ruiz LM, Armengol G, Habeych E and Orduz S. (2006). A theoretical analysis of codon adaptation index of the Boophilus microplus bm86 gene directed to the optimization of a DNA vaccine. J. Theor. Biol. **239:**445–449.

Sharp PM and Li WH. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15(3):**1281–1295.

Sharp PM, Bailes E, Grocock RJ, Peden JF and Sockett RE. (2005).Variation in the strength of selected codon usage bias among bacteria, Nucleic Acids Res. **33:**1141–1153.

Singh V and Somvanshi P. (2009a). Homology modeling of 3-oxoacyl-acyl carrier protein synthase II from *Mycobacterium tuberculosis* H37Rv and molecular docking for exploration of drugs. J. Mol. Model. **15(5):**453-460.

Singh V and Somvanshi P. (2009b). Targeting the peptide deformylase of *Mycobacterium tuberculosis* leads to drug discovery. Letters drug design discov. **6(7)** 487-493.

Singh V and Somvanshi P. (2010). Toward the virtual screening of potential drugs in the homology modeled NAD^+ dependent DNA ligase from *Mycobacterium tuberculosis*. Protein Pept. Lett. **17(2):**269-276.

Singh V, Chaudhary DK, Mani I, Somvanshi P, Rathore G and Sood N. (2010). Molecular identification and codon optimization analysis of major virulence encoding genes of *Aeromonas hydrophila.* African J. Microbiol. Res. **4(10)**: 952-957.

Smith I. (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. Clin. Microbiol. Rev. **16(3)**:463–496.

Somvanshi P, Singh V and Seth PK. (2008). *In silico* prediction of epitopes in virulence proteins of *Mycobacterium tuberculosis* H37Rv for diagnostic and subunit vaccine design. J. Proteomics Bioinform. **1:**143–153.

Tokuoka M, Tanaka M, Ono K, Takagi S, Shintani T and Gomi K. (2008). Codon optimization increases steady-state mRNA levels in *Aspergillus oryzae* heterologous gene expression. Appl. Environ. Microbiol. **74(21):**6538-6546.

Varaldo PB, Miyaji EN, Vilar MM, Campos AS, Dias WO, Armôa GR, Tendler M, Leite LC and McIntosh D. (2006). Mycobacterial codon optimization of the gene encoding the Sm14 antigen of *Schistosoma mansoni* in recombinant Mycobacterium bovis Bacille Calmette-Guérin enhances protein expression but not protection against cercarial challenge in mice. FEMS Immunol. Med. Microbiol. **48(1):**132-139.

Wu ZL, Qiao J, Zhang ZG, Guengerich FP, Liu Y and Pei XQ. (2009). Enhanced bacterial expression of several mammalian cytochrome P450s by codon optimization and chaperone coexpression. Biotechnol. Lett. **31(10):**1589-1593.

Xia X. (2007). An Improved Implementation of Codon Adaptation Index, Evol. Bioinform. **3:**53–58.

Xu L, Jiang X, Yang J, Liu Y and Yan Y. (2010). Cloning of a novel lipase gene, lipJ08, from *Candida rugosa* and expression in *Pichia pastoris* by codon optimization. Biotechnol. Lett. **32(2):**269-276.

****