SHORT COMMUNICATION

Efficient soluble expression and purification of influenza A and B nucleoproteins in E. coli

Nikita D. Yolshin, Aram A. Shaldzhyan, Sergey A. Klotchenko

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russian Federation
#*Corresponding author*: Nikita Yolshin, e-mail: nikita.yolshin@gmail.com

ABSTRACT

Viral nucleoprotein (NP) is an abundant essential protein of an influenza virus that has important functional and structural roles. It participates in genomic organization, nuclear trafficking, RNA transcription, and genome replication. From the research point of view, NP is an important protein that is used in the development of new diagnostic methods and vaccination protocols. NP is a promising target for antiviral chemotherapeutic drugs as well. Successful expression of codon-optimized *NP* genes in *E. coli* has been reported. In this study, we demonstrated the efficient expression and purification of soluble NPs of influenza A and B viruses in *E. coli* without the codon-optimization of DNA sequences. This procedure preserves the co-translational protein folding, protein configuration and function. Obtained NPs of influenza A and B viruses were monomers and reacted well with mouse specific antibodies according to Western blot analysis. Our results show that both influenza A and influenza B virus NPs can be efficiently expressed in *E. coli* without codon-optimization.

INTRODUCTION

Influenza viruses cause respiratory infections in humans that range from asymptomatic to deadly disease. Regular epidemics and pandemics make the influenza virus a global health threat. The nucleoprotein (NP) is a multifunctional viral protein with important structural and functional roles [1, 2]. It plays a critical role in viral replication. NP is involved in important functions such as RNA packaging [3], nuclear trafficking, transcription, and replication of vRNA [4, 5]. Each genomic segment of the influenza virus vRNA is associated with multiple NP molecules and a single polymerase complex (with PB2, PB1, and PA proteins) forming a viral ribonucleoprotein (vRNP) complex. The NP binds to single-stranded vRNA in a non-specific manner [6] with a periodicity of one NP molecule per every 24 bases [7, 8]. The NP contains the nuclear localization signal (NLS), which is essential for the import of vRNP complex to the nucleus where the mRNAs are transcribed. The newly translated NPs participate in the protection of genomic RNA from degradation. The structure of NP is relatively conserved (less than 11% variance) among the different influenza A strains.

Therefore, NP is an appealing target for the development of antiviral drugs [9] and an attractive candidate as a component for broad-spectrum influenza vaccines [10, 11]. For example, it was shown that immunization with a plasmid DNA expressing influenza NP followed by boosting with adenovirus expressing NP leads to the induction of the protective immunity against challenge with highly pathogenic H5N1 influenza virus in animal models [12].

In addition, the NP serves as a target antigen for the diagnosis of influenza virus infections [13, 14].

Therefore, the development of efficient methods for recombinant NP production is an important goal. The expression of recombinant proteins in bacteria and/or eukaryotic cells is the subject of intensive studies [15, 16]. The most prevalent approach for the acceleration of the recombinant protein expression is codon optimization [17]. The successful codon-optimized expression of NP in *E. coli* and in Sf9 insect cells was shown by Huang *et al.* [18] and Yoon *et al.* [19], respectively. At the same time, as it was noted in a number of publications [16, 20, 21], codon optimization can affect the co-translational protein folding, protein configuration and function. Several recent studies proved that the use of the messenger RNA containing rear codons decelerates translation and at the same time ensures the proper co-translational protein folding and consequently post-translational modification and function of expressed proteins [15, 21-23]. Codon optimization can lead to disruption of the co-translational protein folding and consequently to the formation of proteins with impaired functions or

non-functional proteins. Konczal *et al.* [24] showed that introduction of rare codons back to the codon-optimized DNA sequence leads to overall slower translation, but at the same time to enhanced recovery of the soluble protein since it provides the proper co-translational folding of the recombinant protein. Codon optimization in this case led to enhancement of the total protein yield but to lower yield of the desired soluble protein. Therefore, the efficient expression of recombinant proteins using native DNA remains an important goal.

In this study, we showed that influenza A and B virus *NP* genes could be efficiently expressed in *E. coli* without resorting to codon-optimization. Both NPs produced by this method were monomers, were fairly well soluble, and possessed specific immune reactivity.

MATERIALS AND METHODS

Selection of influenza strains and the production of purified viral concentrates

The influenza A/Brisbane/10/2007 (H3N2) and B/Brisbane/46/2015 viruses were used for the production of recombinant NPs. These strains were obtained from the Smorodintsev Research Institute of Influenza (St. Petersburg, Russia) collection. The influenza A and B viruses were grown in the allantoic cavity of 10-day old embryonated chicken eggs at 37°C and 34°C, respectively. Allantoic fluid was harvested in 48 h for influenza A and in 72 h for influenza B strain post inoculation. Virus-containing allantoic fluids were concentrated and the viruses were purified by ultracentrifugation (Beckman, Type 19 rotor) using sucrose density gradients [25] and stored at -80°C.

Construction of pET22b+NP A and pET22b+NP B expression plasmids

Total RNA was isolated from the corresponding purified viral concentrates using the Rneasy Plus Mini Kit (Qiagen, UK). The first strand of cDNA was synthesized using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, USA) with random primers. The cDNA corresponding to both *NP* genes was amplified using forward primers with a NdeI restriction site and the reverse primers containing the XhoI restriction site (Table 1).

Table 1. Primers used for the cDNA amplification of influenza A and B *NP* genes

Primer name	Sequence (5' → 3')
NPA F NdeI	CATATGGCGTCCCAAGGCACCAAACG
NPA R XhoI	CTCGAGATTGTCGTACTCTTCTGCATTGTCTC
NPB F NdeI	CATATGTCCAACATGGATATCGA
NPB R XhoI	CTCGAGATAATCGAGGTCATCAT

Polymerase chain reaction (PCR) was conducted with Encyclo polymerase (Evrogen, Russia) in Bio-Rad C1000 machine (USA). PCR for the influenza B *NP* gene was challenging due to the fact that the reverse primer had homology with a short section in the middle of the nucleotide sequence of this gene (from 299 to 306 bp). That

led to the generation of a predominantly shorter product of 306 bp instead of the desired product of 1682 bp under a variety of conditions. This trend persisted despite the use of different quantities of template, extension temperature gradients, and usage of different polymerases. Therefore, the reamplification of cDNA was not used. Cloning of the influenza A *NP* gene was straightforward.

The *NP* A and *NP* B PCR products were inserted into the pET22b+ commercial vector (Novagen) (Fig. 1). PCR products were gel-purified using the Cleanup Standard Kit (Evrogen, Russia) and digested by two restriction enzymes NdeI and XhoI (Thermo Fisher Scientific, USA). The pET22b+ vector was also digested by the same enzymes, followed by the ligation of the vector/insert mixture.
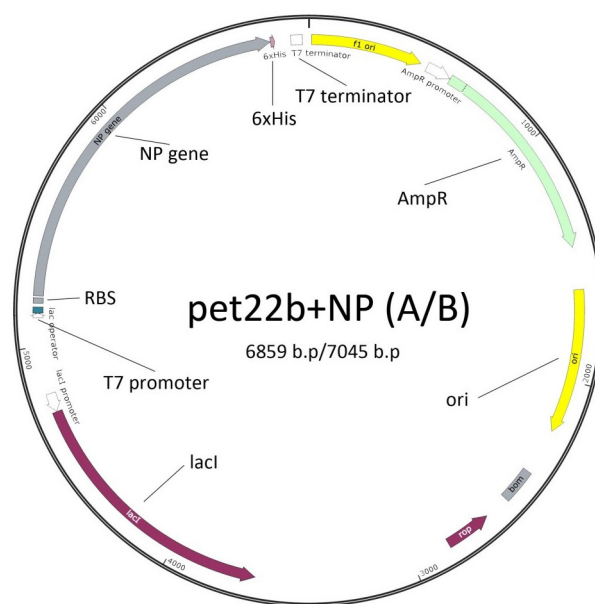
Fig. 1. Map of the pET22b+NP expression plasmid(s).

Electrocompetent *E. coli* BL21(DE3) cells were transformed by the ligation mix, and colonies were selected on LB/ampicillin (100 µg/ml) plates. After the confirmation that bacteria carry the expression plasmid by PCR, the corresponding colonies were selected. The sequencing of the pET22b+ NP A and pET22b+ NP B constructs was performed using ABI PRISM 3100 Genetic Analyzer and their identity was confirmed by comparison to the original sequences of influenza A/Brisbane/10/2007 virus *NP* (PubMed accession number KJ609215.1) and influenza B/Brisbane/46/2015 *NP* (GISAID accession number EPI772456) genes.

Expression and purification of the fusion proteins

Selected clones were cultured overnight on LB/ampicillin plates at 37°C. The obtained cultures were dissolved in the pre-warmed fresh LB with ampicillin (100 µg/ml) to form a solution with OD_{600} of 0.1 and then grown in flasks at 37°C with shaking until the OD_{600} of the solution reached 0.9.

Expression of the NP A and the NP B was initiated by the addition of IPTG to a final concentration of 0.1 mM. The cultures were grown for 3 h at 37°C with 250 rpm shaking. Bacterial pellets were obtained from cultures by

centrifugation at 5000 g (Eppendorf, 5804R) for 15 min. Cell pellets were resuspended in 40 ml of lysis buffer (50 mM sodium phosphate, 500 mM sodium chloride, 20 mM imidazole) and 1 mM of phenylmethylsulfonyl fluoride (PMSF, pH 8.0) protease inhibitor. The cells were subsequently lysed by sonication on ice and centrifuged at 13,000 g for 30 min to remove cell debris and insoluble proteins.

Supernatants containing soluble fractions of recombinant NPs (recNPs) were filtered through 0.45 µm PES syringe filters. A HisTrap HP one milliliter columns (GE Healthcare, USA) were equilibrated with 5 ml of binding buffer (50 mM sodium phosphate, 500 mM sodium chloride, and 20 mM imidazole, pH 8.0) at a flow rate of 1 ml/min. Filtered supernatants were loaded on the columns at the same flow rate. Then columns were washed with 20 ml of binding buffer in order to remove unbound material followed by elution with 10 ml of elution buffer (50 mM sodium phosphate, 500 mM sodium chloride, and 500 mM imidazole, pH 7.8) at the same flow rate. Protein-containing fractions were collected, combined, and passed through a ten milliliter BioScale Mini P-6 desalting cartridge (Bio-Rad, USA), preliminary equilibrated with PBS buffer, at a flow rate of 5 ml/min. Protein concentrations were determined by Lowry assay, and the purity of the recNPs were determined by SDS-PAGE and densitometry.

HPLC purification of recNP was performed using Akta Pure system with the Superdex 200 Increase 10/300 GL column (GE Healthcare) for size-exclusion chromatography. Column was equilibrated with 75 ml of HPLC buffer (150 mM NaCl, 30 mM Na_3PO_4, pH 7.8) at 0.5 ml/min flow rate. The injection volume of a sample was 100 µl, elution volume was 30 ml.

Calibration was done using a series of proteins with known molecular weight (MW). Using the obtained chromatographic data MW for purified recNP A was determined as 57 kDa, that matches the predicted MW of recNP A monomeric form. The chromatogram is shown in Fig. 2.

SDS-PAGE assay

Samples (cell lysates, elution fractions with purified protein, and negative control cell lysates) were mixed with 4x loading buffer containing reducing agent β-mercaptoethanol and incubated at 95℃ for 10 min. Then, the mixtures were separated by SDS-PAGE on the 12% separating gel according to the standard procedure and subsequently stained with Coomassie blue R-250.

Western blot analysis

Purified proteins were analyzed by Western blot using antibodies to His tag (Qiagen AntiHis5, #34660) and monoclonal antibodies 4H1 generated to the NP of influenza virus A/Brisbane/10/2007 as well as antibodies 1B12 generated to the NP of B/Brisbane/46/2015 virus obtained in the Smorodintsev Research Institute of Influenza (St. Petersburg, Russia). Proteins were transferred to nitrocellulose membranes using the Bio-Rad Trans Blot Turbo system. Then, the membrane was blocked with non-fat dried milk (NFDM) in PBS buffer containing 0.05% Tween 20
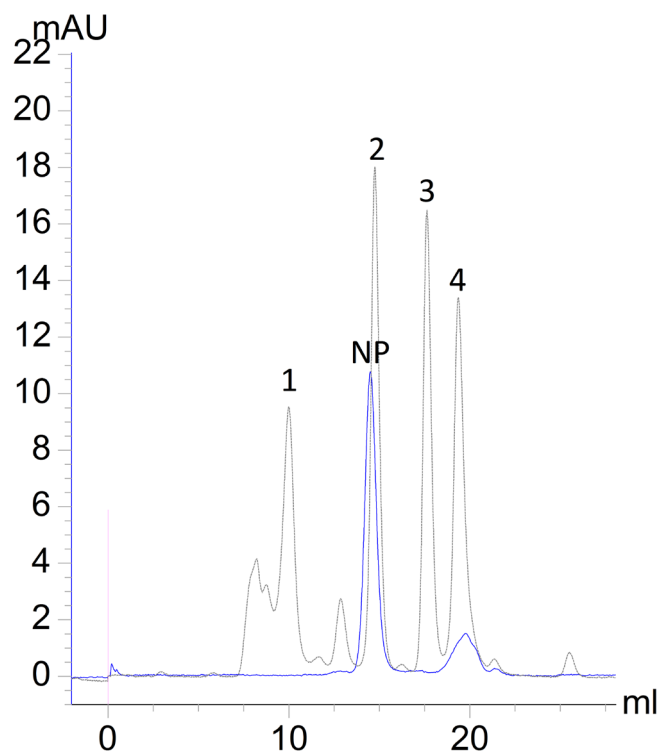

Fig. 2. The HPLC chromatogram of purified recNP A verses a chromatogram of the mixture of molecular mass standards. 1 – ferritin (440 kDa), 2 – ovalbumin (44 kDa), 3 – RNase A (13.5 kDa), 4 – aprotinin (6.5 kDa), NP – purified recNP A.

(PBST) for 1 h and incubated with primary antibody solutions (antibodies to His tag or monoclonal antibodies 4H1 and 1B12) in 4% NFDM (Sigma Aldrich) for 2 h following by incubation with secondary antibodies (GAM, Sigma Aldrich, #A3682) for 1 h in PBST buffer.

RESULTS AND DISCUSSION

The *NP* A and *NP* B PCR products of influenza A/Brisbane/10/2007 (H3N2) and B/Brisbane/46/2015, respectively, were inserted into the pET22b+ commercial vector (Novagen). The expression vectors pET22b+NP A and pET22b+NP B were designed to produce recNP A and recNP B proteins with a c-terminal 6xHis tag as shown in Fig. 1. The *NP* A (1496 bp) and *NP* B (1682 bp) coding sequences were amplified from cDNAs transcribed from the corresponding viral RNAs using the primers shown in the Table 1. As was mentioned hereinabove, the amplification of influenza B *NP* was more problematic due to a partial homology of the reverse primer to an internal region of the gene (from 299 to 306 bp) leading to the amplification of a shorter product. Therefore, the cloning of the *NP* of influenza B virus was performed after one PCR round. The sequencing of the plasmids from the corresponding colonies showed the presence of full-length inserts identical to the original sequences.

The pET22b+NP A and pET22b+NP B plasmids were transformed into *E. coli* BL21(DE3). In the standard induction procedure, cultures were grown for 3 h at 37℃ with shaking (250 rpm) in the presence of IPTG (inductor) at a final concentration of 0.1 mM. In order to maximize

the yield of soluble NPs, different experimental conditions were screened: various IPTG concentrations (0.05, 0.1, and 0.5 mM) and incubation temperatures (23℃, 32℃, and 37℃). No difference in the protein expression level at different experimental conditions was observed. Therefore, the transformed *E. coli* were grown at 37℃ with the addition of IPTG up to 0.1 mM until the OD_{600} of culture reached the value of 0.9. Then, the cells were incubated for an additional 3 h. Bacterial cell pellets were obtained by centrifugation at 5,000 g for 15 min (Eppendorf, 5804R), resuspended in lysis buffer, and lysed by sonication using the homogenizer (MSE Soniprep 150, UK).

Supernatants containing soluble fractions of recNPs were purified using Ni-NTA columns and then the products were eluted with high imidazole buffer. Total lysates, fractions, and purified proteins were analyzed by SDS-PAGE. This analysis showed the presence of bands of approximately 57 kDa and 62 kDa (Fig. 3) that correlate well with the predicted molecular weight of the recNP fusion proteins of influenza A and B viruses, respectively. After Ni^{2+} affinity chromatography, the purity of both proteins was above 95%. From 1 liter of the corresponding cultures, 15 mg of NP A and 5 mg of NP B were isolated.

Western blot analysis indicated that purified recNP A and recNP B reacted well with the monoclonal antibodies obtained to influenza A/Brisbane/10/2007 and B/Brisbane/46/2015 viruses, respectively. Both proteins also reacted well with anti-His tag antibodies (Qiagen AntiHis5, #34660) (Fig. 4).

The codon optimization of DNA sequences is widely used in order to increase the expression level of target proteins. A synonymous codon substitutions – substitutions of codons that are translated slower by faster translated codons – can significantly increase the rate of translation of the desired protein [26]. However, it was shown that this procedure can influence the protein folding and stability and can interfere with post-translational protein modification [23, 27]. Some slower translated
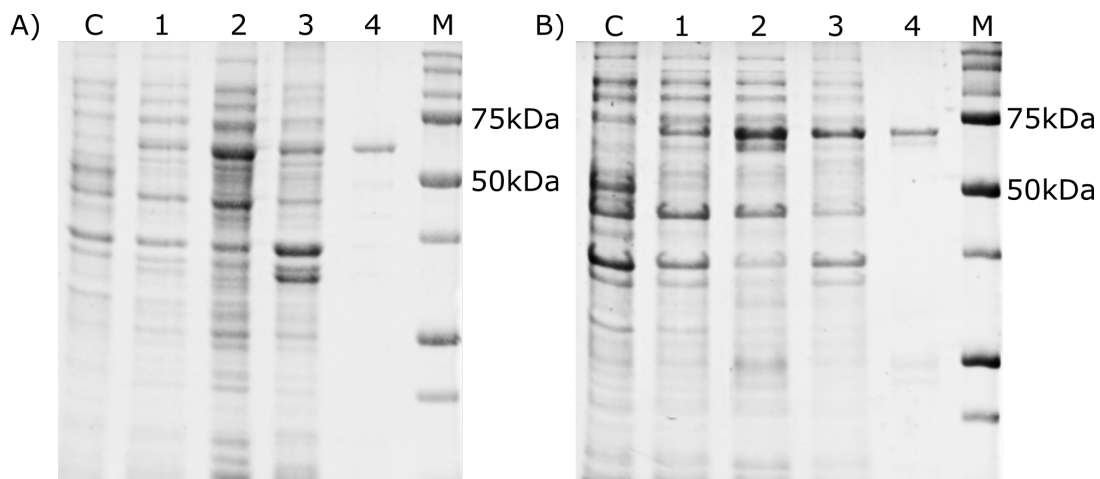
Fig. 3. SDS-PAGE of *E. coli* lysates and purified NPs.
A. NP of influenza A virus: C – negative control, containing *E. coli* BL21(DE3) with pET22b+, 1- total lysate, 2 – soluble fraction, 3 – pellet, 4 – purified recombinant NP, M – molecular marker; **B.** NP of influenza B virus: C – negative control, containing *E. coli* BL21(DE3) with pET22b+, 1 – total lysate, 2 – soluble fraction, 3 – pellet, and 4 – purified recombinant NP, M – molecular marker.
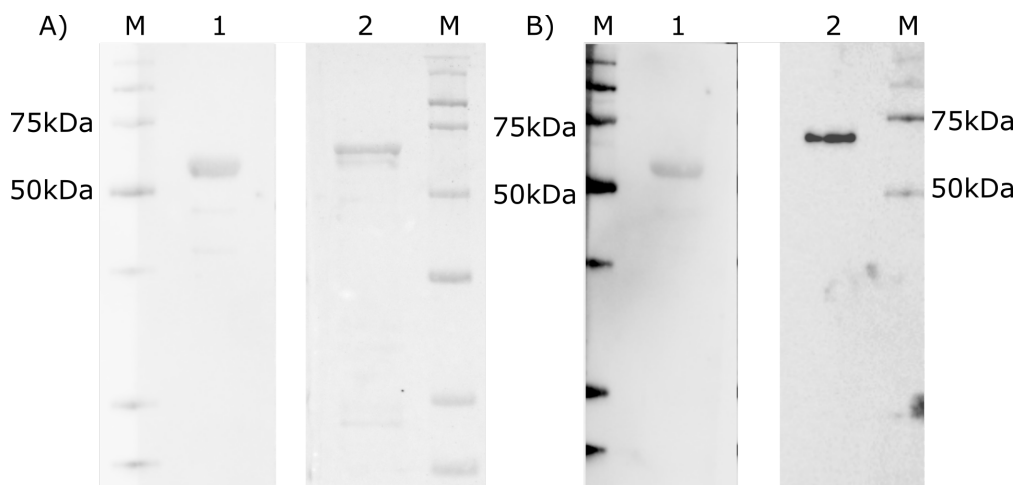
Fig. 4. Western blot analysis of purified recombinant NPs.
A. Staining with antibodies to 6xHis tag. **B.** Staining with mouse monoclonal antibodies 4H1 specific to NP of A/Brisbane/10/2007 and 1B12 specific to NP of influenza B/Brisbane/46/2015 virus. M – molecular marker, 1 – NP of A/Brisbane/10/2007, 2 – NP of B/Brisbane/46/2015 virus.

codons were shown to have advantages for conserving the exact protein structure especially if they encode the amino acids that are located at the key positions of the protein sequence because mutation at these positions would drastically influence the protein properties [28, 29]. Therefore, codon optimization leads not only to the regulation of the protein expression level, but also can influence the structure and function of the expressed proteins. In summary, we showed that influenza A and B heterologous NPs can be successfully expressed in *E. coli* using original sequences. Both expressed nucleoproteins demonstrated good immune reactivity with specific antibodies.

ACKNOWLEDGEMENTS

CONFLICT OF INTERESTS

The authors claim no commercial or financial interests.

CITATION

COPYRIGHT

REFERENCES

1. Gorman OT, Bean WJ, Kawaoka Y, Webster RG. Evolution of the nucleoprotein gene of influenza A virus. J Virol. 1990; 64(4), 1487-97. PubMed PMID: 2319644.

2. Shu LL, Bean WJ, Webster RG. Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990. J Virol. 1993; 67(5), 2723-9. PubMed PMID: 8474171.

3. O'Neill RE, Jaskunas R, Blobel G, Palese P, Moroianu J. Nuclear import of influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import. J Biol Chem. 1995; 270(39), 22701-4. doi: 10.1074/jbc.270.39.22701.

4. Turrell L, Lyall JW, Tiley LS, Fodor E, Vreede FT. The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. Nat Commun. 2013; 4, 1591. doi: 10.1038/ncomms2589.

5. Eisfeld AJ, Neumann G, Kawaoka Y. At the centre: influenza A virus ribonucleoproteins. Nat Rev Microbiol. 2015; 13(1):, 28-41. doi: 10.1038/nrmicro3367.

6. Baudin F, Bach C, Cusack S, Ruigrok RW. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. EMBO J. 1994; 13(13), 3158-65. PubMed PMID: 8039508.

7. Compans RW, Content J, Duesberg PH. Structure of the ribonucleoprotein of influenza virus. J Virol. 1972; 10(4), 795-800. PubMed PMID: 4117350.

8. Ortega J, Martin-Benito J, Zurcher T, Valpuesta JM, Carrascosa JL, Ortin J. Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification. J Virol. 2000; 74(1), 156-63. doi: 10.1128/jvi.74.1.156-163.2000.

9. Cianci C, Gerritz SW, Deminie C, Krystal M. Influenza nucleoprotein: promising target for antiviral chemotherapy. Antivir Chem Chemother. 2012; 23(3), 77-91. doi: 10.3851/IMP2235.

10. Heiny AT, Miotto O, Srinivasan KN, Khan AM, Zhang GL, Brusic V, et al. Evolutionarily conserved protein sequences of influenza a viruses, avian and human, as vaccine targets. PLoS One. 2007; 2(11), e1190. doi: 10.1371/journal.pone.0001190.

11. Huang B, Wang W, Li R, Wang X, Jiang T, Qi X, et al. Influenza A virus nucleoprotein derived from Escherichia coli or recombinant vaccinia (Tiantan) virus elicits robust cross-protection in mice. Virol J. 2012; 9, 322. doi: 10.1186/1743-422X-9-322.

12. Epstein SL, Kong WP, Misplon JA, Lo CY, Tumpey TM, Xu L, et al. Protection against multiple influenza A subtypes by vaccination with highly conserved nucleoprotein. Vaccine. 2005; 23(46-47), 5404-10. doi: 10.1016/j.vaccine.2005.04.047.

13. Vemula SV, Zhao J, Liu J, Wang X, Biswas S, Hewlett I. Current Approaches for Diagnosis of Influenza Virus Infections in Humans. Viruses. 2016; 8(4), 96. doi: 10.3390/v8040096.

14. Phuong NH, Kwak C, Heo CK, Cho EW, Yang J, Poo H. Development and Characterization of Monoclonal Antibodies against Nucleoprotein for Diagnosis of Influenza A Virus. J Microbiol Biotechnol. 2018; 28(5), 809-15. doi: 10.4014/jmb.1801.01002.

15. Brule CE, Grayhack EJ. Synonymous Codons: Choose Wisely for Expression. Trends Genet. 2017; 33(4), 283-97. doi: 10.1016/j.tig.2017.02.001.

16. Hanson G, Coller J. Codon optimality, bias and usage in translation and mRNA decay. Nat Rev Mol Cell Biol. 2018; 19(1), 20-30. doi: 10.1038/nrm.2017.91.

17. Mauro VP, Chappell SA. Considerations in the Use of Codon Optimization for Recombinant Protein Expression. Methods Mol Biol. 2018; 1850, 275-88. doi: 10.1007/978-1-4939-8730-6_18.

18. Huang BY, Wang WL, Wang XP, Jiang T, Tan WJ, Ruan L. [Efficient soluble expression and purification of influenza A nucleoprotein in Escherichia coli]. Bing Du Xue Bao. 2011; 27(1), 50-7. PubMed PMID: 21462506.

19. Yoon SJ, Park YJ, Kim HJ, Jang J, Lee SJ, Koo S, et al. Optimized Expression, Purification, and Rapid Detection of Recombinant Influenza Nucleoproteins Expressed in Sf9 Insect Cells. J Microbiol Biotechnol. 2018; 28(10), 1683-90. doi: 10.4014/jmb.1805.05053.

20. Mauro VP, Chappell SA. A critical analysis of codon optimization in human therapeutics. Trends Mol Med. 2014; 20(11), 604-13. doi: 10.1016/j.molmed.2014.09.003.

21. Zhao F, Yu CH, Liu Y. Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells. Nucleic Acids Res. 2017; 45(14), 8484-92. doi: 10.1093/nar/gkx501.

22. Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in Escherichia coli. Science. 2009; 324(5924), 255-8. doi: 10.1126/science.1170160.

23. Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, et al. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature. 2013; 495(7439), 111-5. doi: 10.1038/nature11833.

24. Konczal J, Bower J, Gray CH. Re-introducing non-optimal synonymous codons into codon-optimized constructs enhances soluble recovery of recombinant proteins from Escherichia coli. PLoS One. 2019; 14(4), e0215892. doi: 10.1371/journal.pone.0215892.

25. Arora DJ, Tremblay P, Bourgault R, Boileau S. Concentration and purification of influenza virus from allantoic fluid. Anal Biochem. 1985; 144(1), 189-92. doi: 10.1016/0003-2697(85)90103-4.

26. Sharp PM, Li WH. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 1987; 15(3), 1281-95. doi: 10.1093/nar/15.3.1281.

27. Komar AA, Lesnik T, Reiss C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. FEBS Lett. 1999; 462(3), 387-91. doi: 10.1016/s0014-5793(99)01566-5.

28. Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell. 2008; 134(2), 341-52. doi: 10.1016/j.cell.2008.05.042.

29. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science. 2007; 315(5811), 525-8. doi: 10.1126/science.1135308.