Cloning and expression of a codon-optimized gene encoding the influenza A virus nucleocapsid protein in *Lactobacillus casei*

Namfon Suebwongsa,1 Marutpong Panya,2 Wises Namwat,1 Saovaluk Sookprasert,1 Begoña Redruello,3 Baltasar Mayo,3 Miguel A. Álvarez,3 Viraphong Lulitanond1*

1Department of Microbiology and Research and Diagnostic Center for Emerging Infectious Diseases (RCEID), Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. 2College of Medicine and Public Health, Ubon Ratchathani University, Ubon Ratchathani, Thailand. 3Dairy Institute of Asturias (IPLA-CSIC), Villaviciosa, Spain

Received 6 June 2013 · Accepted 15 July 2013

Summary. Lactic acid bacteria (LAB) species are envisioned as promising vehicles for the mucosal delivery of therapeutic and prophylactic molecules, including the development of oral vaccines. In this study, we report on the expression of a synthetic nucleocapsid (NP) gene of influenza A virus in *Lactobacillus casei*. The NP gene was re-designed based on the tRNA pool and the codon usage preference of *L. casei* BL23. The codon-optimized NP gene was then cloned and expressed in *L. casei* RCEID02 under the control of a constitutive promoter, that of the lactate dehydrogenase (*ldh*) gene. The synthetic NP gene was further expressed in *L. casei* EM116 under the control of an inducible promoter, that of the structural gene of nisin (*nisA*) from *Lactococcus lactis*. Based on Western blot analysis, the specific protein band of NP, with a molecular mass of 56.0 kDa, was clearly detected in both expression systems. Thus, our study demonstrates the success of expressing a codon-optimized influenza A viral gene in *L. casei*. The suitability of the recombinant LAB strains for immunization purposes is currently under evaluation.

[Int Microbiol 2013; 16(2):93-101]

Keywords: *Lactobacillus casei* · lactic acid bacteria · influenza A virus · viral nucleocapsid proteins · heterologous expression · codon usage

Introduction

There are numerous immunological benefits associated with vaccine administration by the mucosal route [20]. Moreover, because of economic, logistic, and safety reasons, the use of oral vaccines for large-scale immunization programs is an explicit objective of the World Health Organization (WHO) [30]. Among the different strategies for mucosal immunization, lactic acid bacteria (LAB) as live delivery vehicles of vaccine antigens have important advantages [31]. LAB, especially *Lactobacillus* and *Lactococcus* species, have been successfully engineered to express antigenic proteins that could serve as live oral vaccines against various infectious diseases [3]. Several genes derived from bacterial [15], viral [14,19,22], parasitic [9], eukaryotic origin [32], and allergic proteins [24] have been cloned and expressed in these bacteria. However, the low level of production of the heterologous proteins is
one of the major drawbacks and might be a crucial limitation using LAB as an alternative vaccine system. Several strategies have been examined to improve the yield of the expressed proteins, including the use of strong promoters (of both constitutive and inducible classes), high-copy-number plasmid vectors, and efficient Shine–Dalgarno sequences [12]. Codon usage biases in LAB, including that in Lactobacillus casei, were shown to be an important factor affecting the translational efficiency of heterologous proteins [8]. The capability of heterologous protein expression is known to depend on the ability of bacterial tRNA to recognize the codon(s) of newly introduced genes [23]. Some codons not frequently present in L. casei genes (rare codons), such as AGG and AGA, both of which code for arginine (Arg), occur frequently in human genes [27]. This finding suggests the need to modify the coding sequence of the introduced gene according to the sequences frequently used by the bacterial host. For instance, the expression yield of a heterologous protein in Escherichia coli was increased up to 70 % compared to that of wild-type DNA sequence by using a codon-optimized gene [16].

Several complete genomes of LAB have now been sequenced, of which 59 are from the genus Lactobacillus [http://www.ncbi.nlm.nih.gov/genome/browse/]. Genome sequence sizes range from the 1.26 megabase pairs (Mbp) of Lactobacillus plantarum to the 3.43 Mbp of Lactobacillus rhamnosus. Comparative analysis and gene annotation of LAB genomes has allowed the prediction of tRNA in Lactobacillus species [17], which may help in codon optimization and therefore the success of heterologous protein expression in these bacteria.

The influenza virus is a major health problem throughout the world. Although presently available influenza vaccines stimulate the synthesis of neutralizing antibodies against the viral hemagglutinin (HA) and neuraminidase (NA) [2] proteins, they only confer subtype-specific protection immunity. This is a major drawback, as the influenza virus undergoes continuous antigenic changes, which in turn requires continuous influenza surveillance and an annual vaccination program to update protection against the circulating subtype [4]. Consequently, there is an important need to develop a universal influenza vaccine containing conserved viral component(s) eliciting broad-spectrum immunity against all viral subtypes. One of the most conserved influenza components among the different subtypes is the nucleocapsid (NP) protein. Studies on vaccines obtained by expressing the influenza NP gene have shown that cross-protective immune responses are possible [7]. Therefore, influenza vaccines based on the conserved NP protein offer an alternative to confer protection against many, if not all, subtypes of influenza viruses.

In this study, we constructed two recombinant L. casei strains expressing a synthetic NP gene of the influenza A virus. The NP coding sequence was optimized based on the rRNA pool and the codon used by L. casei BL23. The codon-optimized NP gene was cloned and expressed in L. casei under the control of both constitutive and inducible promoters. The product of the synthetic gene in both cases was revealed by using commercial antibodies and Western blot, proving the success of our approach.

Materials and methods

Bacterial strains, culture media, and growth conditions.

Table 1 lists the bacterial strains, cloning vectors, and primers used in this study. Escherichia coli strains were grown in Luria-Bertani (LB) broth (Difco, Franklin Lakes, NJ, USA) at 37 °C with shaking. Lactobacillus casei strains were cultured statically in de Man Rogosa and Sharpe (MRS) medium (Difco) at 37 °C. When needed, 15 g of biological grade agar (Difco) was added to the corresponding medium. Ampicillin (100 μg/ml) and erythromycin (2.5 μg/ml) (Sigma-Aldrich, St. Louis, MO, USA) were added as needed to the media for the selection of transformants in E. coli and L. casei, respectively.

White/blue colony screening was performed for E. coli XL1-blue on LB plates supplemented with the appropriate antibiotic, 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (20 mg/ml; X-Gal, Sigma-Aldrich), and isopropyl β-D-thiogalactopyranoside (0.5 M; IPTG, Sigma-Aldrich).

Isolation of plasmid DNA. Plasmid DNA from L. casei was isolated and purified as previously described [25]. Plasmid DNA from E. coli was isolated and purified using the HiYield plasmid mini kit according to the manufacturer’s instructions (RBC Bioscience, Taiwan).

PCR amplification. The Taq DNA polymerase-based amplification was performed in 50 μl of PCR mixture, containing 50 mM KCl, 75 mM Tris-HCl (pH 9.0), 20 mM (NH₄)₂SO₄, 1.5 mM MgCl₂, 200 μM of each dNTP (dATP, dCTP, dGTP, and dTTP), 0.2 μM of each primer, and 1 U of Taq DNA polymerase. All reagents were molecular grade and were purchased from Invitrogen, Carlsbad, CA, USA. The 576-bp of nica, the multiple cloning site (MCS), and the transcription terminator (TT) were PCR-amplified under the following conditions: pre-denaturation (94 °C), denaturation (94 °C), annealing (60 °C) and final extension (72 °C) for 3 min, 30 s, 30 s, 40 s, and 7 min respectively. The 1842-bp fusion gene consisting of the nucleocapsid gene and the transcription terminator (NP:TT) was amplified under the following conditions: pre-denaturation (94 °C), denaturation (94 °C), annealing (60 °C), extension (68 °C) and final extension (68 °C) for 5 min, 30 s, 30 s, 1.30 min, c and 7 min, respectively.

DNA handling and transformation. The PCR products derived from PCR amplification, or DNA fragments embedded in agarose gels, were purified by using the HiYield Gel PCR DNA Fragments Extraction Kit (RBC). Electroporation of E. coli and L. casei were carried out as described by [6] and [5], respectively. All other DNA manipulations were essentially performed as described [28].

Verification of the constructs. The correct nucleotide sequence and orientation of the cloned genes were confirmed by DNA sequencing with a MegaBACE 1000 sequencer (Biodesign, Bangkok, Thailand) using specific primers (Table 1).
Codon optimization of the NP gene for expression in *Lactobacillus casei*. The native NP gene sequence of influenza A virus [A/NewYork/31/2004(H3N2)] (GenBank accession no. CY000372) was retrieved from the NCBI database [http://www.ncbi.nlm.nih.gov/nuccore/CY000372]. The codon usage of the NP gene was compared with that of *L. casei* BL23 using a reference codon usage database [http://gib.genes.nig.ac.jp/single/codon/main.php?spid=Lcas_BL23]. The codon-optimized NP gene was synthesized (Integrated DNA Technologies, San Diego, CA, USA) with 5’ and 3’ ends having NcoI (CCATGG) and XbaI (TCTAGA) sites, respectively. The synthetic gene was supplied as a clone in pUC19, designated pIDTSMART-AMP-NP. The nucleotide sequence of the codon-optimized NP gene was deposited in NCBI database under accession number KC496021.

**Determination of NP expression in *Lactobacillus casei* under the nisA and ldhL promoters.** Both the recombinant *L. casei* EM116 containing pLC13.9:NisA:NP:TT, designated as *L. casei* EM116:NP, and the
recombinant *L. casei* RCEID02 containing pLC13.9:LdhL:NP:TT, designated as *L. casei* RCEID02:NP, were examined for the production of NP protein. For the nisA-based inducible expression system, recombinant EM116:NP was cultured in MRS medium supplemented with 2.5 μg erythromycin/ml at 37 °C. After the OD₆₀₀ reached 0.3, nisin inducer was added to the culture at a final concentration of 10 ng/ml. NP expression was monitored by sampling the bacterial cultures 1, 3, 5, 7, and 9 h after induction. The bacterial samples were washed twice with phosphate buffer saline (PBS, pH 7.0), and suspended in 300 μl of lysis buffer (0.5 M Na₂HPO₄, 5 M NaCl, 1 M imidazole [Qiagen, Hilden, Germany]), 100 mg lysozyme (Amresco, Salon, OH, USA)/ml, 1× protease inhibitor cocktail (Amresco), and 1 M dithiothreitol [Amresco]). The bacterial cells were further broken by sonication with 10 pulses of 30 s each with intermittent cooling. The whole-cell lysate was subjected to SDS-PAGE using 10 % polyacrylamide gels, followed by Western blot. The expressed NP was immunodetected by the sequential addition of mouse monoclonal anti-H1N1 influenza A virus nucleocapsid protein (Abcam, Cambridge, UK) at a dilution of 1/3000, goat anti-mouse IgG conjugated to horseradish peroxidase (HRP) (Abcam) at a dilution of 1/15,000, and chemiluminescence substrate (SuperSignal West Pico Substrate; Thermo Fisher Scientific, Rockford, IL, USA.). The crude cell lysate derived from MDCK cells (kindly provided by Dr. Parvapan Bhattarakosol, Faculty of Medicine, Chulalongkorn University, Thailand) infected with influenza A virus H1N1 and the whole-cell lysates of *L. casei* strains EM 116 and RCEID02, containing the pRCEID-LC-13.9 empty vector, were used as positive and negative control, respectively.

For the *ldhL*-based constitutive expression system, recombinant RCEID02:NP was cultured in MRS media supplemented with 2.5 μg erythromycin/ml at 37 °C for 18 h. After incubation, the cells were harvested and the processes for sample preparation and detection of the expressed NP were followed as described above.

**Results**

**Codon optimization of the influenza A virus NP gene for expression in *Lactobacillus casei***

We previously cloned and expressed a native NP gene of influenza A virus in *L. casei* under the control of expression cassettes cloned into the pRCEID-LC13.9 cloning vector (Panuya et al., unpublished). The expression cassette included promoter and transcription termination sequences of the *L. casei* lactate dehydrogenase gene. Despite attempts to optimize heterologous protein expression in this species, Western blot analysis failed to show expression of the native NP gene. A literature survey suggested that heterologous gene expression could be enhanced by replacing native codons of the introduced gene with those naturally used by the host for its own highly expressed genes. Thus, in this study, the native codons of the NP gene were replaced with those preferentially used by *L. casei* BL23, as shown in Table 2.

The complete NP gene of influenza A virus is 1497 bp long, with 499 codons that are translated into 498 amino acids. Bioinformatics analysis of the viral NP codons showed that 44 codons are rare codons in *L. casei*.

**Table 2.** Native- and codon-optimized nucleocapsid protein gene

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Native*</th>
<th>Optimized</th>
<th>Amino acid</th>
<th>Native*</th>
<th>Optimized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg (R)</td>
<td>AGA (22); AGG (14); CGA (4); CGC (3); CGG (5); CGT (1)</td>
<td>CGT</td>
<td>Ser (S)</td>
<td>AGC (8); AGT (8); TCA (3); TCC (6); TCG (2); TCT (12)</td>
<td>TCA</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>GGA (18); GCC (7); GGG (11); GGT (6)</td>
<td>GGC</td>
<td>Val (V)</td>
<td>GTA (4); GTC (5); GTG (9); GTT (4)</td>
<td>GTT</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>CTA (4); CTC (11); CTG (6); CTT (6); TTA (2); TTTG (4)</td>
<td>TTG</td>
<td>Ala (A)</td>
<td>GCA (18); GCC (7); GCG (4); GCT (10)</td>
<td>GCC</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>CCA (5); CCC (2); CCG (1); CCU (9)</td>
<td>CCA</td>
<td>Ile (I)</td>
<td>ATA (8); ATC (13); ATT (6)</td>
<td>ATC</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>TGC (5); TGT (1)</td>
<td>TGC</td>
<td>Gin (Q)</td>
<td>CAA (15); CAG (6)</td>
<td>CAA</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>ACA (8); ACC (7); ACG (3); ACT (10); ACT (10)</td>
<td>ACC</td>
<td>Asn (N)</td>
<td>AAC (7); AAT (20)</td>
<td>AAC</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>AAA (14); AAG (7)</td>
<td>AAA</td>
<td>Asp (D)</td>
<td>GAC (11); GAT (11)</td>
<td>GAT</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>TTC (11); TTT (7)</td>
<td>TTT</td>
<td>His (H)</td>
<td>CAC (3); CAT (3)</td>
<td>CAT</td>
</tr>
<tr>
<td>Tyr (Y)</td>
<td>TAC (9); TAT (6)</td>
<td>TAT</td>
<td>Glu (E)</td>
<td>GAA (23); GAG (13)</td>
<td>GAA</td>
</tr>
<tr>
<td>Met (M)</td>
<td>ATG (25)</td>
<td>ATG</td>
<td>Trp (W)</td>
<td>TGG (6)</td>
<td>TGG</td>
</tr>
</tbody>
</table>

*In parentheses, the number of each codon found in the native NP sequence. Rare codons found in the native NP gene are shown in bold.
and 8 ATA codons (Table 2). For codon optimization of the NP gene, the AGA and AGG, which code for arginine (R), were replaced with CGT; ATA, which codes for isoleucine (I), was replaced with ATC. In addition, some of the remaining native NP codons were replaced with the corresponding codons predicted to be highly used by *L. casei* BL23. For example, all codons that code for arginine, i.e. CGA, CGC, CGG, including two rare codons (AGA and AGG), were replaced with the optimized codon CGT. With this approach, a total of 321 codons were changed, as shown in Table 2. Some restriction endonuclease sites in the native NP gene, including four *Pst* I, three *Hind* III, two *Bam* HI and a single *Eco* RV site, were removed by this replacement. To facilitate cloning of the optimized NP gene into the expression vector, the gene was synthesized with the addition of the *NcoI* (CCATGG), overlapping the initial ATG codon, and *XbaI* (TCTAGA) recognition sequences at the 5'- and 3'-end, respectively. This optimized NP gene, with a total length of 1505 base pairs, was cloned into *pUC19* digested with both *NcoI* and *XbaI*, and the resulting construct was designated *pIDTSMART-AMP-NP*. The synthesis was externally constructed (Integrated DNA Technologies, San Diego, CA, USA).
Construction of recombinant expression plasmids containing the codon-optimized NP gene.

In this study, two NP-expression systems were constructed, one based on the nisin gene (nisA)-inducible promoter and the other on the constitutive promoter of the lactate dehydrogenase gene (ldhL). For the NisA-based expression system (Fig. 1), the PCR product containing the nisA gene promoter, the multiple cloning site (MCS), and the transcription terminator signals (TT) from pNZ8048 were amplified with the primer pair pnisAatII and pnisNdeI (Table 1) and the obtained amplicons were cloned into the pGEM-T-Easy vector to generate pNisA. The codon-optimized NP gene was released from pIDTSMART-AMP-NP by NcoI and XbaI digestion and the fragment was isolated and purified from an agarose gel. The DNA segment was cloned into pNisA digested with the same restriction enzymes; the resulting construct was named pNisA:NP:TT. A DNA fragment from this construct digested with AatII and NdeI was isolated from a gel and subcloned into the E. coli/L. casei shuttle vector pRCEID-LC13.9 (Table 1) digested with the same enzymes. The new construct was designated pLC13.9:NisA:NP:TT.
For the construction of the ldhL-based expression system (Fig. 2), the NP:TT fragment was amplified from pNisA:NP:TT using the primer pair NPsynF1 and pnisNdeI (Table 1) and then inserted into KpnI/NdeI-digested pGEM:LdhL:GFPuv (Table 1) to generate pLdhL:NP:TT. The LdhL:NP:TT fragment was isolated and purified from pLdhL:NP:TT after digestion with AatII and NdeI and cloned into the pRCEID-LC13.9 vector. The final construct, designated pLC13.9:LdhL:NP:TT, and pLC13.9:NisA:NP:TT were verified by DNA sequencing and analysis. Purified plasmids from these verified constructs were transformed into L. casei EM116 and RCEID02, respectively, to generate the recombinant lactobacilli strains L. casei EM116:NP and L. casei RCEID:NP, carrying the former and the latter construct, respectively. As the negative control, the empty plasmid vector, pRCEID-LC13.9, was also transformed into both EM116 and RCEID02.

Expression of NP protein in Lactobacillus casei under different expression systems. Western blotting using a specific antibody against NP showed that both L. casei EM116:NP and RCEID:NP expressed NP (Fig. 3A,B). The expression of NP in the nisin inducible system was confirmed by the detection of a protein with the expected size of 56 kDa 3 and 5 h after nisin induction (10 ng nisin/ml). Expression declined slightly at 7 h and was completely absent at 9 h (Fig. 3A). Figure 3B shows NP expression by L. casei strain RCEID02:NP. Taken together, our results indicated that codon-optimization is a successful strategy for the expression of the NP gene in L. casei, both under the nisA and the ldhL promoters.

Using the same constructs, we further analyzed the expression of codon-optimized NP gene in E. coli XL1. As shown in Fig. 3B, recombinant E. coli cells containing pLC13.9:LdhL:NP:TT (XL1:NP) synthesized the 56-kDa NP protein under the ldhL promoter, whereas expression was not observed in E. coli under the control of the nisA-inducible expression system (data not shown).

Discussion

Currently, heterologous gene expression is being actively investigated in LAB as live vehicles for the mucosal delivery of therapeutic and prophylactic proteins. However, the low-level expression of heterologous genes in these bacteria is an important drawback that limits their application for such purposes.

Codon usage bias in LAB has been shown to reduce translational efficiency and, thus, the total amount of protein synthesized [8]. The expression level of genes containing codons matching those of the bacterial hosts is higher than that of genes containing native codons [16]. In L. casei, three codons (AGG and AGA, coding for arginine, and ATA, coding for isoleucine) present in the wild-type NP gene are rare codons...
Bioinformatic analysis showed that the influenza A virus NP gene contains 22 AGA, 14 AGG, and 8 ATA codons. All these rare codons were replaced by the codons used by *L. casei* BL23. To obtain more effective translation, additional codon replacement was achieved with codons that are frequently used by *L. casei* BL23 (Table 2). Codon optimization thus resulted in the removal of four *PstI*, three *HindIII*, two *BamHI*, and a single *EcoRV* site from the optimized NP gene. The loss of these restriction sites allowed for greater flexibility in the choice of restriction sites used for cloning the NP gene in the cloning vector. In a previous study, we were unable to obtain the expression of the native NP coding sequence as protein in *L. casei*, although expression was not quantitatively determined (data not shown). As reported previously by other authors, the translation of heterologous genes containing a high percentage of rare codons is generally stalled or prematurely terminated, which results in a low level of expression [1].

Since the stability (both segregational and structural) of a plasmid in a bacterial host is a necessary requirement in the generation of live vaccines. In this study we used the pRCEID-LC13.9 *E. coli/L. casei* shuttle vector, which is highly stable in lactobacilli, as the backbone plasmid [26]. Codon-optimized NP was cloned under the control of the lactate dehydrogenase (*ldhL*) gene promoter and under the *nisA* promoter in *L. casei* RCEID02 and EM116, respectively. Further stability was assured by the presence of a rho-independent terminator at the 3′N-terminal of the NP gene. Note that the growth rate of transformants containing recombinant plasmids was similar to that of wild-type lactobacilli. Furthermore, the constructs were highly stable, without appreciable gene rearrangements (structural stability) or plasmid loss during bacterial growth (segregational stability), even after 80 generations (data not shown).

The *ldhL* promoter is a constitutive gene expression system that does not require an inducer, and it has been successfully used for both homologous and heterologous gene expression in many lactobacilli species [10,11,29]. The *nisA*-based expression system, by contrast, is a regulated gene expression system that is often used to prevent the toxicity of the expressed proteins [21]. Thus, the expression of a given gene can be modulated by the use of a convenient expression system.

Western blot assays showed that the specific protein band of NP, with a molecular mass of 56.0 kDa, was clearly detected when expressed in either system. However, higher levels of expression were obtained with the *nisA* expression system in *L. casei* EM116. Our study therefore clearly demonstrates the cloning and expression of an influenza A gene in *L. casei*. Similar results were previously reported for the expression in *L. casei* of the VP60 protein of Norwalk virus [19].

In conclusion, in this study we describe the use of codon-optimization in the NP gene of influenza A virus, which allowed its successful cloning and expression in *L. casei* strains, using the *E. coli/L. casei* shuttle vector pRCEID-LC13.9. Expression of the NP gene in *L. casei* would facilitate the use of this bacterium in the immunization against influenza A virus.

**Acknowledgement.** This study was supported by the Higher Education Research Promotion and the National Research University Project of Thailand, Office of the Higher Education Commission, and by The National Center for Genetic Engineering and Biotechnology, Thailand.

**Competing interests.** None declared.

**References**