GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Molecular cloning and bioinformatic analysis of the *Streptococcus agalactiae neuA* gene isolated from tilapia

E.L. Wang[1], K.Y. Wang[1,2], D.F. Chen[1,3], Y. Geng[1,2], L.Y. Huang[1], J. Wang[1] and Y. He[1]

[1]Department of Basic Veterinary Medicine, College of Veterinary Medicine, Sichuan Agricultural University, Ya'an, Sichuan, China
[2]Key Laboratory of Animal Disease and Human Health of Sichuan Province, Sichuan Agricultural University, Ya'an, Sichuan, China
[3]Department of Aquaculture, College of Animal Science and Technology, Sichuan Agricultural University, Ya'an, Sichuan, China

Corresponding author: K.Y. Wang
E-mail: kywangsicau@126.com

Genet. Mol. Res. 14 (2): 6003-6017 (2015)
Received August 26, 2014
Accepted January 12, 2015
Published June 1, 2015
DOI http://dx.doi.org/10.4238/2015.June.1.18

ABSTRACT. Cytidine monophosphate (CMP) N-acetylneuraminic acid (NeuNAc) synthetase, which is encoded by the *neuA* gene, can catalyze the activation of sialic acid with CMP, and plays an important role in *Streptococcus agalactiae* infection pathogenesis. To study the structure and function of the *S. agalactiae neuA* gene, we isolated it from diseased tilapia, amplified it using polymerase chain reaction (PCR) with specific primers, and cloned it into a pMD19-T vector. The recombinant plasmid was confirmed by PCR and restriction enzyme digestion, and identified by sequencing. Molecular characterization analyses of the *neuA* nucleotide amino acid sequence were performed using bioinformatic tools and an online server. The results showed that the *neuA* nucleotide sequence contained a complete coding region, which comprised 1242 bp, encoding 413 amino acids (aa). The aa sequence was highly conserved and contained a Glyco_tranf_GTA_type superfamily and an SGNH_hydrolase superfam-

ily conserved domain, which are related to sialic acid activation catalysis. The NeuA protein possessed many important sites related to post-translational modification, including 28 potential phosphorylation sites and 2 potential N-glycosylation sites, had no signal peptides or transmembrane regions, and was predicted to reside in the cytoplasm. Moreover, the protein had some B-cell epitopes, which suggests its potential in development of a vaccine against *S. agalactiae* infection. The codon usage frequency of *neuA* differed greatly in *Escherichia coli* and *Homo sapiens* genes, and *neuA* may be more efficiently expressed in eukaryotes (yeast). *S. agalactiae neuA* from tilapia maintains high structural homology and sequence identity with CMP-NeuNAc synthetases from other bacteria.

Key words: Cloning; Bioinformatic analyses; *Streptococcus agalactiae*; *neuA* gene; NeuA amino acid sequences

INTRODUCTION

Fish streptococcosis, which is characterized by spinning, pop-eye, and hemorrhage, is caused by *Streptococcus* infection and is a contagious and highly lethal disease. Streptococcal disease in fish was first reported in 1957 in cultured rainbow trout in Japan (Hoshina et al., 1958). Since then, numerous other species of fish have been found to be susceptible to the infection, including salmon, mullet, golden shiner, pinfish, eel, sea trout, tilapia, sturgeon, and striped bass (Kitao, 1993). In addition, some streptococci species such as Group A streptococci and Group B streptococci can also cause disease in humans. *Streptococcus agalactiae* was first reported to cause disease in fish in 1966, infecting golden shiners in USA (Robinson and Meyer, 1966). Since then, many researchers have isolated the pathogen from rainbow trout, dentex, tilapia, catfish, menhaden, mullet, and silvery pomfret (Amal and Zamri-Saad, 2011). *S. agalactiae* infection has caused tremendous economic losses and brought obstacles to the steady development of tilapia culture, with high infectiousness, fast outbreak, and high mortality.

In the summer of 2009, there was an outbreak of tilapia streptococcosis in Hainan Province, China, and the cumulative mortality reached 85%. A Gram-positive bacterium (designated as 3-BY) was isolated and identified as *S. agalactiae* on the basis of its morphological, biochemical, and physiological properties, and through 16S rDNA analysis in our laboratory. Challenge experiments following Koch's postulates which included four criteria to identify the causative agent of an infectious disease revealed that *S. agalactiae* was the pathogen in the Hainan outbreak (Huang et al., 2011). However, more research is required on the pathogenesis of *S. agalactiae.*

Previous research conducted on *S. agalactiae* virulence factors and immunogenic components included work on sialic acid-capped capsular polysaccharides (Wessels, 1997), the C proteins (Payne et al., 1987), surface immunogenic proteins (Brodeur et al., 2000), and C5a peptidase (Cheng et al., 2002). The glycohydrolytic enzyme sialidase can specifically catalyze the release of terminal sialic acid residues from various sialo-derivatives such as glycoproteins and glycolipids (Saito and Robert, 1995), which limit the deposition of alternative complement pathway component C3b onto the bacterial surface. This inhibition results in decreased phagocytosis by macrophages and neutrophils (Marques et al., 1992), thus enhancing bacterial survival in the host. The group of genes encoding sialidase comprises the genes *neuA*, *neuB*, *neuC*, and *neuD*, and NeuA sialidase can catalyze the activation of sialic acid (Haft et al., 1996). Therefore, NeuA

sialidase plays an important role in the pathogenesis of *S. agalactiae* infection. However, there is little information about the molecular characteristics of NeuA sialidase or its encoding gene.

The objective of this study was to clone and identify the *neuA* gene of *S. agalactiae* 3-BY, and report the bioinformatic analysis of the *neuA* nucleotide sequence and derived amino acid sequence. We also intend to conduct some related experiments to study the immunogenicity and protection of NeuA sialidase as a candidate vaccine based on this study. These studies might provide some insight into the gene and the corresponding protein for further research, and lay the foundation of genetic information for the study of *neuA* gene function, offering a theoretical basis for screening NeuA as a vaccine for *S. agalactiae* infection in tilapia.

MATERIAL AND METHODS

S. agalactiae genomic DNA extraction

S. agalactiae strain 3-BY was recovered in brain heart infusion broth at 37°C for 24 h. The genomic DNA of *S. agalactiae* was then extracted using a bacterial genomic DNA extraction kit, which was purchased from the Tiangen Biotech Company, China, according to manufacturer instructions.

Polymerase chain reaction (PCR) amplification of the *neuA* gene

The coding region of the *neuA* gene was amplified with specific primers by PCR. The forward primer (P1) 5'-GGATCCATGAAGCCAATTTGTATTAT-3' and the reverse primer (P2) 5'-CTCGAGTTATAAGGTTTTAACTTCGTC-3' contained the *Bam*HI and *Xho*I restriction sites (underlined), respectively. The two primers were synthesized by TaKaRa Bio Inc. (Dalian, China). PCR was carried out in a 25-μL reaction mixture containing 1.0 μL of each primer (100 μM each), 1.0 μL DNA template, 12.5 μL Master Mix, and 9.5 μL double-distilled H_2O. The PCR conditions were: 95°C for 5 min; 30 cycles of 94°C for 1 min each; 55°C for 1 min; 72°C for 1.5 min; and then a final extension at 72°C for 10 min. The PCR products were fractionated on 1.0% agarose gel by electrophoresis and stained with GoldView (TIANGEN, Beijing, China).

Cloning and sequencing of the *neuA* gene

The PCR products were purified using the Agarose Gel DNA Extraction Kit (TaKaRa). The purified PCR products were cloned into the pMD19-T vector, followed by transformation into *Escherichia coli* DH5α-competent cells. The positive recombinant clone was then selected using an Amp/IPTG/X-Gal agar plate. The recombinant plasmid was identified by PCR after the aforementioned conditions, digested with restriction enzymes *Bam*HI and *Xho*I, and fractionated on 1% agarose gels. DNA sequencing was also conducted by TaKaRa Bio Inc.

Bioinformatic analyses of *neuA* nucleotide and amino acid sequences

Alignment and base composition analysis of the nucleotide sequence of the *neuA* gene were performed with the BLASTn software (Gotea et al., 2003) and DNASTAR version 7.0, respectively. The physical and chemical properties of the deduced amino acid sequence from the *neuA* gene, including the amino acid composition, relative molecular mass, and isoelectric point,

were calculated using the ProtParam software from the ExPASy online system (Gasteiger et al., 2003). Conserved domains in the NeuA protein were identified using the National Center for Biotechnology Information (NCBI) Conserved Domains software (Marchler-Bauer et al., 2005). The phylogenetic analysis of the NeuA protein and the construction of the NeuA evolutionary tree were carried out using the Clustal X 2.0 and MEGA 5.0 softwares. Hydrophobicity analysis was performed using the Bioedit version 7.0 software, followed by solubility prediction at http://www.biotech.ou.edu/ (Smialowski et al., 2007). Phosphorylation sites and N-glycosylation sites were predicted using NetPhos version 2.0 and NetNGlyc version 1.0 online programs, respectively (Blom et al., 1999). Transmembrane region prediction was carried out using the TMHMM online program (Möller et al., 2001), followed by signal peptide searching using the SignalP 4.0 Server online system (Petersen et al., 2011). Protein subcellular localization prediction was performed with PSORTb v3.0 (Yu et al., 2010). Secondary structure prediction of the NeuA protein, i.e., α-helix, extended strand, and random coil prediction, was carried out by accessing the web server PSIPRED (Jones, 1999). The prediction of the three-dimensional structure of the NeuA protein was performed using the SWISS-MODEL online program (Schwede et al., 2003).

The B-cell epitopes of the NeuA protein were directly predicted using the BCPREDS online server at http://crdd.osdd.net/raghava/lbtope (El-Manzalawy et al., 2008), and indirectly predicted on the basis of secondary structures, flexibility, hydrophilicity, and surface probability of the NeuA protein using the DNASTAR software (Burland, 2000).

Codon usage bias analysis of the *neuA* gene

The effective number of codons (ENC) value of the *neuA* gene was computed using the European Molecular Biology Open Software Suite (EMBOSS) CHIPS online service program (Comeron and Aguadé, 1998), followed by calculation of the codon usage bias of the *neuA* gene with the CUSP program of EMBOSS (Popov et al., 2009). To examine whether different species followed the same codon usage rule, we compared *neuA* gene codon usage bias with *E. coli*, yeast, and *Homo sapiens*. The database of the codon usage in *E. coli*, yeast, and *H. sapiens* is available at http://www.kazusa.or.jp/codon (Nakamura et al., 2000).

RESULTS

Amplification, cloning, and nucleotide sequence analysis of the *neuA* gene

To amplify the *neuA* gene, PCR was conducted on genomic DNA extracted from *S. agalactiae* 3-BY using primers P1 and P2, which are specific to the *neuA* gene. A band of about 1242 bp was observed upon electrophoresis of the PCR products on agarose gel (Figure 1A). The PCR product was purified and cloned into the pMD19-T vector, followed by identification through PCR and digestion with restriction enzymes *Bam*HI and *Xho*I (Figure 1B). Thus, the positive recombinant plasmid was constructed and designated as pMD19-T-*neuA*. We obtained the *neuA* nucleotide sequence from TaKaRa Bio Inc. by sequencing, and the sequence was aligned using the BLASTn software. Sequence analysis using DNASTAR indicated that the nucleotide sequence of the *neuA* gene was 1242 bp in length with a G+C content of 32.53%, and encoded a complete open reading frame from the start codon ATG to the stop codon TAA (Figure 2). We submitted the *neuA* nucleotide sequence to NCBI and obtained GenBank accession No. KF447585.

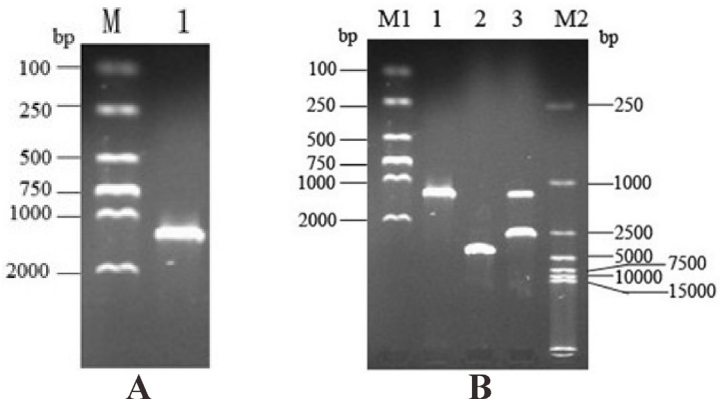

Figure 1. Polymerase chain reaction (PCR) amplification of the *neuA* gene and identification of recombinant plasmid pMD19-T-neuA by PCR and restriction enzyme digestion. **A.** PCR amplification result of the *neuA* gene. *Lane M*: DNA marker (DL2000); *Lane 1*: PCR product of the *neuA* gene (1242 bp). **B.** Identification of recombinant plasmid pMD19-T-*neuA* by PCR and restriction enzyme digestion. *Lane M1*: DNA marker (DL2000); *Lane 1*: PCR products of recombination plasmid pMD19-T-*neuA*; *Lane 2*: digestion of recombinant plasmid with *Bam*HI; *Lane 3*: digestion of recombinant plasmid with *BamH*I + *Xho*I; *Lane M2*: DNA marker (DL15000).

```
1     atgaagccaatttgtattattcctgcgcgatcagggtcaaaaggattaccagataagaatatgttattcctagcaggtaaacccatgatt
      M  K  P  I  C  I  I  P  A  R  S  G  S  K  G  L  P  D  K  N  M  L  I  L  R  G  K  P  M  I
91    ttccatactatagatgctgcaattgaatctggaatgtttgataagaaagatatttttgttagtacagattcagaattgtacagagagatt
      F  H  T  I  D  A  A  I  E  S  G  M  F  D  K  K  D  I  F  V  S  T  D  S  E  L  Y  R  E  I
181   tgtttagaacgcggtatttcagtggtgatgagaaaaccggaactttcaactgatcaggcaacttcgtatgatatgttaaaagatttttta
      C  L  E  R  G  I  S  V  V  M  R  K  P  E  L  S  T  D  Q  A  T  S  Y  D  M  L  K  D  F  L
271   tctgactatgaagataatcaggagtttgtgttacttcaagtaacctctccactaagaaaatcatggcatataaaggaagcaatggagtat
      S  D  Y  E  D  N  Q  E  F  V  L  L  Q  V  T  S  P  L  R  K  S  W  H  I  K  E  A  M  E  Y
361   tattcttcacatgatgttgacaatgttgtaagtttttctgaagttgagaaacaccctggtctgtttacgacattgtctgataaaggctat
      Y  S  S  H  D  V  D  N  I  V  S  F  S  F  Y  E  K  H  P  G  L  F  T  T  L  S  D  K  G  Y
451   gctatagatatggtgggagcagataaaggttatcgtcgccaagatttacaacctttatactatccgaacggcgctatttttatttctaat
      A  I  D  M  V  G  A  D  K  G  Y  R  R  Q  D  L  Q  P  L  Y  Y  R  N  G  A  I  F  I  S  N
541   aaagaaacttacttaagggaaaaaagcttttcacctctaggacatatgcttatcaaatggcaaaggaattttcattagatgttgatacg
      K  E  T  Y  L  R  E  K  S  F  F  T  S  R  I  Y  I  Y  Q  M  A  K  E  F  S  L  D  V  I  T
631   agagatgattttatccacgtcatcggtcacctctttttcgattatgccattagggaaaaagagaataaagtttttatataagaaggctat
      R  D  D  F  I  H  V  I  G  H  L  F  F  D  Y  A  I  R  E  K  E  N  K  V  F  Y  K  E  G  Y
721   agtcgtttgttcaatagagaagcttcaaagataatttaggtgattcaaaaacgatatctatctcactagaaaattaccataattattcc
      S  R  L  F  N  R  E  A  S  K  I  I  L  G  D  E  K  T  I  S  I  S  L  E  N  Y  H  N  Y  S
811   caaggtggtgtaacattagcaacgatgttggaaaacttacctaacttttgacagctaatgtaacagaggctttgtttctatagggta
      Q  G  G  V  T  L  A  T  M  L  E  N  L  P  N  F  L  T  A  N  V  T  E  A  F  V  S  I  G  V
901   aatgatctcattacaggttatagtgtggaagaaatattcagcaatttccaaaaactctactcattattagctgagaataaaataaagatg
      N  D  L  I  T  G  Y  S  V  E  E  I  F  S  N  F  Q  K  L  Y  S  L  L  A  E  N  K  I  K  M
991   agatttacaactattgcttacacccttttagagaaactgtcaataatgcagatattgagaaaatcaatcaatggctaacagaattttgt
      R  F  T  T  I  A  Y  T  L  F  R  E  T  V  N  N  A  D  I  E  K  I  N  Q  W  L  T  E  F  C
1081  tatcaaaatcagattccactgttagatattaatagatttttatcaaaggatggtaatctaaattatcatttaactagtgatggattacat
      Y  Q  N  Q  I  P  L  L  D  I  N  R  F  L  S  K  D  G  N  L  N  Y  H  L  T  S  D  G  L  H
1171  tttactcaagaggctaatgatttgttacaaagtcaatatcaattatttgtagacgaagttaaaaccttataa
      F  T  Q  E  A  N  D  L  L  Q  S  Q  Y  Q  L  F  V  D  E  V  K  T  L  *
```

Figure 2. Nucleotide and putative amino acid sequences of *neuA*. In the nucleotide sequence, letters with gray shading indicate the start codon (atg) and the stop codon (taa), suggesting that the *neuA* nucleotide sequence comprised a complete open reading frame. In the deduced protein sequence, the putative CMP-NeuAc_Synthase domain (amino acids 2-217) and the sialate_O-acetylesterase_like2 domain (amino acids 251-411) are indicated by single and double underlining, respectively. The ligand-binding sites, tetramer interface, and active sites are highlighted in yellow, red, and green in the amino acid sequence, respectively.

Bioinformatic analyses of the NeuA protein

The *neuA* gene is expected to encode a protein comprising 413 amino acids (aa) with a molecular formula of $C_{2153}H_{3304}N_{546}O_{649}S_{14}$, a theoretical isoelectric point of 5.14, and a putative relative molecular mass of about 47.67 kDa. The protein encoded by the *neuA* gene comprises 140 hydrophobic aa, 122 hydrophilic aa, 43 basic aa, and 58 acidic aa, accounting for 33.9, 29.5, 10.4, and 14.0%, respectively. The most abundant aa in the polypeptide was Leu (9.9%), and the least abundant aa were Cys (0.7%) and Trp (0.5%) (Figure 3).

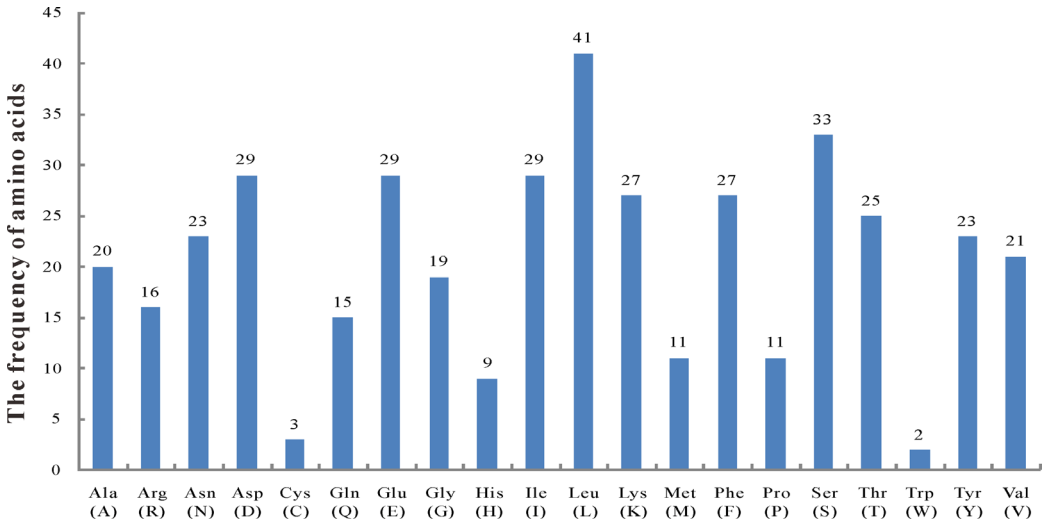

Figure 3. Amino acid composition of the peptide deduced from *Streptococcus agalactiae* NeuA.

The conserved domains analysis of the deduced *neuA* amino acid sequence using the NCBI CDD program indicated that it contained a CMP-NeuAc_Synthase conserved domain belonging to the Glyco_tranf_GTA_type superfamily and a sialate_O-acetylesterase_like2 domain which is conserved in SGNH_hydrolase or the GDSL_hydrolase superfamily at 2-217 and 251-411 aa, respectively. In addition, there were 13 ligand-binding sites (PAR^{8-10}, K^{14}, N^{20}, R^{71}, L^{75}, S^{82}, Q^{103}, T^{105}, P^{172}, F^{177}, D^{209}), 17 tetramer interface sites (F^{23}, A^{25}, G^{26}, V^{130}, E^{134}, V^{135}, $LYYP^{169-172}$, F^{177}, T^{195}, A^{197}, D^{209}, D^{212}, D^{213}, H^{216}), and five active sites containing three catalytic triad sites (S^{256}, D^{387}, H^{390}) and three oxyanion holes (S^{256}, G^{273}, N^{301}) in the NeuA peptide sequence (Figure 2).

Fourteen NeuA aa sequences of different strains were selected to conduct multiple sequence alignment (Figure 4A) using Clustal X 2.0 software, and the evolutionary tree (Figure 4B) was constructed using the Mega 5.0 software. It was clear that the evolutionary relationship between the *neuA* amino acid sequence of *S. agalactiae* 3-BY and other *S. agalactiae* strains (GenBank Nos. AAR29919, WP_000802348, AAD53077, and NP_735677) was close, with high homology; the consistency was 99.5-100% and there was a cluster in the evolutionary tree. However, a distant relationship existed with other streptococci (GenBank Nos. WP_000645006, BAM94564, and WP_003097221) with a consistency of 43.3-51.0%.

Furthermore, by using different softwares and online web servers, we gained more information about the NeuA protein, as follows. First, the result of hydrophilicity/hydropho-

bicity prediction analysis revealed that there were more hydrophilic than hydrophobic regions in the polypeptide (Figure 5A), which indicates that NeuA is a hydrophilic protein. Secondly, 28 potential phosphorylation sites (including 16 Ser phosphorylation sites, five Thr phosphorylation sites and seven Tyr phosphorylation sites) and two potential N-linked glycosylation sites were also identified (Figure 5B, 5C). Thirdly, the signal peptide prediction revealed that there was no signal peptide or transmembrane region in the NeuA aa sequence. Finally, the protein subcellular localization prediction indicated that NeuA was located in the cytoplasm, which matched 100% with N-acylneuraminate cytidylyltransferase (CMP-N-acetylneuraminic acid synthetase) (CMP-NeuNAc synthetase) (CMP-sialic acid synthetase) (GI: 27808669), and the solubility of the recombinant protein was only 61.1% when *E. coli* was selected for induced expression.

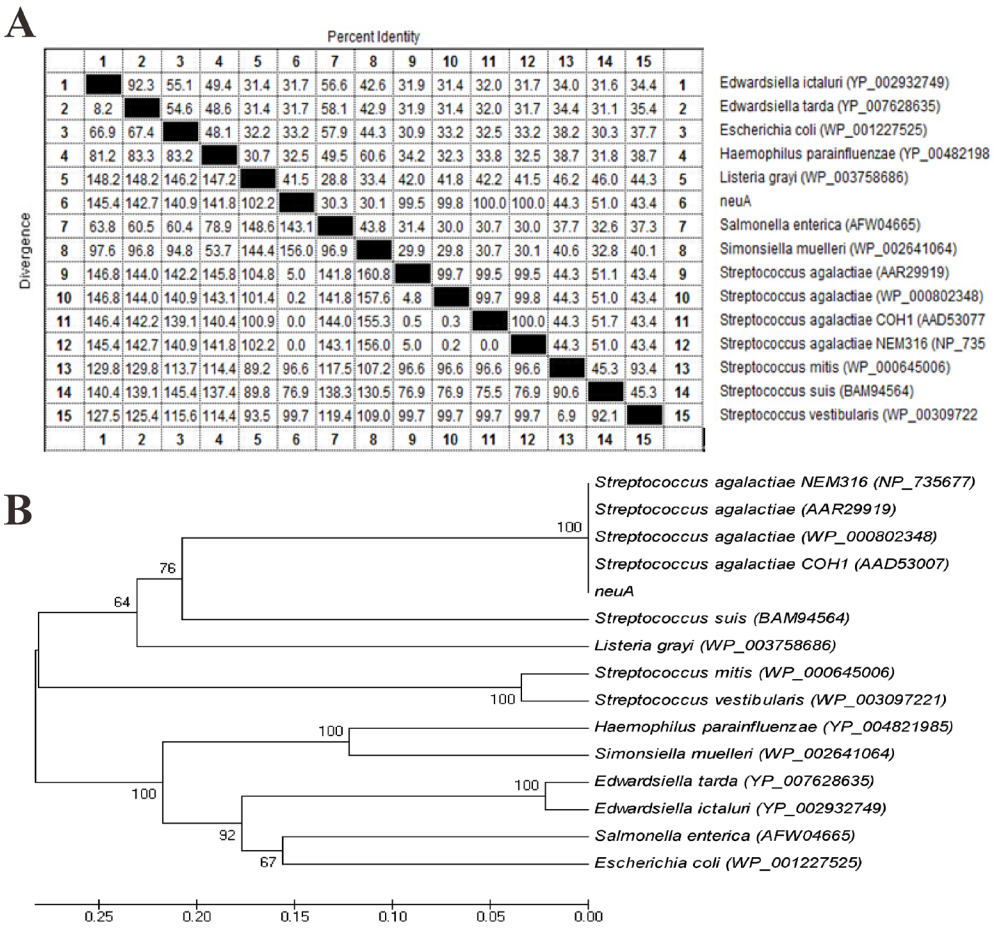

Figure 4. Percent identity matrix and phylogenetic tree analysis of *Streptococcus agalactiae* 3-BY NeuA amino acid sequences with the homologous NeuA from reference strains. **A.** Percent identity matrix analysis of NeuA amino acid sequences. **B.** Phylogenetic tree analysis of NeuA amino acid sequences. Values at the node (deduced amino acid) indicate the percentage of times that the particular node occurred in 1000 trees generated by bootstrapping the original deduced protein sequences. "*neuA*" represents the amino acid sequence of *S. agalactiae* 3-BY NeuA.

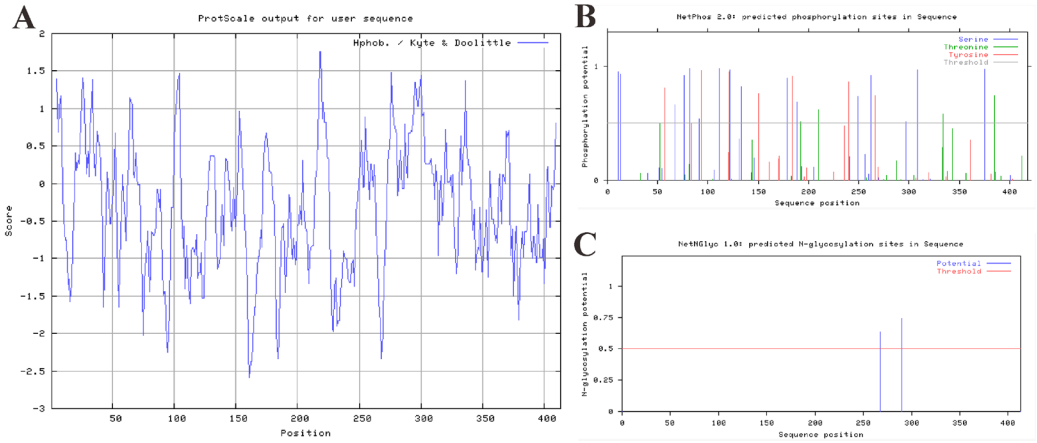

Figure 5. Result of hydrophilicity/hydrophobicity analyses and post-translational modification sites prediction of NeuA protein. **A.** Hydrophilicity/hydrophobicity analyses of NeuA protein by ProtScale. **B.** Potential phosphorylation sites prediction of NeuA protein by NetPhos 2.0. **C.** Potential N-glycosylation sites prediction of NeuA protein by NetNGlyc 1.0.

The results of the secondary structure prediction of the NeuA protein using the PSIPRED web server indicated that it comprised α-helix, extended strand, β-turn, and random coil elements with 40.19, 19.85, 6.54, and 33.41% frequency, respectively (Figure 6A). The tertiary structure prediction of the NeuA protein using the SWISS-MODEL online program showed that the three-dimensional structure of the NeuA protein (Figure 6B) was highly homologous with the sialic acid-activated synthetase NEUA_STAR1 (UniProt Database Accession Nos. P0A4V1) of *S. agalactiae* Ia serotypes.

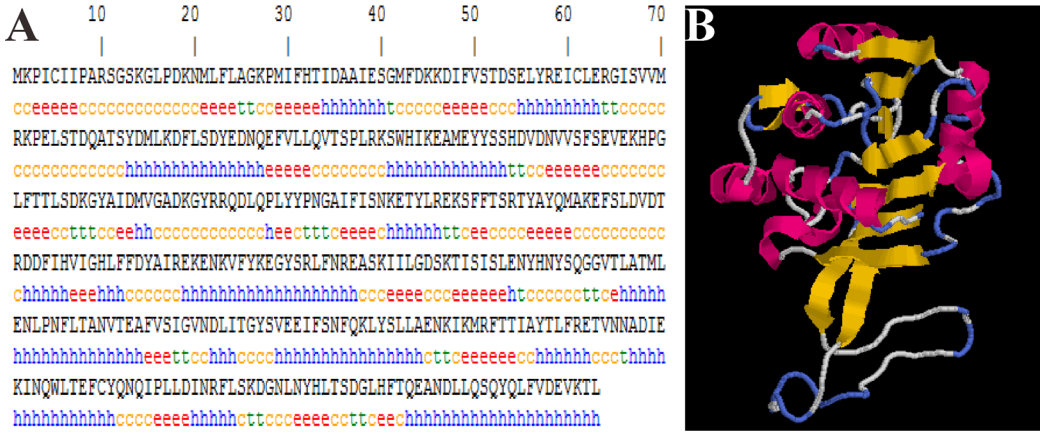

Figure 6. Secondary and tertiary structure prediction of NeuA protein. **A.** "c" indicates random coil, "e" indicates extended strand, "t" indicates β-turn, and "h" indicates α-helix. **B.** The α-helices are represented in red, extended strands are yellow, random coils are blue, and gray represents other structures.

B-cell epitope direct prediction of the NeuA protein using the BCPREDS online server revealed that the potential B-cell epitopes were located at the 3-22nd, 40-59th, 125-144th, and 155-174th aa, or their neighboring regions (Figure 7A). By using the DNASTAR software to perform a comprehensive analysis on the prediction of secondary structures, hydrophilicity, flexibility, surface probability, and antigenic index, we consistently found that the potential B-cell epitopes were located at sites 8-21, 36-56, 71-98, 134-151, 156-174, 254-272, and 369-388 of the aa sequence (Figure 7B).

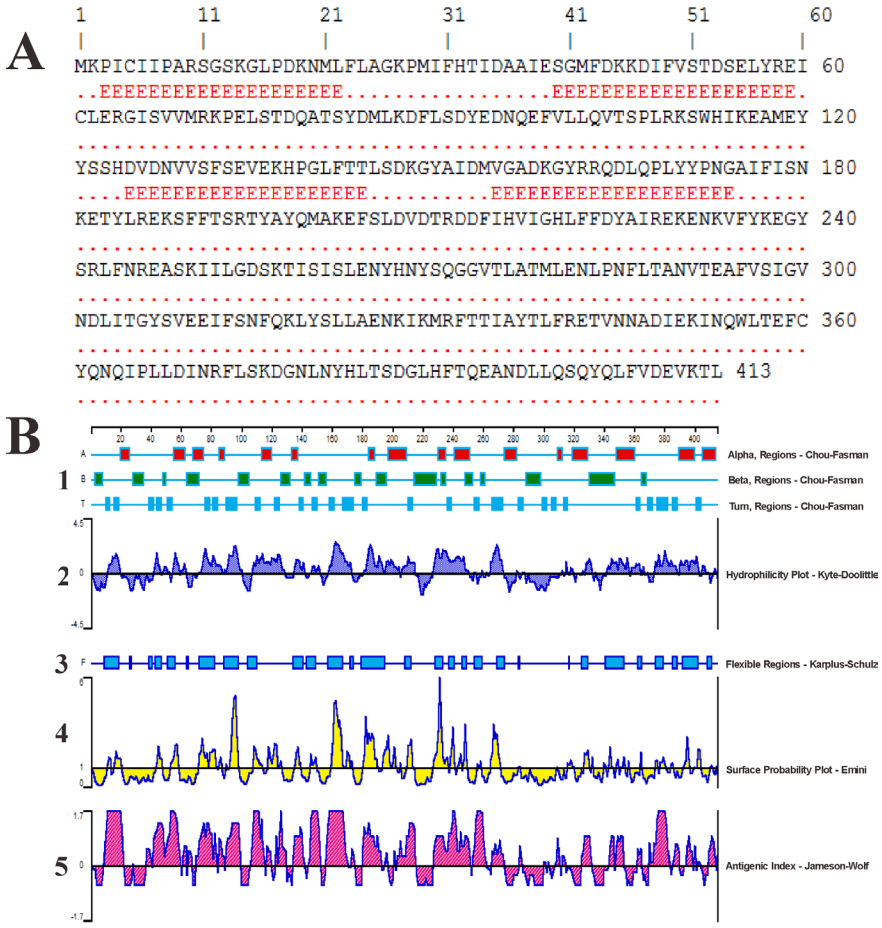

Figure 7. Direct and indirect prediction of B-cell epitopes of NeuA protein. **A** Direct prediction of B-cell epitopes of NeuA protein using the BCPREDS online server. The predicted positions of potential B-cell epitopes are represented by red 'E's. **B**. Indirect prediction of B-cell epitopes of NeuA protein using the Protean program of the DNASTAR software. **1.** Secondary structure predictions for NeuA protein. The regions of α-helix (A), β-sheet (B), and β-turns (T) are represented by red, green, and blue strips, respectively. **2.** Hydrophilicity predictions for NeuA protein. The regions of hydrophilicity are shown above the horizontal scale, whereas hydrophobicity regions are shown under the horizontal scale. **3.** Flexibility prediction for NeuA protein. The regions of flexibility are represented by blue strips. **4.** Surface probability prediction for NeuA protein. The regions of surface probability are shown above the horizontal scale. **5.** Antigenic index prediction for NeuA protein. The regions of antigenic index are shown above the horizontal scale.

Codon usage bias analysis of the *neuA* gene

In general, the ENC is used to quantify how far the codon usage of a gene departs from equal usage of synonymous codons, without dependence on sequence length or specific knowledge of preferred codons, although it is affected by base composition (Wright, 1990). Values of the ENC can range from 20 (when only 1 codon is used per amino acid) to 61 (when all synonyms are used with equal frequency). The ENC value of the *neuA* gene calculated by using the CHIPS program was 45.465, suggesting that the codon usage bias was a little higher. Therefore, it was essential to check the codon usage with the CUSP program. The results indicated that the *neuA* gene did not contain codons TGC or CGG, and the stop codon only appeared with TAA. A high level of diversity in codon usage bias existed in the *neuA* gene, with particular preference for GAT, TTA, GAA, TTT, AAT, AAA, ATT, TAT, TCA, and CAA for coding Asp, Leu, Glu, Phe, Asn, Lys, Ile, Tyr, Ser, and Gln, respectively (Table 1).

Table 1. Results of codon preferences in *Streptococcus agalactiae neuA* using the CUSP program.

Codon	AA	Fraction	Frequency /1000	No.	Codon	AA	Fraction	Frequency / 1000	No.
GCA	A(Ala)	0.4	19.324	8	**CCA***	P(Pro)	0.364	9.662	4
GCC	A	0.05	2.415	1	CCC	P	0.091	2.415	1
GCG	A	0.05	2.415	1	CCG	P	0.182	4.831	2
GCT*	A	0.5	24.155	10	**CCT***	P	0.364	9.662	4
TGC	C(Cys)	0	0	0	**CAA***	Q(Gln)	0.8	28.986	12
TGT*	C	1	7.246	3	CAG	Q	0.2	7.246	3
GAC	D(Asp)	0.103	7.246	3	**AGA***	R(Arg)	0.5	19.324	8
GAT*	D	0.897	62.802	26	AGG	R	0.188	7.246	3
GAA*	E(Glu)	0.69	48.309	20	CGA	R	0.062	2.415	1
GAG	E	0.31	21.739	9	CGC	R	0.125	4.831	2
TTC	F(Phe)	0.259	16.908	7	CGG	R	0	0	0
TTT*	F	0.741	48.309	20	CGT	R	0.125	4.831	2
GGA	G(Gly)	0.211	9.662	4	AGC	S(Ser)	0.061	4.831	2
GGC	G	0.158	7.246	3	AGT	S	0.182	14.493	6
GGG	G	0.105	4.831	2	**TCA***	S	0.394	31.401	13
GGT*	G	0.526	24.155	10	TCC	S	0.03	2.415	1
CAC	H(His)	0.333	7.246	3	TCG	S	0.03	2.415	1
CAT*	H	0.667	14.493	6	TCT	S	0.303	24.155	10
ATA	I(Ile)	0.276	19.324	8	**ACA***	T(Thr)	0.36	21.739	9
ATC	I	0.138	9.662	4	ACC	T	0.16	9.662	4
ATT*	I	0.586	41.063	17	ACG	T	0.16	9.662	4
AAA*	K(Lys)	0.704	45.894	19	ACT	T	0.32	19.324	8
AAG	K	0.296	19.324	8	GTA	V(Val)	0.286	14.493	6
CTA	L(Leu)	0.122	12.077	5	GTC	V	0.095	4.831	2
CTC	L	0.073	7.246	3	GTG	V	0.238	12.077	5
CTG	L	0.049	4.831	2	**GTT***	V	0.381	19.324	8
CTT	L	0.073	7.246	3	TGG	W(Trp)	1	4.831	2
TTA*	L	0.537	53.14	22	TAC	Y(Tyr)	0.261	14.493	6
TTG	L	0.146	14.493	6	**TAT***	Y	0.739	41.063	17
ATG	M(Met)	1	26.57	11	TAA	#	1	2.415	1
AAC	N(Asn)	0.13	7.246	3	TAG	#	0	0	0
AAT*	N	0.87	48.309	20	TGA	#	0	0	0

*Strong bias towards the codons with A or T at the third codon position. # = Stop codons.

Comparison of codon usage of *neuA* to *E. coli*, yeast, and *H. sapiens*

The results of comparison of codon usage between *neuA* to *E. coli*, yeast, and *H. sapiens* showed that there were 16 codons whose ratio was lower than 0.5 or higher than 2 between *neuA*-to-yeast, but up to 36 codons between *neuA*-to-*E. coli*, and 35 codons between *neuA*-to-*H. sapiens* (Table 2).

Table 2. Comparison of codon preferences among *Streptococcus agalactiae neuA* and *Escherichia coli*, yeast, and *Homo sapiens*.

Codon	Amino acid	*E. coli* (1/1000)	Yeast (1/1000)	*H. sapiens* (1/1000)	*neuA* (1/1000)	*neuA/E. coli*	*neuA*/Yeast	*neuA/H. sapiens*
GCA	A(Ala)	20.6	16.1	16.1	19.32	0.94	1.20	1.20
GCC	A	25.5	12.5	28.4	2.42	**0.09**	**0.19**	**0.09**
GCG	A	31.7	6.1	7.5	2.42	**0.08**	**0.40**	**0.32**
GCT	A	15.6	21.1	18.6	24.16	1.55	1.14	1.30
TGC	C(Cys)	6.9	4.7	12.2	0.00	**0.00**	**0.00**	**0.00**
TGT	C	5.5	8	10	7.25	1.32	0.91	0.72
GAC	D(Asp)	18.6	20.2	25.6	7.25	**0.39**	**0.36**	**0.28**
GAT	D	32.1	37.8	21.99	62.80	1.96	1.66	**2.86**
GAA	E(Glu)	38.2	48.5	29	48.31	1.26	1.00	1.67
GAG	E	17.7	19.1	39.9	21.74	1.23	1.14	0.54
TTC	F(Phe)	16.9	18.2	20.6	16.91	1.00	0.93	0.82
TTT	F	23.2	26.1	17.1	48.31	**2.08**	1.85	**2.83**
GGA	G(Gly)	9	10.9	16.4	9.66	1.07	0.89	0.59
GGC	G	27.9	9.7	22.5	7.25	**0.26**	0.75	**0.32**
GGG	G	11.3	6	16.3	4.83	**0.43**	0.81	**0.30**
GGT	G	24.4	24	10.8	24.16	0.99	1.01	**2.24**
CAC	H(His)	9.8	7.7	15	7.25	0.74	0.94	**0.48**
CAT	H	13.6	13.7	10.5	14.49	1.07	1.06	1.38
ATA	I(Ile)	5.4	17.8	7.7	19.32	**3.58**	1.09	**2.51**
ATC	I	24.2	17	21.6	9.66	**0.40**	0.57	**0.45**
ATT	I	29.8	30.4	16.1	41.06	1.38	1.35	**2.55**
AAA	K(Lys)	33.2	42.2	24.1	45.89	1.38	1.09	1.90
AAG	K	10.7	30.7	32.2	19.32	1.81	0.63	0.60
CTA	L(Leu)	4	13.3	7.8	12.08	**3.02**	0.91	1.55
CTC	L	11	5.4	19.8	7.25	0.66	1.34	**0.37**
CTG	L	50.9	10.4	39.8	4.83	**0.09**	**0.46**	**0.12**
CTT	L	11.7	12.1	13	7.25	0.62	0.60	0.56
TTA	L	13.9	26.7	7.5	53.14	**3.82**	1.99	**7.09**
TTG	L	14	27	12.6	14.49	1.04	0.54	1.15
ATG	M(Met)	27	20.9	22.2	26.57	0.98	1.27	1.20
AAC	N(Asn)	21.4	24.9	19.5	7.25	**0.34**	**0.29**	**0.37**
AAT	N	18.6	36.3	16.7	48.31	**2.60**	1.33	**2.89**
CCA	P(Pro)	8.5	18.2	16.7	9.66	1.14	0.53	0.58
CCC	P	5.8	6.8	20.1	2.42	**0.42**	**0.36**	**0.12**
CCG	P	21.8	5.3	6.9	4.83	**0.22**	0.91	0.70
CCT	P	7.3	13.6	17.3	9.66	1.32	0.71	0.56
CAA	Q(Gln)	15	27.5	12	28.99	1.93	1.05	**2.42**
CAG	Q	29.5	12.1	34.1	7.25	**0.25**	0.60	**0.21**
AGA	R(Arg)	2.9	21.3	11.5	19.32	**6.66**	0.91	1.68
AGG	R	1.9	9.2	11.4	7.25	**3.81**	0.79	0.64
CGA	R	3.9	3	6.3	2.42	0.62	0.81	**0.38**
CGC	R	21	2.6	10.7	4.83	**0.23**	1.86	**0.45**
CGG	R	6.3	1.7	11.6	0.00	**0.00**	**0.00**	**0.00**
CGT	R	20.3	6.5	4.6	4.83	**0.24**	0.74	1.05
AGC	S(Ser)	16	9.7	19.3	4.83	**0.30**	0.50	**0.25**
AGT	S	9.5	14.2	11.9	14.49	1.53	1.02	1.22
TCA	S	7.8	18.8	12	31.40	**4.03**	1.67	**2.62**
TCC	S	8.9	14.2	17.6	2.42	**0.27**	**0.17**	**0.14**
TCG	S	8.7	8.5	4.4	2.42	**0.28**	**0.28**	0.55
TCT	S	8.7	23.5	14.7	24.16	**2.78**	1.03	1.64
ACA	T(Thr)	8.2	17.8	15.1	21.74	**2.65**	1.22	1.44
ACC	T	22.8	12.6	19.4	9.66	**0.42**	0.77	0.50
ACG	T	14.8	7.9	6.1	9.66	0.65	1.22	1.58
ACT	T	9.1	20.3	13	19.32	**2.12**	0.95	1.49
GTA	V(Val)	11.1	11.8	7.2	14.49	1.31	1.23	**2.01**
GTC	V	15.1	11.6	14.6	4.83	**0.32**	**0.42**	**0.33**
GTG	V	25.5	10.6	28.4	12.08	**0.47**	1.14	**0.43**
GTT	V	18.5	22	11	19.32	1.04	0.88	1.76
TGG	W(Trp)	15.2	10.3	12.7	4.83	**0.32**	**0.47**	**0.38**
TAC	Y(Tyr)	12.1	14.6	15.5	14.49	1.20	0.99	0.94
TAT	Y	16.5	18.9	12.1	41.06	**2.49**	**2.17**	**3.39**
TAA	#	2	1	0.7	2.42	1.21	**2.42**	**3.45**
TAG	#	0.3	0.5	0.6	0.00	**0.00**	**0.00**	**0.00**
TGA	#	1.1	0.7	1.5	0.00	**0.00**	**0.00**	**0.00**

[#] = Stop codons. Underlined enhanced scores designate ratios higher than 2 or lower than 0.5.

DISCUSSION

Sialic acid is a generic term that represents a wide family of related nine-carbon sugar acids that occupy the terminal position within glycan molecules on the surfaces of many vertebrate cells. It has functions in diverse cellular processes such as intercellular adhesion and cell signaling. Pathogenic bacteria have evolved to use this molecule beneficially in at least two different ways: one is to provide their cell surfaces with sialic acid, conferring resistance to their host's innate immune response and an ability to interact specifically with different host-cell surfaces; another is to use it as a nutrient (Severi et al., 2007). Sialidase is a glycohydrolytic enzyme that can remove terminal sialic acid residues from various sialo-derivatives, such as glycoproteins, glycolipids, and oligosaccharides. These exoglycosidases are widely distributed in nature, including in viruses, protozoa, bacteria, fungi, mycoplasma, other microorganisms, and vertebrates (Saito and Robert, 1995).

S. agalactiae, also known as GBS, is a common cause of neonatal sepsis and meningitis in human beings, and also an important fish pathogen. The most extensively studied virulence factor of GBS is its sialic acid (Sia)-capped capsular polysaccharide (CPS). The Sia coated on the CPS of GBS can be specifically catalyzed and activated by sialidase, which limits the deposition of alternative complement pathway component C3b onto the cell surface. This inhibition results in decreasing phagocytosis by macrophages and neutrophils (Marques et al., 1992), thereby enhancing bacterial survival in the host. To our knowledge, the group of genes encoding sialidase comprises *neuA*, *neuB*, *neuC*, and *neuD*, and the *neuA* gene encodes CMP-NeuNAc synthetase, which catalyzes the activation of sialic acid with CMP (Haft et al., 1996). Thus, NeuA sialidase plays an important role in the pathogenesis of *S. agalactiae* infection, and there is an urgent need to study the *neuA* gene and its corresponding products for a better understanding of the pathogenic mechanism caused by *S. agalactiae.*

In the present paper, the *neuA* nucleotide sequence analysis indicated that it was 1242 bp in length, which was longer than other reported *neuA* nucleotide sequences (Ramaswamy et al., 2006). The protein derived from the *neuA* gene was 413 aa in length and the aa contents varied greatly. We identified the conserved domains of the NeuA protein containing a Glyco_tranf_GTA_type superfamily conserved domain and an SGNH_hydrolase superfamily conserved domain. The Glyco_tranf_GTA_type superfamily in the NeuA protein was characterized by a CMP-NeuAc_Synthase conserved domain, which can catalyze the activation of N-acetylneuraminic acid (Sia) and play an important role in the sialylation of the oligosaccharide component of glycoconjugates. Thus, the discovery of the CMP-NeuAc-Synthase conserved domain in the NeuA protein in this study is useful and provides important theoretical reference information for research on the NeuA protein in the immune protection and toxicity assessment to tilapia.

The inherent hydrophobicity of aa is an important factor for the three-dimensional conformation of the corresponding protein (Cao et al., 2010). The hydrophobicity prediction analysis showed that hydrophilic regions were more common than hydrophobic regions in the polypeptide, indicating that NeuA is a hydrophilic protein. Post-translational modifications are necessary for protein formation, and alter the protein characteristics through cleaving the protein by proteolysis or adding a modifying group to one or more aa (Mann and Jensen, 2003; Blom et al., 2004). Phosphorylation of the serine/threonine/tyrosine in the protein is the important post-translational modification in this case. Glycosylation is a complex process, involving the participation of 13 different types of monosaccharides and eight types of aa

residues (Spiro, 2002), and the shortage of some glycosylation sites may affect the correct folding of the protein (Shental-Bechor and Levy, 2008). The modification of the protein can often greatly affect protein function. We identified 28 potential phosphorylation sites and two potential N-linked glycosylation sites in the NeuA protein, which may be modified by the post-translational modifications to some extent and play an important role in the regulation of the biological function of the NeuA protein.

A signal peptide is a sequence that facilitates the secretion of a precursor protein through cell membranes. When the signal peptide has been removed from the precursor protein by signal peptidase, the mature protein is ready to fulfill its normal functions (Nielsen and Krogh, 1998). However, the NeuA polypeptide in this study does not contain a signal peptide, suggesting that the mature NeuA protein has 413 aa. The transmembrane region prediction showed that the NeuA protein also does not contain a transmembrane region. Moreover, recombinant protein solubility prediction indicated that the solubility of the NeuA recombinant protein was only 61.1% when *E. coli* was selected for induced expression**,** suggesting that the NeuA fusion protein might be mostly packaged in inclusion bodies. Thus, when *E. coli* is selected as the expression host to construct a recombinant plasmid, it is very important to add the appropriate fusion protein labels at both sides of the *neuA* gene, which will improve the recombinant protein solubility and aid its purification (Huang et al., 2011).

It is well known that B-cells are able to bind with antigens through their B-cell receptors (BCR) or secrete antibodies that possess this ability (Pier et al., 2004). The antigen-binding portion of antibodies is equivalent to the BCR, and the BCR-binding portion of an antigen is called the B-cell epitope. B-cell epitopes are frequently located on the surface of the antigen, and are highly flexible, which facilitates the binding between antigen and BCR (Xiang et al., 2010). Moreover, secondary structures have a close relationship with the distribution of B-cell epitopes. Owing to the regular structure imposed by hydrogen bonds, α-helix and β-sheet elements are frequently located in the interior of the protein to stabilize its structure. The β-turn regions (T) and random coil regions (C), however, are located on the protein's surface and are easily deformed, which contributes to BCR binding (Barlow et al., 1986; Apostolopoulos et al., 2002). In addition, the hydrophobic residues are often embedded in the protein interior while the hydrophilic residues are exposed on the surface. The regions with the largest surface charge and polarity in local regions consisting of hydrophilic residues are often thought to be B-cell epitopes. In the present study, we combined the direct prediction of the B-cell epitopes using the BCPREDS online server with indirect prediction using DNASTAR. Considering the above comprehensive analyses, we concluded that the B-cell epitopes were located at residue positions 3-22, 40-59, 125-144, and 155-174, or their neighboring regions in the aa sequence. This indicates that the NeuA protein possesses sufficient immunogenicity to function as a vaccine against *S. agalactiae* infection in tilapia.

In general, codon usage bias in genes remains at a certain level across species (Wang et al., 2012). However, there is little available information on codon usage bias of the *S. agalactiae neuA* gene. In the present study, we compared the codon preferences of the *neuA* gene with those of *E. coli*, yeast, and *H. sapiens* to check which would be the most suitable host for the optimal expression of the *neuA* gene. The result of codon usage bias analysis of the *neuA* gene suggested that the codon usage bias in synonymous codons was a little higher, with 45.465 of the ENC value of the *neuA* gene. A comparison of the codon preferences of the *neuA* gene with those of *E. coli*, yeast, and *H. sapiens* shows that the codon preference differs greatly in *E. coli* and *H. sapiens* genes, and the *neuA* gene may be more efficiently expressed

in yeast systems. This lays the foundation for research on prokaryotic expression of the *neuA* gene and immunogenicity analysis of the NeuA protein.

In conclusion, we successfully amplified the *neuA* gene of *S. agalactiae* 3-BY, isolated from diseased tilapia, by PCR with specific primers. We identified the gene by sequencing and obtained the GenBank accession No. KF447585. When we conducted multiple sequence alignment with 14 other NeuA aa sequences from different bacterial strains, our NeuA sequence showed great homology with the other *S. agalactiae* strains, but much less homology to other streptococci species. Based on the bioinformatic analyses and homology modeling to generate similar tertiary structures, we identified the NeuA protein as belonging to the CMP-N-acetylneuraminic acid synthetase family with a CMP-NeuAc_Synthase conserved domain, which relates to catalyzing the activation of N-acetylneuraminic acid (sialic acid). In addition, there are some B-cell epitopes in the NeuA protein, although it is predicted to reside in the cytoplasm, which suggests that it could be useful as a vaccine for *S. agalactiae* infection. We hope the results will lead to a better understanding of the relationship between the structure and function of NeuA and its role in the pathogenic mechanism caused by *S. agalactiae.* Furthermore, the results will be helpful in the study of the immunoprotection and effectiveness of NeuA as a candidate vaccine.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by grants from the Changjiang Scholars and Innovative Team Development Plans of the Ministry of Education, China (#IRT0848).

REFERENCES

Amal MNA and Zamri-Saad M (2011). Streptococcosis in Tilapia (*Oreochromis niloticus*): A Review. *Pertanika J. Trop. Agric. Sci.* 34: 195-206.

Apostolopoulos V, Yu M, Corper AL, Teyton L, et al. (2002). Crystal structure of a non-canonical low-affinity peptide complexed with MHC class I: a new approach for vaccine design. *J. Mol. Biol.* 318: 1293-1305.

Barlow DJ, Edwards MS and Thornton JM (1986). Continuous and discontinuous protein antigenic determinants. *Nature* 322: 747-748.

Blom N, Gammeltoft S and Brunak S (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 294: 1351-1362.

Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, et al. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4: 1633-1649.

Brodeur BR, Boyer M, Charlebois I, Hamel J, et al. (2000). Identification of group B streptococcal Sip protein, which elicits cross-protective immunity. *Infect. Immun.* 68: 5610-5618.

Burland TG (2000). DNASTAR's Lasergene sequence analysis software. *Methods Mol. Biol.* 132: 71-91.

Cao H, Wang D and Zhang Y (2010). Cloning and bioinformatic analyses of BmLITAF gene from *Bombyx mori*. *J. Zhejiang Sci-Tech Univ.* 149-154. Available at [http://en.cnki.com.cn/Article_en/CJFDTOTAL-ZJSG201001030.htm].

Cheng Q, Stafslien D, Purushothaman SS and Cleary P (2002). The group B streptococcal C5a peptidase is both a specific protease and an invasin. *Infect. Immun.* 70: 2408-2413.

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

El-Manzalawy Y, Dobbs D and Honavar V (2008). Predicting flexible length linear B-cell epitopes. *Comput. Syst. Bioinformatics Conf.* 7: 121-32.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, et al. (2003). ExPASy: The proteomics server for in-depth protein

knowledge and analysis. *Nucleic Acids Res.* 31: 3784-3788.
Gotea V, Veeramachaneni V and Makałowski W (2003). Mastering seeds for genomic size nucleotide BLAST searches. *Nucleic Acids Res.* 31: 6935-6941.
Haft RF, Wessels MR, Mebane MF, Conaty N, et al. (1996). Characterization of cpsF and its product CMP-N-acetylneuraminic acid synthetase, a group B streptococcal enzyme that can function in K1 capsular polysaccharide biosynthesis in Escherichia coli. *Mol. Microbiol.* 19: 555-563.
Hoshina T, Sano T and Morimoto Y (1958). A *Streptococcus* pathogenic to fish. *J. Tokyo Univ. Fish.* 44: 57-68.
Huang JL, Wang KY, Xiao D, Wang J, et al. (2011). Characterization of the sip gene in the strain *Streptococcus agalactiae* isolated from tilapia. *Oceanol. Limnol. Sinica* 42: 554-560. Available at [http://en.cnki.com.cn/Article_en/CJFDTOTAL-HYFZ201104015.htm].
Jones DT (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292: 195-202.
Kitao T (1993). Streptococcal infections. In: Bacterial diseases of fish (Inglis V, Roberts RJ and Bromage NR, eds.). Blackwell Press, Landon, 196-210.
Mann M and Jensen ON (2003). Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* 21: 255-261.
Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, et al. (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res.* 33(Database issue): D192-D196.
Marques MB, Kasper DL, Pangburn MK and Wessels MR (1992). Prevention of C3 deposition by capsular polysaccharide is a virulence mechanism of type III group B streptococci. *Infect. Immun.* 60: 3986-3993.
Möller S, Croning MD and Apweiler R (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17: 646-653.
Nakamura Y, Gojobori T and Ikemura T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 28: 292.
Nielsen H and Krogh A (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6: 122-130.
Payne NR, Kim YK and Ferrieri P (1987). Effect of differences in antibody and complement requirements on phagocytic uptake and intracellular killing of "c" protein-positive and -negative strains of type II group B streptococci. *Infect. Immun.* 55: 1243-1251.
Petersen TN, Brunak S, von Heijne G and Nielsen H (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8: 785-786.
Pier GB, Lyczak JB and Wetzler LM (2004). Immunology, infection, and immunity: Washington, DC: American Society for Microbiology Press.
Popov I, Nenov A, Petrov P and Vassilev D (2009). Detecting Prospective Mutation Regions by Codon Usage-A Bioinformatics Approach. *Biotechnol. Biotechnological Equip.* 23 (Suppl. 1): 655-657.
Ramaswamy SV, Ferrieri P, Madoff LC, Flores AE, et al. (2006). Identification of novel cps locus polymorphisms in nontypable group B Streptococcus. *J. Med. Microbiol.* 55: 775-783.
Robinson JA and Meyer FP (1966). Streptococcal fish pathogen. *J. Bacteriol.* 92: 512.
Saito M and Robert KY (1995). Biochemistry and function of sialidases. In: Biology of the sialic acids (Rosenberg A, Ed.). Springer, New York, pp. 261-313.
Schwede T, Kopp J, Guex N and Peitsch MC (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 31: 3381-3385.
Severi E, Hood DW and Thomas GH (2007). Sialic acid utilization by bacterial pathogens. *Microbiology* 153: 2817-2822.
Shental-Bechor D and Levy Y (2008). Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci. U. S. A.* 105: 8256-8261.
Smialowski P, Martin-Galiano AJ, Mikolajka A, Girschick T, et al. (2007). Protein solubility: sequence based prediction and experimental verification. *Bioinformatics* 23: 2536-2542.
Spiro RG (2002). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 12: 43R-56R.
Wang K, Lian H, Chen D, Huang J, et al. (2012). Cloning, identification and molecular characteristics analysis of p1 gene of *Yersinia ruckeri* isolated from channel catfish (*Ictalurus punctatus*). *Asian J. An. and Vet. Adv.* 7: 1067-1078.
Wessels MR (1997). Biology of streptococcal capsular polysaccharides. *Soc. Appl. Bacteriol. Symp. Ser.* 26: 20S-31S.
Wright F (1990). The 'effective number of codons' used in a gene. *Gene* 87: 23-29.
Xiang J, Cheng A, Wang M, Chang H, et al. (2010). Prediction of B Cell Epitopes and Overexpression of Truncated VP19c of Duck Enteritis Virus in *Escherichia coli* Proceedings of Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference: IEEE: 1-5.
Yu NY, Wagner JR, Laird MR, Melli G, et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26: 1608-1615.