# aJava

Asian Journal of Animal and Veterinary Advances

Academic
Journals Inc.

www.academicjournals.com

Asian Journal of Animal and Veterinary Advances 7 (11): 1067-1078, 2012
ISSN 1683-9919 / DOI: 10.3923/ajava.2012.1067.1078


# Cloning, Identification and Molecular Characteristics Analysis of *p1* Gene of *Yersinia ruckeri* isolated from Channel Catfish (*Ictalurus punctatus*)

[1,3]Kai-Yu Wang, [1,3]Hai Lian, [2,3]De-Fang Chen, [1,3]Jin-Lu Huang, [1,3]Jun Wang and [1,3]Qiao-Feng Mou

[1]Department of Basic Veterinary, College of Veterinary Medicine, Sichuan Agricultural University, Ya'an 625014, Sichuan, People's Republic of China

[2]Department of Aquaculture, College of Animal Science and Technology, Sichuan Agricultural University, Ya'an 625014, Sichuan, People's Republic of China

[3]Key Laboratory of Animal Disease and Human Health of Sichuan Province, Sichuan Agricultural University, Ya'an 625014, Sichuan, People's Republic of China

*Corresponding Author: Kai-Yu Wang, Department of Basic Veterinary, College of Veterinary Medicine, Sichuan Agricultural University, Ya'an 625014, Sichuan, People's Republic of China Tel: +86-0835-2885753 Fax: +86-0835-2885302*

## ABSTRACT

The *p1* gene of *Yersinia ruckeri* (*yrp1*) which was isolated from channel catfish was amplified by PCR with specific primers and inserted into pMD19-T vector. The positive recombinant plasmid was selected and sequenced. Molecular characteristics analysis of the *yrp1* gene and the protein which is encoded by it was performed. The results showed that the *yrp1* gene is 1434 bp in length with G+C content of 44.35%. The analysis of codon bias indicated that the codon usage frequency of the *yrp1* gene is distinctly different and it is preferable to perform in *E. coli* and yeast. The theoretical relative molecular mass and iso-electric point of the Yrp1 amino acid sequence are about 51.5 kDa and 4.48 separately. The polypeptide has some important sites related to post-translational modification, including 35 potential phosphorylation sites and 4 potential N-glycosylation sites. The polypeptide analyzed in this study contains a ZnMc superfamily conserved domains and does not contain a signal peptide, even though it is a secretory protein.

**Key words:** *Yersinia ruckeri*, *p1* gene, cloning, molecular analysis, catfish, secretory protein

## INTRODUCTION

Yersiniosis or Enteric Redmouth Disease (ERM) which is caused by *Yersinia ruckeri* (*Y. ruckeri*) is an acute, contagious and highly lethal disease in all ages of fish (Ross *et al*., 1966). Since the first report of ERM in rainbow trout in 1952 (Rucker, 1966), more outbreaks were reported in North America, Australia, South Africa and Europe and caused significant economic losses in the salmonid farming industry (Furones *et al*., 1993). In 1990, *Y. ruckeri* was first isolated from silver carp (*Hypophthalmichthys molitrix*) and bighead carp (*Aristichthys nobilis*) in China (Bo-Hai *et al*., 1991). Since then, ERM began to break out in China and the susceptible hosts expanded to some other aquatic animals such as channel catfish, carp and shrimp etc. according to the reports in recent decades (Xue-Feng *et al*., 1997; Fang-Ling *et al*., 2010; Bo-Hai *et al*., 1991).

In April 2008, a severe and highly fatal disease broke out in a channel catfish farm in Sichuan province, China and the cumulative mortality reached 85%. A gram-negative bacterium

(designated as FF003) was isolated and identified as *Y. ruckeri* on the basis of its morphological, biochemical and physiological properties and molecular identity through 16S rDNA analysis (GeneBank Accession Number: FJ908709). The challenge experiments followed by the Koch's postulates revealed that *Y. ruckeri* was the pathogen in this outbreak (Fang-Ling *et al*., 2010). However, more research is required about the pathogenesis of *Y. ruckeri*.

Extracellular products of *Y. ruckeri*, including lipases (Romalde and Toranzo, 1993), proteases (Secades and Guijarro, 1999) and haemolysins (Fernandez *et al*., 2007), reproduce some characteristic signs of ERM, such as haemorrhage in the mouth and the intestine, when injected into fish. An extracellular metalloprotease termed Yrp1 (*Yersinia ruckeri* protease 1) can digest a wide variety of extracellular matrix and muscle proteins (e.g., laminin, fibrinogen, gelatine, actin and myosin) and lead to membrane alterations and pores in the capillary vessels. This may result in the leaking of blood from these vessels and hence cause the typical haemorrhages especially around the mouth and intestine (Fernandez *et al*., 2003). Therefore, the Yrp1 protein seems to play an important role in the pathogenesis of *Y. ruckeri* infection. However, there is little information about the molecular characteristics of the Yrp1 protein or the gene encoding it.

The objective of this study was to report the cloning, identification and nucleotide sequence analysis of the *yrp1* gene and the molecular characteristics of the Yrp1 protein. These works might provide some insight for further research about the gene and the protein and laid the foundation for further study on the pathogenic mechanism caused by *Y. ruckeri*.

## MATERIALS AND METHODS

**Bacterial strain, plasmid, chemicals and kits:** *Y. ruckeri* strain FF003 was isolated from channel catfish farm outbreaks in Sichuan Province, China and preserved in the laboratory. *Escherichia coli* (*E. coli*) DH5α competent cell was purchased from Tian Gen Biotech company, China. Cloning vector pMD19-T was purchased from TaKaRa company, China, as well as all chemicals and kits used in this study.

***Y. ruckeri* genomic DNA extraction:** *Y. ruckeri* strain FF003 was routinely cultured on nutrient broth at 28°C for 24 h. Then the genomic DNA of *Y. ruckeri* was extracted by using Bacterial Genomic DNA Extraction Kit according to the manufacturer's instruction.

**PCR amplification of the *yrp1* gene:** The coding region of the *yrp1* gene was amplified with one pair of primers. Forward primer (P1) 5'-GGATCCATGAAAGCAAGTAGTAATAAA-3' and the reverse primer (P2) 5'-AAGCTTCGAATACATCCAAACAATA-3', containing the *Bam*H I and *Hind* III restriction sites (underlined), respectively. The two primers were synthesized in TaKaRa company and PCR was carried out in a 25 μL reaction mixture containing 0.5 μL of each primer (20 pmol each), 1.0 μL DNA template (5 ng), 12.5 μL MasterMix and 10.5 μL water. The PCR conditions were: 95°C for 5 min; 30 cycles of 94°C for 1 min, 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 electrophoresis and stained with gold view.

**Cloning and sequencing of the *yrp1* gene:** The PCR products were purified by using Agarose Gel DNA Extraction Kit according to the manufacturer's instruction. The purified PCR products were cloned into pMD19-T vector, followed by transformation into *E. coli* DH5α competent cell.

Then the positive recombinant clone was selected by the Amp/IPTG/X-Gal agar plate. Recombinant plasmid was identified by bacterial colony PCR with aforementioned conditions, digested with restriction enzymes *Bam*H I and *Hind* III and fractionated in 1% agarose gels. DNA sequencing was also conducted in TaKaRa company.

**Nucleotide sequence analysis of the *yrp1* gene:** Alignment and base composition analysis of nucleotide sequence of the *yrp1* gene were performed with BLAST N software (Gotea *et al*., 2003) and DNAStar version 7.0, respectively.

**Codon usage bias analysis of the *yrp1* gene:** The Effective Number of Codons (ENC) value of the *yrp1* gene was computed with the European Molecular Biology Open Software Suite (EMBOSS) CHIPS online service program (Comeron and Aguade, 1998), followed by calculation of the codon usage bias of the *yrp1* gene with the CUSP program of EMBOSS (Popov *et al*., 2009). To examine whether different species follow with the same codon usage rule, the comparison of codon usage bias between the *yrp1* gene and *E. coli*, yeast and *Homo sapiens* (*H. sapiens*) were conducted. The database of the codon usage in *E. coli*, yeast and *H. sapiens* is available (Nakamura *et al*., 2000).

**Molecular characteristics analysis of the Yrp1 protein:** Amino acid composition, relative molecular mass and iso-electric point of the deduced Yrp1 amino acid sequence were calculated by using ProtParam software in ExPASy online system (Gasteiger *et al*., 2003). Conserved domains of the Yrp1 protein was identified by NCBI Conserved Domains software (Marchler-Bauer *et al*., 2005). Hydrophobicity analysis was performed in Bioedit version 7.0 software, followed by solubility prediction (Smialowski *et al*., 2007). Phosphorylation sites and N-glycosylation sites were predicted in NetPhos version 2.0 (Blom *et al*., 1999) and NetNGlyc version 1.0 online program, respectively. Transmembrane region prediction was carried out by using TMHMM online program (Moller *et al*., 2001), followed by searching signal peptide in SignalP 4.0 Server online system (Petersen *et al*., 2011).

**Secondary structure and tertiary structure prediction:** Logging in the web server Psipred (Jones, 1999), alpha helix, extended strand and random coil of the Yrp1 protein could be obtained. The presumption of 3D structure of the Yrp1 protein was performed through SWISS-MODEL (Schwede *et al*., 2003) online program.

## RESULTS AND DISCUSSION

**Amplification, cloning and sequencing of the *yrp1* gene:** To isolate the *yrp1* gene, PCR was conducted on DNA from *Y. ruckeri* genome, by using primers P1 and P2 which were specific to the *yrp1* gene. A band which was about 1400 bp was observed upon electrophoresis of the PCR products on agarose gel (Fig. 1). The approximate 1400 bp PCR product was purified and cloned into pMD19-T, followed by identification through bacterial colony PCR and digestion with restriction enzymes *Bam*H I and *Hind* III (Fig. 2), thus the positive recombinant plasmid was constructed, designated as pMD19-T-*yrp1*. After cloning and sequencing, the identical nucleotide sequence was analyzed by BLAST N software. Sequence alignment indicated that the DNA sequence obtained by PCR displayed similarity of 100% to the *p1* gene of *Y. ruckeri* ATCC standard

strain (GeneBank Accession Number: AJ 318052) and lower than 80% compared with other nucleotide sequences, suggesting that the *yrp1* gene is very conservative. Sequence analysis indicated that the nucleotide sequence of the *yrp1* gene which is 1434 bp in length has base composition of 392 adenine (27.34%), 286 cytosine (19.94%), 350 guanine (24.41%) and 406 thymine (28.31%) (Table 1) and the G+C content of 44.35%.

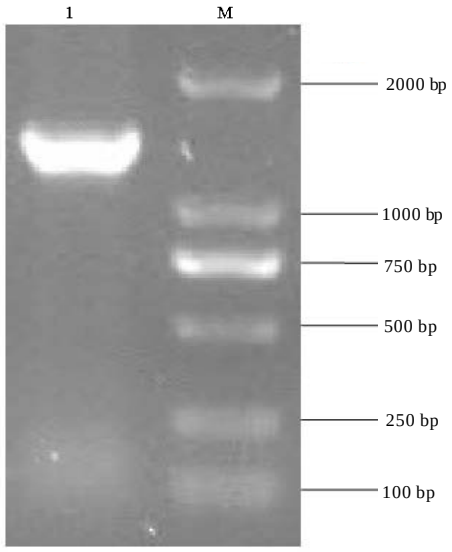


Fig. 1: PCR amplification result of the *yrp1* gene, M: DNA marker (DL2000), Lane 1: PCR product of the *yrp1* gene

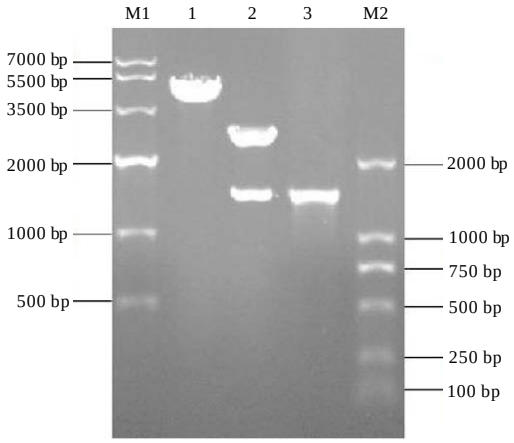


Fig. 2: Identification of the positive recombination plasmid through bacterial colony PCR and digestion with restriction enzymes *Bam*H I and *Hin*d III, 1, DNA marker (DL7000), Lane 1: Digestion of the recombinant plasmid with *Bam*H I, Lane 2: Digestion of the recombinant plasmid with *Bam*H I and *Hin*d III, Lane 3: Product of the bacterial colony PCR, M 2: DNA Marker (DL2000)

**Codon usage bias analysis of the *yrp1* gene:** Generally, the ENC is the effective number of codons which is used to quantify how far the codon usage of a gene departs from equal usage of synonymous codons and 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 between 20 (when only 1 codon is used for per amino acid) and 61 (when all synonyms are used with equal frequency). The ENC value of the *yrp1* gene calculated by using CHIPS program is 47.250, suggesting that the codon usage bias of it is a little higher. Therefore it was essential to check the codon usage in CUSP program. The results indicated that the *yrp1* gene does not contain codons TGT, CAC, AGG, CGA and CGG and the stop codon only appears with TGA. A high level of diversity in codon usage bias existed in the *yrp1* gene, especially preferring to use GAT, GAA, TTT, GGT, ATT, AAA, TTA, CCT, CGT and ACC for coding the Asp, Glu, Phe, Gly, Ile, Lys, Leu, Pro, Arg and Thr amino acid separately (Table 2).

Table 1: Base composition of the *yrp1* gene

| Base | No. | Percentage |
|---|---|---|
| Adenine | 392 | 27.34 |
| Cytosine | 286 | 19.94 |
| Guanine | 350 | 24.41 |
| Thymine | 406 | 28.31 |

Table 2: The result of codon bias analysis of the *yrp1* gene

| Codon | AA | Fraction[a] | Freq.[b] | No. | Codon | AA | Fraction[a] | Freq.[b] | No. |
|---|---|---|---|---|---|---|---|---|---|
| GCA | A(Alanine) | 0.225 | 18.828 | 9 | CCA | P(Proline) | 0.125 | 4.184 | 2 |
| **GCC** | A | 0.325 | 27.197 | 13 | *CCC* | P | 0.062 | 2.092 | 1 |
| GCG | A | 0.275 | 23.013 | 11 | **CCG** | P | 0.375 | 12.552 | 6 |
| *GCT* | A | 0.175 | 14.544 | 7 | **CCT** | P | 0.438 | 14.644 | 7 |
| **TGC** | C(Cysteine) | 1.000 | 2.092 | 1 | CAA | Q(Glutamine) | 0.500 | 20.921 | 10 |
| *TGT* | C | 0.000 | 0.000 | 0 | CAG | Q | 0.500 | 20.921 | 10 |
| *GAC* | D(Aspartate) | 0.205 | 16.736 | 8 | AGA | R(Arginine) | 0.100 | 2.092 | 1 |
| **GAT** | D | 0.795 | 64.854 | 31 | *AGG* | R | 0.000 | 0.000 | 0 |
| **GAA** | E(Glutamate) | 0.750 | 25.105 | 12 | *CGA* | R | 0.000 | 0.000 | 0 |
| *GAG* | E | 0.250 | 8.368 | 4 | CGC | R | 0.200 | 4.184 | 2 |
| *TTC* | F(Phenylalanine) | 0.333 | 16.736 | 8 | *CGG* | R | 0.000 | 0.000 | 0 |
| **TTT** | F | 0.667 | 33.473 | 16 | **CGT** | R | 0.700 | 14.644 | 7 |
| *GGA* | G(Glycine) | 0.038 | 4.184 | 2 | AGC | S(Serine) | 0.200 | 18.828 | 9 |
| GGC | G | 0.269 | 29.289 | 14 | **AGT** | S | 0.289 | 27.197 | 13 |
| GGG | G | 0.173 | 18.828 | 9 | TCA | S | 0.156 | 14.644 | 7 |
| **GGT** | G | 0.519 | 56.485 | 27 | TCC | S | 0.089 | 8.368 | 4 |
| *CAC* | H(Histidine) | 0.222 | 4.184 | 0 | *TCG* | S | 0.067 | 6.276 | 3 |
| **CAT** | H | 0.778 | 14.644 | 2 | TCT | S | 0.200 | 18.828 | 9 |
| *ATA* | I(Isoleucine) | 0.188 | 12.552 | 6 | ACA | T(Threonine) | 0.207 | 12.552 | 6 |
| *ATC* | I | 0.188 | 12.552 | 6 | **ACC** | T | 0.448 | 27.197 | 13 |
| **ATT** | I | 0.625 | 41.841 | 20 | *ACG* | T | 0.172 | 10.460 | 5 |
| **AAA** | K(Lysine) | 0.722 | 27.197 | 13 | *ACT* | T | 0.172 | 10.460 | 5 |
| *AAG* | K | 0.278 | 10.460 | 5 | *GTA* | V(Valine) | 0.125 | 6.276 | 3 |
| *CTA* | L(Leucine) | 0.037 | 2.092 | 1 | GTC | V | 0.333 | 16.736 | 8 |

Table 2: Continue

| Codon | AA | Fraction[a] | Freq.[b] | No. | Codon | AA | Fraction[a] | Freq.[b] | No. |
|---|---|---|---|---|---|---|---|---|---|
| *CTC* | L | 0.037 | 2.092 | 1 | GTG | V | 0.167 | 8.368 | 4 |
| CTG | L | 0.222 | 12.552 | 6 | **GTT** | V | 0.375 | 18.828 | 9 |
| *CTT* | L | 0.037 | 2.092 | 1 | **TGG** | W(Tryptophane) | 1.000 | 16.736 | 8 |
| **TTA** | L | 0.407 | 23.013 | 11 | *TAC* | Y(Tyrosine) | 0.391 | 18.828 | 9 |
| TTG | L | 0.259 | 14.644 | 7 | **TAT** | Y | 0.609 | 29.289 | 14 |
| **ATG** | M(Methionine) | 1.000 | 14.644 | 7 | *TAA* | Stop codon | 0.000 | 0.000 | 0 |
| AAC | N(Asparagine) | 0.432 | 33.473 | 16 | *TAG* | Stop codon | 0.000 | 0.000 | 0 |
| AAT | N | 0.568 | 43.933 | 21 | **TGA** | Stop codon | 1.000 | 2.092 | 1 |

[a]Proportion of all synonymous codons encoding the same amino acid, [b]The frequency of each codon that appears in the coding sequence of the individual gene is 1/1,000, Bold codous: High frequency in coding the amino acid, Italic codous: Low frequency in coding the amino acid

**Comparison of codon usage between *yrp1* and *E. coli*, yeast and *H. sapien*:** Generally, the codon usage bias in genes remains at a certain level across species. Thus, the codon preferences of the *yrp1* gene were compared with those of *E. coli*, yeast and *H. sapiens* to check which will be the suitable host for the optimal expression of the *yrp1* gene. The results showed there were 24 codons whose ratio was higher than 2 or lower than 0.5 between *yrp1*-to-*E. coli*, as well as *yrp1*-to-yeast, but up to 28 codons between *yrp1*-to-*H. sapiens*, suggesting that codon usage of the *yrp1* gene more closely resembles that of *E. coli* and yeast genes than that of *H. sapiens* genes (Table 3). Thus, it can be speculated that the *yrp1* gene may be more efficiently expressed in the *E. coli* or yeast systems which will lay the foundation for further research about the pathogenesis of *Y. ruckeri*, such as prokaryotic expression of the *yrp1* gene followed by studies on pathogenicity and immunogenicity of the Yrp1 protein.

**Molecular characteristics analysis of the Yrp1 protein:** The *yrp1* gene is expected to encode a protein comprising 477 amino acids with a molecular formula of $C_{2284}H_{3422}N_{608}O_{742}S_8$, theoretical isoelectric point of 4.48 and putative relative molecular mass of about 51.5 kDa which is in agreement with that observed for the purified metalloprotease (Secades and Guijarro, 1999). The protein encoded by the *yrp1* gene comprises 230 hydrophobic amino acids, 247 hydrophilic amino acids, 37 basic amino acids and 55 acidic amino acids, accounting for 48.2, 51.8, 7.8 and 11.5% separately. The polypeptide with higher content of Gly (10.9%) and Ser (9.4%) and lower content of Trp (1.7%), Met (1.5%) and Cys (0.2%) does not contain Pyl and Sec (Table 4).

Furthermore, by using different software and online web server, more information about the Yrp1 protein were gained as follows: Firstly, the Yrp1 protein contains a ZnMc superfamily conserved domains (between amino acid residue 70th and 255th) which is related to its virulence. Secondly, 3 potential N-linked glycosylation sites and 35 potential phosphorylation sites (including 22 Ser phosphorylation sites, 5 Thr phosphorylation sites and 8 Tyr phosphorylation sites) were also identified. Thirdly, the result of hydrophobicity prediction analysis revealed that the hydrophilic regions of the polypeptide are more than the hydrophobic regions, but recombinant protein solubility prediction indicated that the solubility of the recombinant protein is only 35.9% when *E. coli* was selected for induced expression, suggesting that the fusion protein might be mostly packaged into inclusion bodies. Fourthly, the Yrp1 protein does not contain transmembrane region and all the amino acids of it are located in the extracellular. However, the polypeptide analyzed in

Table 3: Comparison of codon bias between the *yrp1* gene and *E. coli*, yeast and *H. sapiens*

| | | | | | | Codon usage frequency | | |
|---|---|---|---|---|---|---|---|---|
| Codon | Amino acid | *E. coli* (1/1000) | Yeast (1/1000) | *H. sapiens* (1/1000) | *yrp1* (1/1000) | *yrp1*/*E. coli* | *yrp1*/Yeast | *yrp1*/*H. sapiens* |
| GCA | A(Alanine) | 20.6 | 16.1 | 16.1 | 18.828 | 0.914 | 1.169 | 1.169 |
| GCC | A | 25.5 | 12.5 | 28.4 | 27.197 | 1.067 | 2.176 | 0.958 |
| GCG | A | 31.7 | 6.1 | 7.5 | 23.013 | 0.726 | 3.773 | 3.068 |
| GCT | A | 15.6 | 21.1 | 18.6 | 14.644 | 0.939 | 0.694 | 0.787 |
| TGC | C(Cysteine) | 6.9 | 4.7 | 12.2 | 2.092 | 0.303 | 0.445 | 0.171 |
| TGT | C | 5.5 | 8.0 | 10.0 | 0.000 | 0.000 | *0.000* | 0.000 |
| GAC | D(Aspartate) | 18.6 | 20.2 | 25.6 | 16.736 | 0.900 | 0.829 | 0.654 |
| GAT | D | 32.1 | 37.8 | 21.9 | 64.854 | 2.020 | 1.716 | 2.961 |
| GAA | E(Glutamate) | 38.2 | 48.5 | 29.0 | 25.105 | 0.657 | 0.518 | 0.866 |
| GAG | E | 17.7 | 19.1 | 39.9 | 8.368 | 0.473 | 0.438 | 0.21 |
| TTC | F(Phenylalanine) | 16.9 | 18.2 | 20.6 | 16.736 | 0.990 | 0.92 | 0.812 |
| TTT | F | 23.2 | 26.1 | 17.1 | 33.473 | 1.443 | 1.282 | 1.957 |
| GGA | G(Glycine) | 9.0 | 10.9 | 16.4 | 4.184 | 0.465 | 0.384 | 0.255 |
| GGC | G | 27.9 | 9.7 | 22.5 | 29.289 | 1.05 | 3.019 | 1.302 |
| GGG | G | 11.3 | 6.0 | 16.3 | 18.828 | 1.666 | 3.138 | 1.155 |
| GGT | G | 24.4 | 24.0 | 10.8 | 56.485 | 2.315 | 2.354 | 5.23 |
| CAC | H(Histidine) | 9.8 | 7.7 | 15.0 | 4.184 | 0.427 | 0.543 | 0.279 |
| CAT | H | 13.6 | 13.7 | 10.5 | 14.644 | 1.077 | 1.069 | 1.395 |
| ATA | I(Isoleucine) | 5.4 | 17.8 | 7.7 | 12.552 | 2.324 | 0.705 | 1.63 |
| ATC | I | 24.2 | 17.0 | 21.6 | 12.552 | 0.519 | 0.738 | 0.581 |
| ATT | I | 29.8 | 30.4 | 16.1 | 41.841 | 1.404 | 1.376 | 2.599 |
| AAA | K(Lysine) | 33.2 | 42.2 | 24.1 | 27.197 | 0.819 | 0.644 | 1.129 |
| AAG | K | 10.7 | 30.7 | 32.2 | 10.46 | 0.978 | 0.341 | 0.325 |
| CTA | L(Leucine) | 4.0 | 13.3 | 7.8 | 2.092 | 0.523 | 0.157 | 0.268 |
| CTC | L | 11.0 | 5.4 | 19.8 | 2.092 | 0.190 | 0.387 | 0.106 |
| CTG | L | 50.9 | 10.4 | 39.8 | 12.552 | 0.247 | 1.207 | 0.315 |
| CTT | L | 11.7 | 12.1 | 13.0 | 2.092 | 0.179 | 0.173 | 0.161 |
| TTA | L | 13.9 | 26.7 | 7.5 | 23.013 | 1.656 | 0.862 | 3.068 |
| TTG | L | 14.0 | 27.0 | 12.6 | 14.644 | 1.046 | 0.542 | 1.162 |
| ATG | M(Methionine) | 27.0 | 20.9 | 22.2 | 14.644 | 0.542 | 0.701 | 0.66 |
| AAC | N(Asparagine) | 21.4 | 24.9 | 19.5 | 33.473 | 1.564 | 1.344 | 1.717 |
| AAT | N | 18.6 | 36.3 | 16.7 | 43.933 | 2.362 | 1.21 | 2.631 |
| CCA | P(Proline) | 8.5 | 18.2 | 16.7 | 4.184 | 0.492 | 0.230 | 0.251 |
| CCC | P | 5.8 | 6.8 | 20.1 | 2.092 | 0.361 | 0.308 | 0.104 |
| CCG | P | 21.8 | 5.3 | 6.9 | 12.552 | 0.576 | 2.368 | 1.819 |
| CCT | P | 7.3 | 13.6 | 17.3 | 14.644 | 2.006 | 1.077 | 0.846 |
| CAA | Q(Glutamine) | 15.0 | 27.5 | 12.0 | 20.921 | 1.395 | 0.761 | 1.743 |
| CAG | Q | 29.5 | 12.1 | 34.1 | 20.921 | 0.709 | 1.729 | 0.614 |
| AGA | R(Arginine) | 2.9 | 21.3 | 11.5 | 2.092 | 0.721 | 0.098 | 0.182 |
| AGG | R | 1.9 | 9.2 | 11.4 | 0.000 | 0.000 | 0.000 | 0.000 |
| CGA | R | 3.9 | 3.0 | 6.3 | 0.000 | 0.000 | 0.000 | 0.000 |
| CGC | R | 21.0 | 2.6 | 10.7 | 4.184 | 0.199 | 1.609 | 0.391 |
| CGG | R | 6.3 | 1.7 | 11.6 | 0.000 | 0.000 | 0.000 | 0.000 |
| CGT | R | 20.3 | 6.5 | 4.6 | 14.644 | 0.721 | 2.253 | 3.183 |

Table 3: Continue

| Codon | Amino acid | *E. coli* (1/1000) | Yeast (1/1000) | *H. sapiens* (1/1000) | *yrp1* (1/1000) | Codon wage frequency | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | *yrp1*/*E. coli* | *yrp1*/Yeast | *yrp1*/*H. sapiens* |
| AGC | S(Serine) | 16.0 | 9.7 | 19.3 | 18.828 | 1.177 | 1.941 | 0.976 |
| AGT | S | 9.5 | 14.2 | 11.9 | 27.197 | 2.863 | 1.915 | 2.285 |
| TCA | S | 7.8 | 18.8 | 12.0 | 14.644 | 1.877 | 0.779 | 1.22 |
| TCC | S | 8.9 | 14.2 | 11.9 | 8.368 | 0.940 | 0.589 | 0.703 |
| TCG | S | 8.7 | 8.5 | 4.4 | 6.276 | 0.721 | 0.738 | 1.426 |
| TCT | S | 8.7 | 23.5 | 14.7 | 18.828 | 2.164 | 0.801 | 1.281 |
| ACA | T(Threonine) | 8.2 | 17.8 | 15.1 | 12.552 | 1.531 | 0.705 | 0.831 |
| ACC | T | 22.8 | 12.6 | 19.4 | 27.197 | 1.193 | 2.158 | 1.402 |
| ACG | T | 14.8 | 7.9 | 6.1 | 10.46 | 0.707 | 1.324 | 1.715 |
| ACT | T | 9.1 | 20.3 | 13.0 | 10.46 | 1.149 | 0.515 | 0.805 |
| GTA | V(Valine) | 11.1 | 11.8 | 7.2 | 6.276 | 0.565 | 0.532 | 0.872 |
| GTC | V | 15.1 | 11.6 | 14.6 | 16.736 | 1.108 | 1.443 | 1.146 |
| GTG | V | 25.5 | 10.6 | 28.4 | 8.368 | 0.328 | 0.789 | 0.295 |
| GTT | V | 18.5 | 22.0 | 11.0 | 18.828 | 1.018 | 0.856 | 1.712 |
| TGG | W(Tryptophan) | 15.2 | 10.3 | 12.7 | 16.736 | 1.101 | 1.625 | 1.318 |
| TAC | Y(Tyrosine) | 12.1 | 14.6 | 15.5 | 18.828 | 1.556 | 1.29 | 1.215 |
| TAT | Y | 16.5 | 18.9 | 12.1 | 29.289 | 1.775 | 1.55 | 2.421 |
| TAA | Stop codon | 2.0 | 1.0 | 0.7 | 0.000 | 0.000 | 0.000 | 0.000 |
| TAG | Stop codon | 0.3 | 0.5 | 0.6 | 0.000 | 0.000 | 0.000 | 0.000 |
| TGA | Stop codon | 1.1 | 0.7 | 1.5 | 1.000 | 0.909 | 1.429 | 0.667 |

Underlined values higher than 2 or lower than 0.5 codon preference differs greatly and vice versa

Table 4: Amino acid analysis of the *yrp1* gene deduced protein

| AA | No. | Percentage | AA | No. | Percentage |
|---|---|---|---|---|---|
| Alanine | 40 | 8.4 | Lencine | 27 | 5.7 |
| Arginine | 10 | 2.1 | Lysine | 18 | 3.8 |
| Asparagine | 37 | 7.8 | Methionine | 7 | 1.5 |
| Aspartate | 39 | 8.2 | Phenylalanine | 24 | 5.0 |
| Cysteine | 1 | 0.2 | Proline | 16 | 3.4 |
| Glutamine | 20 | 4.2 | Serine | 45 | 9.4 |
| Glutamate | 16 | 3.4 | Threonine | 29 | 6.1 |
| Glycine | 52 | 10.9 | Tryptophane | 8 | 1.7 |
| Histidine | 9 | 1.9 | Tyrosine | 23 | 4.8 |
| Isolencine | 32 | 6.7 | Valine | 24 | 5.0 |

this study does not contain a signal peptide. Actually, the protein is secreted by a type I Gram-negative bacterial ABC exporter protein secretion system composed of three genes, *yrpD*, *yrpE* and *yrpF* and a protease inhibitor *inh* (Fernandez *et al*., 2002). That is why the polypeptide without a signal peptide can be secreted outside.

**Secondary structure and tertiary structure prediction:** The result of secondary structure prediction (Fig. 3) of the Yrp1 protein indicated it consists of alpha-helix with 17.4%, extended strand with 29.8% and random coil with 52.6%. The tertiary structure prediction of the Yrp1 protein was shown in Fig. 4.

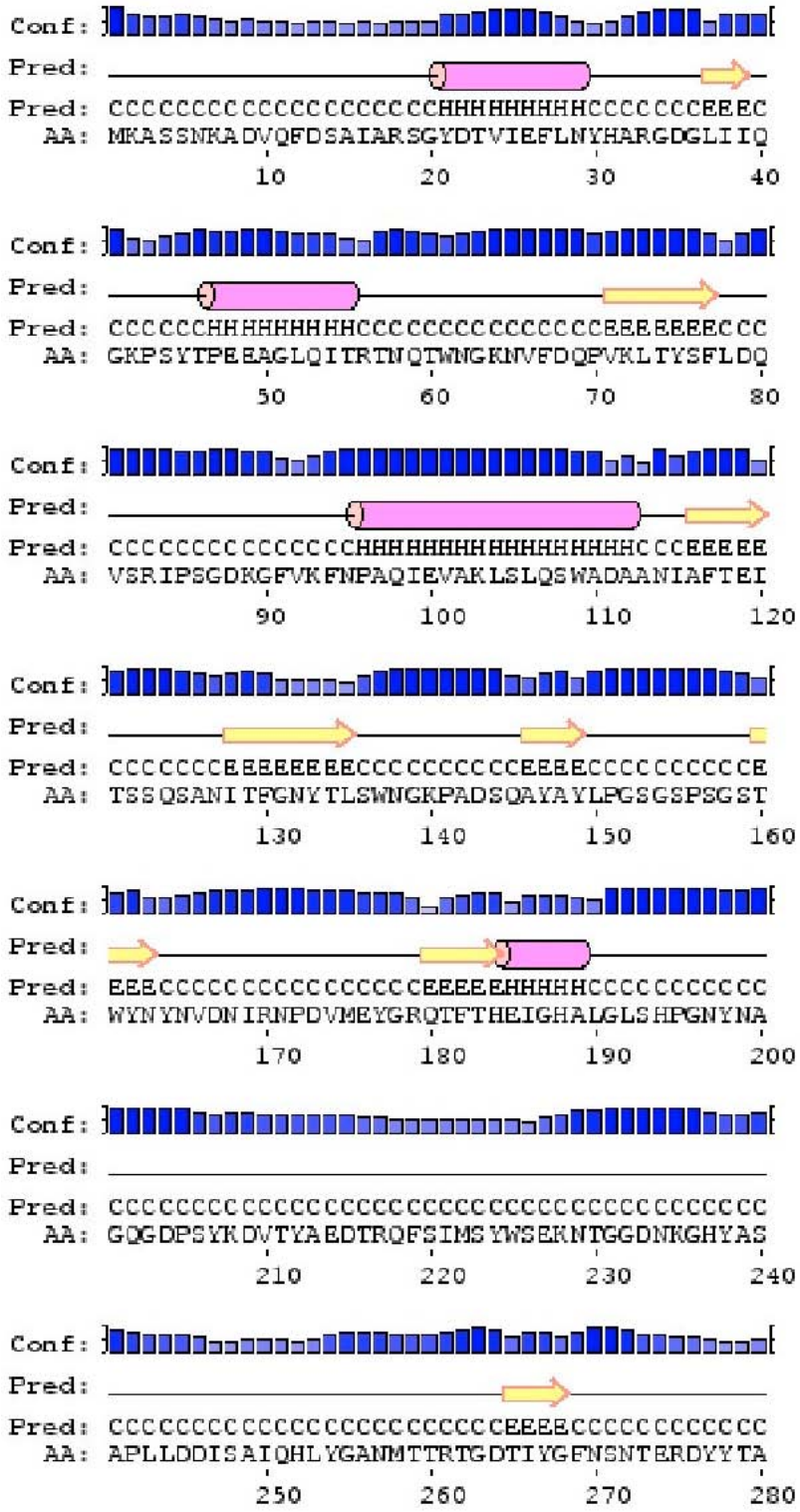


Fig. 3: Continue

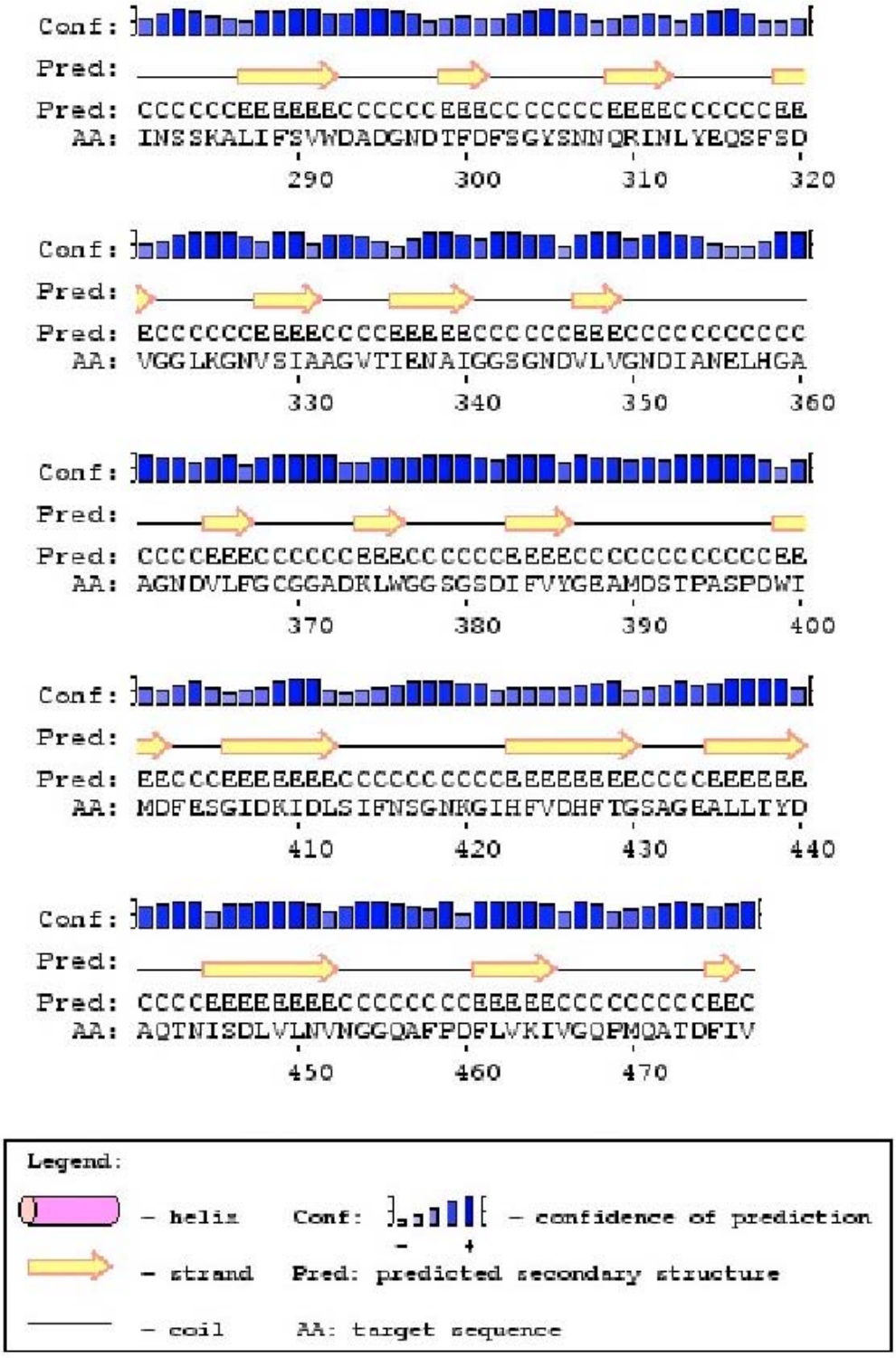


Fig. 3: Secondary structure prediction of the Yrp1 protein

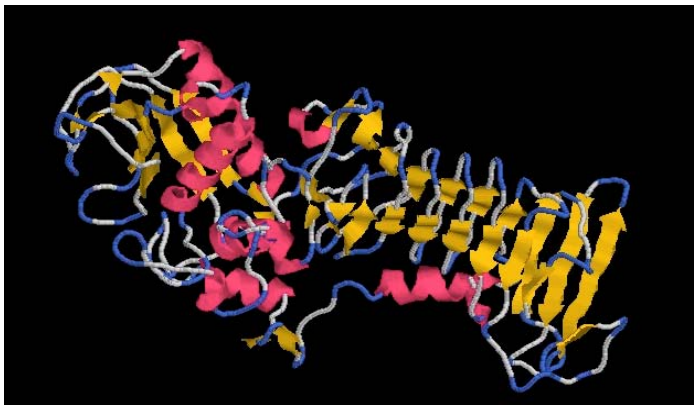

Fig. 4: Tertiary structure prediction of the Yrp1 protein, The alpha-helix is colored in red, the extended strand is colored in yellow and the random coil is colored in blue and grey

## CONCLUSION

The *p1* gene of *Yersinia ruckeri* isolated from channel catfish was successfully amplified by PCR with specific primers and identified by sequencing, followed by bioinformatics analysis with software and online web service. The results showed that the *yrp1* gene is 1434 bp in length with G+C content of 44.35%. The analysis of codon bias indicated that the codon usage frequency of the *yrp1* gene was distinctly different and it preferred to perform in yeast and *E. coli*. The theoretical relative molecular mass and iso-electric point of the deduced amino acid sequence are about 51.5 kDa and 4.48 separately, with a molecular formula of $C_{2284}H_{3422}N_{608}O_{742}S_8$. The polypeptide has some important sites related to post-translational modification, including 35 potential phosphorylation sites and 4 potential N-glycosylation sites. It contains a ZnMc superfamily conserved domains and does not contain a signal peptide, even though it is a secretory protein.

## ACKNOWLEDGMENT

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

## REFERENCES

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

Bo-Hai, X., Y. Zhan, C. Yan-Shen and W. Yu-Shen, 1991. Silver carp, bighead carp, a New infectious disease-*Yersinia ruckeri*, a new silver carp, bighead carp pathogen. Chin. Sci., Bull., 8: 620-622.

Comeron, J.M. and M. Aguade, 1998. An evaluation of measures of synonymous codon usage bias. J. Mol. Evol., 47: 268-274.

Fang-Ling, F., W. Kai-Yu, G. Yi, H. Xiao-Li and C. De-Fang, 2010. Isolation, identification and phylogenetic analysis of *Yersinia ruckeri* in channel catfish (*Ictalunes punctatus*). Oceanologia Et Limnologia Sincia, 41: 862-868.

Fernandez, L., J.R. Lopez, P. Secades, A. Menendez, I. Marquez and J.A. Guijarro, 2003. *In vitro* and *in vivo* studies of the Yrp1 protease from *Yersinia ruckeri* and its role in protective immunity against enteric red mouth disease of salmonids. Applied Environ. Microbiol., 69: 7328-7335.

Fernandez, L., M. Prieto and J.A. Guijarro, 2007. The iron- and temperature-regulated haemolysin YhlA is a virulence factor of *Yersinia ruckeri*. Microbiology, 153: 483-489.

Fernandez, L., P. Secades, J.R. Lopez, I. Marquez and J.A. Guijarro, 2002. Isolation and analysis of a protease gene with an ABC transport system in the fish pathogen *Yersinia ruckeri*: Insertional mutagenesis and involvement in virulence. Microbiology, 148: 2233-2243.

Furones, M., C. Rodgers and C. Munn, 1993. *Yersinia ruckeri*, the causal agent of enteric redmouth disease (ERM) in fish. Annua. Rev. Fish Dis., 3: 105-125.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel and A. Bairoch, 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucl. Acids Res., 31: 3784-3788.

Gotea, V., V. Veeramachaneni and W. Makalowski, 2003. Mastering seeds for genomic size nucleotide BLAST searches. Nucleic Acids Res., 31: 6935-6941.

Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol., 292: 195-202.

Marchler-Bauer, A., J.B. Anderson, P.F. Cherukuri, C. DeWeese-Scott and L.Y. Geer *et al.*, 2005. CDD: A conserved domain database for protein classification. Nucl. Acids Res., 33: D192-D196.

Moller, S, M.D. Croning and R. Apweiler, 2001. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics, 17: 646-653.

Nakamura, Y., T. Gojobori and T. Ikemura, 2000. Codon usage tabulated from international DNA sequence databases: Status for the year 2000. Nucleic Acids Res., 28: 292-292.

Petersen, T.N., S. Brunak, G. von Heijne and H. Nielsen, 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Natur. Methods, 8: 785-786.

Popov, I., A. Nenov, P. Petrov and D. Vassilev, 2009. Detecting prospective mutation regions by codon usage-a bioinformatics approach. Methods, 15: 424-424.

Romalde, J.L. and A.E. Toranzo, 1993. Pathological activities of *Yersinia ruckeri*, the enteric redmouth (ERM) bacterium. FEMS Microbiol. Lett., 112: 291-299.

Ross, A.J., R.R. Rucker and W.H. Ewing, 1966. Description of a bacterium associated with redmouth disease of rainbow trout (*Salmo gairdneri*). Can. J. Microbiol., 12: 763-770.

Rucker, R.R., 1966. Redmouth disease of rainbow trout (*Salmo gairdneri*). Bull. Off. Int. Epizoot, 65: 825-830.

Schwede, T., J. Kopp, N. Guex and M.C. Peitsch, 2003. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res., 31: 3381-3385.

Secades, P. and J. Guijarro, 1999. Purification and characterization of an extracellular protease from the fish pathogen *Yersinia ruckeri* and effect of culture conditions on production. Applied Environ. Microbiol., 65: 3969-3975.

Smialowski, P., A.J. Martin-Gahano, A. Mikolajka, T. Girschick, T.A. Holak and D. Frishman, 2007. Protein solubility: Sequence based prediction and experimental verification. Bioinformatics, 23: 2536-2542.

Wright, F., 1990. The effective number of codons used in a gene. Gene, 87: 23-29.

Xue-Feng, C., Y. Kai-Kang and M. Ing-Xian, 1997. Studies on a pathogen (*Yersinia ruckeri*) and its effects on hemolymph of *Penaeus orientalis*. J. Southwest Agric. Univ., 19: 458-461.