

Characterization of *Riemerella anatipestifer* CH-1 *gldJ* gene and GldJ protein

B. Yuan^{1,2}, A.C. Cheng^{1,2,3} and M.S. Wang^{1,2,3}

¹Institute of Preventive Veterinary Medicine, Sichuan Agricultural University, Wenjiang, Chengdu, Sichuan, China

²Avian Disease Research Center, College of Veterinary Medicine of Sichuan Agricultural University, Chengdu, Sichuan, China

³Key Laboratory of Animal Disease and Human Health of Sichuan Province, Sichuan Agricultural University, Chengdu, Sichuan, China

Corresponding authors: A.C. Cheng / M.S. Wang
E-mail: chenganchun@vip.163.com / mshwang@163.com

Genet. Mol. Res. 13 (4): 8329-8341 (2014)

Received July 31, 2013

Accepted December 9, 2013

Published October 20, 2014

DOI <http://dx.doi.org/10.4238/2014.October.20.9>

ABSTRACT. *Riemerella anatipestifer* (RA) CH-1, a highly virulent field strain, was isolated and identified by our laboratory. The *gldJ* gene was conserved in RA, and it had a typical TATA promoter region and AU-rich sequence within the 5' untranslated region. The GldJ protein was an outer-membrane lipoprotein with a signal peptide cleavage site between amino acids 20 and 21. GldJ was also a member of proteins involved in gliding motility. The RA GldJ protein had 16 phosphorylation sites and 4 N-glycosylation sites. These functional sites played an important role in the posttranslational modification of the GldJ protein. In addition, the GldJ protein of RA CH-1 had strong immunogenicity. Fifteen B-cell epitopes were identified in the GldJ protein, which might be a good biological material for the development of a subunit vaccine and diagnostic reagents. The GldJ protein also had the activity of formylglycine-generating sulfatase enzyme. The 3-dimensional structure models of GldJ were constructed based on

2 templates using the SwissModel automatic modeling mode in the SWISS-MODEL workplace. Here, we characterized the RA *gldJ* gene and GldJ protein to contribute to the functional annotation for the GldJ protein.

Key words: *Riemerella anatipestifer* CH-1; GldJ protein; Bioinformatic analysis

INTRODUCTION

Riemerella anatipestifer (RA) is a non-spore-forming, rod-shaped, and atrichous gram-negative bacterium belonging to the Flavobacteriaceae family (Subramaniam et al., 1997). It can cause a contagious disease in domestic ducks, geese, turkeys, and various other domestic and wild birds (Sarver et al., 2005; Rubbenstroth et al., 2009; Li et al., 2011). RA can cause large economic losses in the duck industry throughout the world. To date, more than 21 serovars have been identified (Pathanasophon et al., 2002). Unfortunately, no cross-protection has been observed with inactivated bacterins made from different serotypes of RA. Therefore, there are no effective inactivated bacterins for controlling the disease, and chemotherapy is currently a very important method in the treatment of ducks infected with RA (Chang et al., 2003). However, the application of antibiotics accelerates the emergence of drug-resistant strains of RA. Some resistant RA isolates have been reported whose aminoglycoside resistance genes, florfenicol resistance genes, and gene cassettes containing integrons also have been identified and investigated (Zhong et al., 2009; Chen et al., 2010; Sun et al., 2012).

Antibiotic therapy is not the best treatment in the long run. Some conserved immunogenic outer-membrane proteins have been identified in RA, including OmpA (Subramaniam et al., 2000), chaperonin GroEL (Han et al., 2012), and TonB-dependent outer-membrane receptor (Hu et al., 2012), which have good immunogenicity and are cross-immunogenic antigens in some different serotypes of RA. These proteins may be appropriate for developing novel vaccine candidates and serological diagnosis markers. China has the largest duck population over the world, and the economic losses of the duck industry that are caused by RA are serious in China.

Therefore, there is an urgent need to study the biological characteristics of these pathogens. The complete genome sequence of RA CH-1 was first determined by our laboratory (Wang et al., 2012). The structures and functions of hypothetical proteins of RA CH-1 need to be annotated. In most genome sequences, a large portion of unknown biochemical, biophysical, and functional hypothetical proteins exist. Structural genomics may be a viable approach to understand the biochemical function of hypothetical proteins. The functional annotations of hypothetical proteins are important in the process of illustrating the biological characteristics of RA. We can preliminarily annotate the function of the hypothetical proteins *in silico* and confirm the function of the hypothetical proteins by experimental evidence later. In this paper, we use bioinformatic tools to predict the function of the hypothetical protein GldJ of RA CH-1. Doing this, we can lay a foundation for the experimental study of the GldJ protein from RA.

MATERIAL AND METHODS

Genetic characteristics of the *gldJ* gene

The *gldJ* gene (GeneID: 13716554) of RA CH-1 was a hypothetical gene, which had not been described and annotated. The nucleotide sequence of the *gldJ* gene was submitted to the National Center for Biotechnology Information (NCBI) for open reading frame (ORF) finder and nucleotide basic local alignment search tool (BLASTN) analysis. Looking through the nucleotide sequence upstream of the initial codon of the *gldJ* gene, it was observed whether the TATA promoter region and Shine-Dalgarno (SD) sequence were present. The SD sequence, typically AGGAGG, is located approximately 10 nucleotides upstream of the initiator codon. In prokaryotes, the SD sequence in an mRNA is well known as the initiator element of translation.

Codon usage analyses of the *gldJ* gene

In order to investigate the pattern of codon usage in the *gldJ* gene of RA CH-1, the relative synonymous codon usage (RSCU) (Zhou et al., 2010) values of each codon were calculated. RSCU is the number observed of codon occurrences divided by the number expected when synonymous codons are used randomly and equally. If the RSCU value is close to 1.0, it indicates a lack of bias.

In order to know which system is suitable to express the RA *gldJ* gene efficiently, the ratio of *gldJ* codon usage frequency in RA to the complete genomes of *Escherichia coli* (*E. coli*), yeast, and human were calculated. The codon usage frequency (1/1000) of *gldJ* was calculated with the EMBOSS CUSP program. A ratio higher than 2 or lower than 0.5 indicates that the codon usage preference differs. The database of the codon usage in *E. coli*, yeast, and human is available at <http://www.kazusa.or.jp/codon>. Correlation analysis was used to identify the relationship between the codon usage bias and synonymous codon usage pattern. This analysis was conducted based on Spearman's rank correlation analysis. Statistical analysis was carried out using the statistical analysis software SPSS Version 19.0.

Prediction of phosphorylation and glycosylation sites of the GldJ protein

Phosphorylation sites of GldJ (GenBank accession No. YP_006724515.1) were predicted by NetPhosBac 1.0 Server. NetPhosBac 1.0 is a predictor of serine/threonine phosphorylation sites in bacterial proteins (Miller et al., 2009). N-glycosylation and O-glycosylation sites of the GldJ protein were predicted by NetNGlyc 1.0 Server and NetOGlyc 3.1 Server, respectively. These programs all use a threshold value of 0.5 to indicate potential phosphorylation and glycosylation sites.

Prediction of B-cell epitopes of the GldJ protein

Using the Protean program DNASTar, the alpha-helix and beta-sheet regions, hydrophilicity, flexibility, antigenic index, and surface probability of the GldJ protein were analyzed

using the Chou-Fasman method, Kyte-Doolittle method, Karplus-Schulz method, Jameson-Wolf method, and Emini method, respectively. B-cell epitopes have good hydrophilicity, good flexibility, high antigenic index, and good surface probability. However, B-cell epitopes are not located in the alpha-helix and beta-sheet regions (Liu et al., 2012).

BLAST-based search and conserved domains search

The GldJ protein was searched with protein BLAST (BLASTP) against the non-redundant protein sequences database. CD-Search is a bioinformatic tool that can search the conserved domains of proteins in the NCBI database. To find the conserved domains and annotate the functional units of the hypothetical protein, the hypothetical protein was searched against the conserved domain database.

Physicochemical characterization of the GldJ protein

The ProtParam procedure in ExPASy was used to analyze physical and chemical properties of the hypothetical protein. The molecular weight, theoretical isoelectric point, extinction coefficient, instability index, and half-life were predicted. The percent chance of insolubility when expressed in *E. coli* was predicted using the recombinant protein solubility prediction of the University of Oklahoma.

Structure and subcellular localization prediction of the GldJ protein

The secondary structure of the hypothetical protein was analyzed by using SOPMA. The 3-dimensional structure model of the hypothetical protein was constructed using the SwissModel automatic modeling mode in the SWISS-MODEL workplace. The subcellular localization of the hypothetical protein was predicted by the PSLpred program, which has an accuracy of 91.2% for predicting the subcellular localization of gram-negative bacterial proteins (Bhasin et al., 2005).

Transmembrane and signal peptide domains prediction of the GldJ protein

The transmembrane domains of the GldJ protein were predicted using the TM-HMM 2.0 program. The program adopts the hidden Markov model, and the predicting accuracy is high (Möller et al., 2001). The LipoP 1.0 Server is used to predict lipoproteins and signal peptides in gram-negative bacteria (Juncker et al., 2003). Here, we used the LipoP 1.0 Server to predict whether the GldJ protein was a lipoprotein with a signal peptide.

RESULTS

Genetic characteristics of the *gldJ* gene

The nucleotide sequence of the *gldJ* gene of RA CH-1 was a complete ORF. By

BLASTN analysis, we found that the *gldJ* ORF had 100% query coverage and 100% identities with the *gldJ* gene in RA ATCC 11845 and RA RA-GD. The TATA promoter region was found 5-8 nucleotides upstream of the initial codon of the *gldJ* gene. The mRNA of the *gldJ* gene did not possess the SD sequence. However, an AU-rich sequence, 5'-UUUAAAUUUU AAUAAA-3' (forward), was found in the mRNA 18-34 nucleotides upstream of the initiator codon.

Codon usage analyses of the *gldJ* gene

The RSCU values of codons in the *gldJ* gene are listed in Table 1, and 24 codons have high RSCU values. Thirty-one codons show an RA/*E. coli* ratio that is higher than 2 or lower than 0.50, along with 28 codons in the RA/yeast ratio and 38 codons in the RA/human ratio (Figure 1). Therefore, the codon usage of the *gldJ* gene is most closely similar to that of the yeast genome. In addition, the correlations of codon usage among the *gldJ* gene in RA and the complete genomes of *E. coli*, yeast, and human are highly significant (Table 2). The *gldJ* gene can be more efficiently expressed in the 3-expression system.

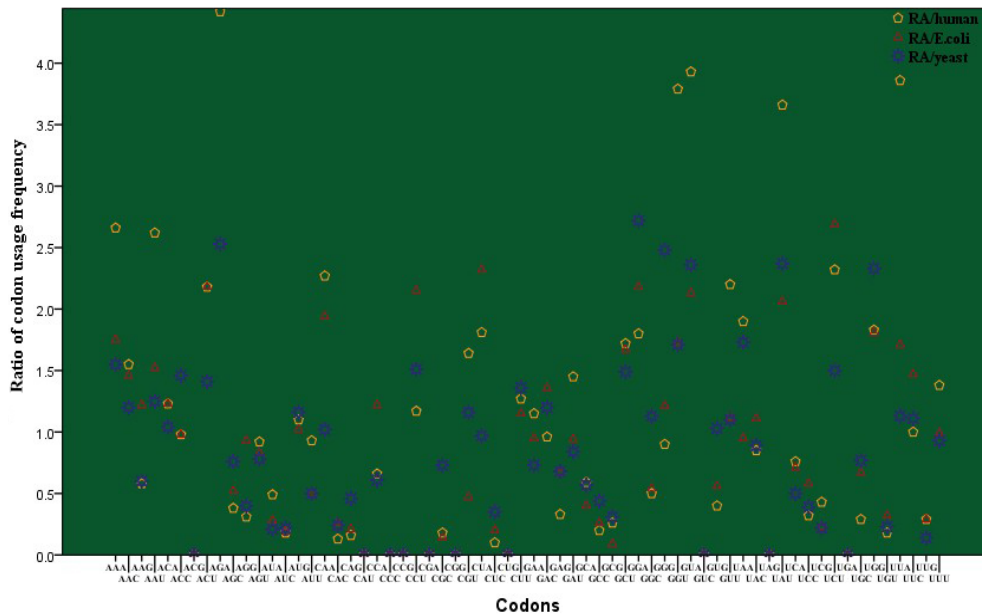


Figure 1. Ratio of *gldJ* codon usage frequency in *Riemerella anatipestifer* (RA) to the complete genome of *Escherichia coli*, yeast, and human.

Table 1. Synonymous codon usage in RA CH-1 *gldJ* gene.

AA ^a	Codon	RSCU ^b	AA ^a	Codon	RSCU ^b
Phe	UUU	1.08	Tyr	UAU ^c	1.55
	UUC	0.92		UAC	0.45
Leu	UUA ^c	2.74	His	CAC ^c	2.00
	UUG	0.34		Gln	CAA ^c
	CUU ^c	1.50	Asn	CAG	0.33
	CUC	0.17		AAU ^c	1.62
Ile	CUA ^c	1.20	Lys	AAC	0.38
	AUU ^c	2.00		AAA ^c	1.56
	AUC	0.50	Asp	AAG	0.44
	AUA	0.50		GAU ^c	1.35
Met	AUG	1.00	GAC	0.65	
Val	GUU ^c	1.53	Glu	GAA ^c	1.44
	GUA ^c	1.76		GAG	0.56
	GUG	0.71	Cys	UGU	0.67
Ser	UCU ^c	3.00		UGC ^c	1.33
	UCC	0.47	Arg	CGU	0.67
	UCA	0.79		CGC	0.17
	UCG	0.16	Ser	AGU	0.95
Pro	CCU ^c	2.59	Arg	AGC	0.63
	CCA ^c	1.41		AGA	4.83
Thr	ACU ^c	1.65	Gly	AGG	0.33
	ACC ^c	1.18		GGU ^c	1.39
	ACA ^c	1.18		GGC	0.46
Ala	GCU ^c	2.62	Gly	GGA ^c	1.23
	GCC	0.46		GGG	0.62
	GCA	0.77			
	GCG	0.15			

^aAA = abbreviation of amino acid. ^bRSCU value = fraction of the relative synonymous codon usage. ^cStrong bias on codons with A/U at the third codon position.

Table 2. Correlation analysis among *gldJ*, *Escherichia coli*, yeast, and human.

r	<i>gldJ</i>	<i>E. coli</i>	Yeast	Human
<i>gldJ</i>	1	0.477**	0.764**	0.324**
<i>E. coli</i>		1	0.567**	0.536**
Yeast			1	0.415**
Human				1

**Significant correlation at the 0.01 level (2-tailed).

Prediction of phosphorylation and glycosylation sites of the GldJ protein

The prediction of generic phosphorylation sites in the GldJ protein was performed using NetPhosBac 1.0 Server. Sixteen possible phosphorylation sites were identified in the GldJ protein, which included 14 serine residues (Figure 2; positions: 8, 15, 32, 123, 151, 169, 196, 244, 318, 446, 481, 484, 487, and 494) and 2 threonine phosphorylation sites (Figure 2 positions: 241 and 458). N-glycosylation and O-glycosylation sites of the GldJ protein were predicted using NetNGlyc 1.0 Server and NetOGlyc 3.1 Server, respectively. There were only 4 N-glycosylation sites (Figure 3; positions: 202, 396, 456, and 482) in the GldJ protein.

Prediction of B-cell epitopes of the GldJ protein

The alpha-helix and beta-sheet regions, hydrophilicity, flexibility, antigenic index, and surface probability of the GldJ protein are shown in Figure 4. The GldJ protein had good hydrophilicity and strong antigenicity. There were many flexible regions in the amino acid sequence of the GldJ protein, and few alpha-helix and beta-sheet regions were identified in the GldJ protein. The hydrophilicity and antigenicity index regions shown in Figure 1 revealed that the immunogenicity of GldJ should be very good. B-cell epitopes were screened by comprehensive comparison and might be located at amino acid residues 24-28, 41-46, 86-91, 122-128, 144-151, 181-208, 215-224, 297-302, 311-318, 325-339, 374-378, 385-393, 423-428, 504-514, and 530-534.

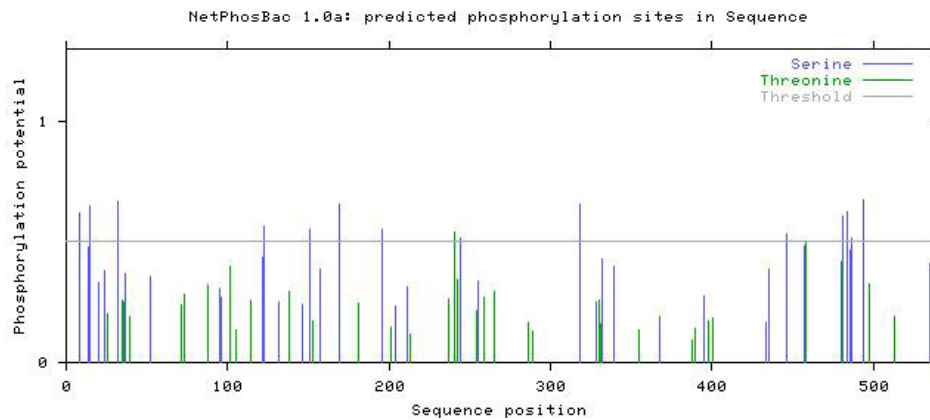


Figure 2. Phosphorylation sites of the GldJ protein were predicted by NetPhosBac 1.0.

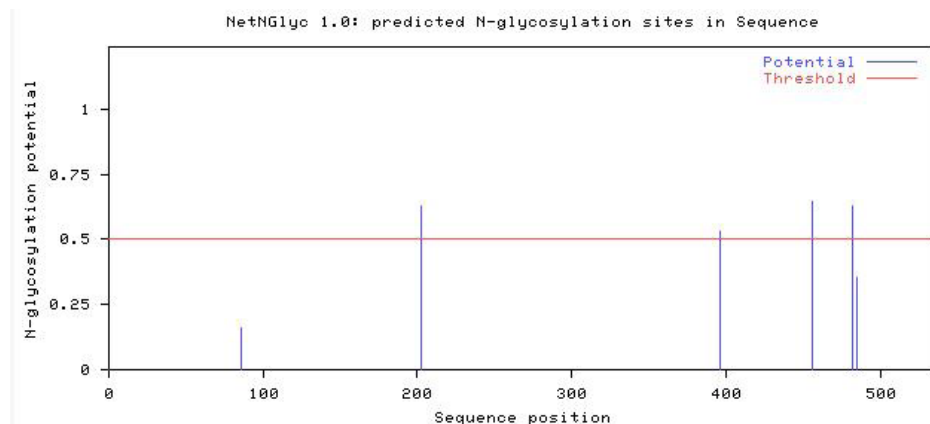


Figure 3. N-glycosylation sites of the GldJ protein were predicted by NetNGlyc 1.0.

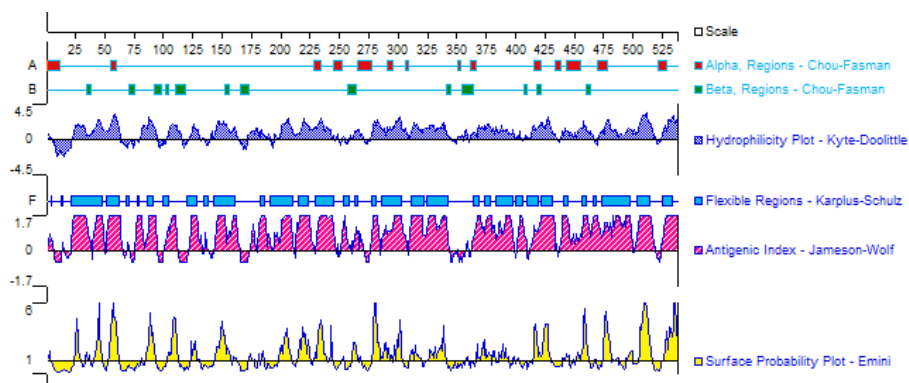


Figure 4. Alpha-helix and beta-sheet regions, hydrophilicity, flexibility, antigenic index, and surface probability of the GldJ protein were analyzed using the Protean program of DNASTar.

BLAST-based search and conserved domains search

The GldJ protein was searched with BLASTP against the non-redundant protein sequence database. The identities of the hypothetical protein with RA ATCC 11845, RA RA-GD, and RA RA-YM GldJ protein are 100, 100, and 99%, respectively. In addition, there are many homologous GldJ proteins in other bacteria that have high identity with the GldJ protein of RA CH-1 (identities >40%). The GldJ protein should be a conserved protein in RA. The GldJ protein belongs to the formylglycine-generating sulfatase enzyme (FGE-sulfatase) superfamily. The homologous proteins found in eukaryotes were required for posttranslational sulphatase modification (SUMF1) (Schlotawa et al., 2011). Moreover, the GldJ protein was involved in the gliding motility of bacteria (Braun and McBride, 2005).

Physicochemical characterization of the GldJ protein

The GldJ protein of RA CH-1 was composed of 537 amino acids, which included 205 hydrophobic amino acids (38.18%), 185 hydrophilic amino acids (34.45%), 82 basic amino acids (15.27%), and 65 acidic amino acids (12.10%). The molecular weight of the GldJ protein was 61.175 kDa, and the theoretical isoelectric point was 9.31. The estimated half-life of the GldJ protein was 30 h (mammalian reticulocytes, *in vitro*), more than 20 h (yeast, *in vivo*), and more than 10 h (*E. coli*, *in vivo*). Assuming all pairs of cysteine residues formed cystines, the extinction coefficient was 117,815 L·mol⁻¹·cm⁻¹ at 280 nm in water. The instability index (II) was computed to be 27.33 (II < 40). Therefore, the properties of the GldJ protein are stable (Guruprasad et al., 1990). The amino acid sequence had a 93.9% chance of insolubility when expressed in *E. coli*. When the *gldJ* gene was expressed in the *E. coli* system, the expressed GldJ protein would be insoluble.

Structure, transmembrane domains, signal peptide domain, and subcellular localization of the GldJ protein

The GldJ protein had 120 (22.35%) alpha helices (Hh), 108 (20.11%) extended

strands (Ee), 30 (5.59%) beta turns (Tt), and 279 (51.96%) random coils (Cc). The distributions of Hh, Ee, Tt, and Cc are shown in the Figure 5. The 3-dimensional structure models of the GldJ protein were constructed based on the crystal structure of human and *Streptomyces coelicolor* FGE (Figures 6 and 7, respectively). Prediction scores of cytoplasm, extracellular, inner-membrane, outer-membrane, and periplasmic location were -1.1192642, -1.6058245, -1.0932696, 0.92141513, and -0.85227888 using PSLpred, respectively. The prediction result of the TMHMM2.0 program revealed that there were no transmembrane domains in the GldJ protein. The prediction result of the LipoP 1.0 Server showed that GldJ was a lipoprotein with a signal peptide cleavage site between amino acids 20 and 21. Thus, the GldJ protein may be an outer-membrane lipoprotein.

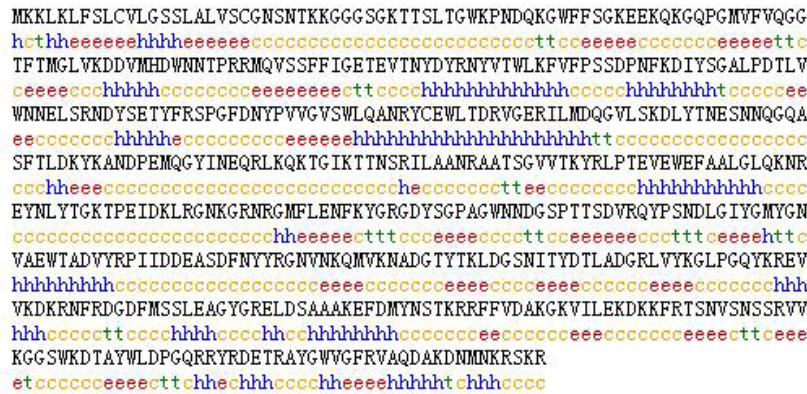


Figure 5. Secondary structure of the GldJ protein of RA was predicted by SOPMA. h = alpha helice; c = random coil; t = beta turn; e = extended strand.

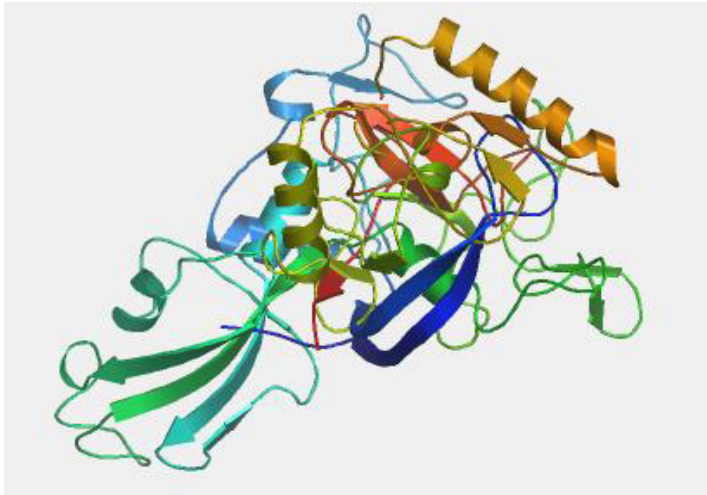


Figure 6. Three-dimensional structure mode of RA CH-1 GldJ based on the 2AII_X template of human formylglycine-generating enzyme (FGE). The 3-dimensional structure model was predicted by SWISS-MODEL, which is an automated comparative protein modeling server.

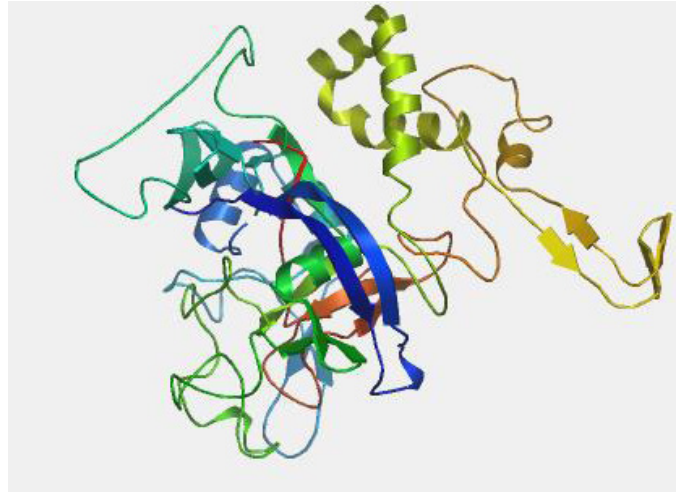


Figure 7. Three-dimensional structure model of RA CH-1 GldJ based on the 2Q17_A template of *Streptomyces coelicolor* FGE. The 3-dimensional structure model was predicted by SWISS-MODEL, which is an automated comparative protein modeling server.

DISCUSSION

Gliding motility allows some bacteria to glide over surfaces without the aid of flagella. It is an important biological characteristic of bacteria, which is related to biofilm formation and bacterial virulence (Mignot et al., 2007). The GldJ protein of *Flavobacterium johnsoniae* is responsible for gliding motility (Braun and McBride, 2005). The GldJ protein of RA is 43% identical to that of *F. johnsoniae*. Generally speaking, if two proteins are 25% or more identical, they may have a similar function. The result of the BLASTP analysis showed that RA CH-1 GldJ had high identities with GldJ proteins from other bacteria. GldJ should be conserved in some bacteria that need the GldJ protein to accomplish gliding motility. GldJ may also play an important role in the gliding motility of RA, and it is worth being investigated. Many bacterial GldJ amino acid sequences have been submitted to NCBI. However, the mechanism of gliding motility in many bacteria is not clear. The *gldJ* gene of RA had an analogous TATA promoter region and an AU-rich sequence within the 5' untranslated region. The TATA promoter region is the binding site of general transcription factors, and it is critical for transcription initiation and veracity (Yang et al., 2007). The AU-rich sequence can enhance and stabilize mRNA (Komarova et al., 2005). Moreover, the sequence can be bound by ribosomal protein S1 to instigate translation initiation in the case of mRNA (Komarova et al., 2005; Nakagawa et al., 2010).

Protein phosphorylation and glycosylation are major regulatory post-translational modifications in bacteria. In bacteria, protein phosphorylation is the most important post-translational modification on tyrosine, serine, and threonine residues. Protein phosphorylation can control bacterial virulence at the level of gene expression (Jers et al., 2008; Whitmore and Lamont, 2012), and there is also a definite relation between protein phosphorylation and bacterial pathogenicity (Ge and Shan, 2011). N-Glycosylation can introduce significant changes in the biological characteristics of a protein, including the localization, activity, solubility, and

cell-to-cell interactions (Dall'Olio et al., 2012). Additionally, the glycosylation of bacterial proteins may serve specific functions in infection and pathogenesis and interfere with inflammatory immune responses (Schmidt et al., 2003). The phosphorylation and N-glycosylation sites of the GldJ protein have been identified. However, the functional effects of these modification sites on the GldJ protein need to be further investigated.

B-cell epitopes can be categorized into 2 types: linear (continuous) epitopes and conformational (discontinuous) epitopes (Mahdavi et al., 2012). Although the majority of B-cell epitopes are conformational epitopes, the experimental determination of epitopes has primarily focused on the identification of linear B-cell epitopes (El-Manzalawy and Honavar, 2010). Therefore, the prediction of linear B-cell epitopes of the GldJ protein may help us develop a vaccine and diagnostic reagents and understand the pathogenesis and immunological action of RA (Tian et al., 2013).

The GldJ protein of RA CH-1 was likely to participate in the gliding motility of bacterial cells. The instability index (II) was computed to be 27.33 ($II < 40$), which revealed that the GldJ protein had a stable biological function. However, when expressed in *E. coli*, GldJ was insoluble. The biological activity of the GldJ protein expressed in *E. coli* must be influenced. The 3-dimensional structural model of the GldJ protein represents a stable folding conformation. Mutations in *gldA*, *gldB*, *gldD*, *gldF*, *gldG*, *gldH*, *gldI*, and *secDF* all resulted in decreased levels of GldJ protein (Braun and McBride, 2005; Nelson and McBride, 2006). Therefore, the function of GldJ and the interaction of gliding motility proteins will be better understood by studying the 3-dimensional structural model of the GldJ protein. Besides, it can pave the way for generating computer-generated molecular models for proteins whose crystal structures are not available. The prediction of PSLpred showed that GldJ was an outer-membrane protein. However, the prediction result of the TMHMM2.0 program revealed that there were no transmembrane domains in the GldJ protein. Nevertheless, the prediction of transmembrane domains is not sufficient to illustrate whether GldJ is an outer-membrane protein. GldA of *F. johnsoniae* has no obvious transmembrane domains, but cell fractionation followed by Western blot analysis demonstrated its presence in membrane fractions (Hunnicuttt et al., 2002). In addition, the GldJ protein of *F. johnsoniae* was shown to be an outer-membrane lipoprotein (Braun and McBride, 2005). Besides, the prediction result of LipoP 1.0 Server revealed that GldJ was a lipoprotein with a signal peptide cleavage site between amino acids 20 and 21. Thus, the GldJ protein of RA should be an outer-membrane lipoprotein.

The GldJ protein of RA had the activity of FGE-sulfatase, which was a member of the FGE-sulfatase superfamily. FGE, which belongs to the oxygen-dependent oxygenases family, is an important modifying factor for maturing cysteine-type sulfatases in eukaryotes (Cosma et al., 2003). To date, FGE is mainly investigated in eukaryotes. In prokaryotes, the research on FGE is very sparse. FGE can recognize and mature cysteine-type or serine-type sulfatases in bacteria (Marquardt et al., 2003). In summary, the GldJ protein not only have the enzyme activity of FGE but is also involved in the gliding motility of RA. To better understand the unique enzymatic mechanism and biological function of the GldJ protein, we need to pay more attention to studies of the GldJ protein.

ACKNOWLEDGMENTS

Research supported by the Special Fund for Agro-Scientific Research in the Public Interest (#201003012), National Science and Technology Support Program for Ag-

riculture (#2011BAD34B03), China Agricultural Research System (CARS-43-8), and Innovative Research Team Program in Education Department of Sichuan Province (#12TD005/#2013TD0015).

REFERENCES

- Bhasin M, Garg A and Raghava GP (2005). PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics* 21: 2522-2524.
- Braun TF and McBride MJ (2005). *Flavobacterium johnsoniae* GldJ is a lipoprotein that is required for gliding motility. *J. Bacteriol.* 187: 2628-2637.
- Chang CF, Lin WH, Yeh TM, Chiang TS, et al. (2003). Antimicrobial susceptibility of *Riemerella anatipestifer* isolated from ducks and the efficacy of ceftiofur treatment. *J. Vet. Diagn. Invest.* 15: 26-29.
- Chen YP, Tsao MY, Lee SH, Chou CH, et al. (2010). Prevalence and molecular characterization of chloramphenicol resistance in *Riemerella anatipestifer* isolated from ducks and geese in Taiwan. *Avian Pathol.* 39: 333-338.
- Cosma MP, Pepe S, Annunziata I, Newbold RF, et al. (2003). The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. *Cell* 113: 445-456.
- Dall'Olio GM, Laayouni H, Luisi P, Sikora M, et al. (2012). Distribution of events of positive selection and population differentiation in a metabolic pathway: the case of asparagine N-glycosylation. *BMC Evol. Biol.* 12: 98.
- El-Manzalawy Y and Honavar V (2010). Recent advances in B-cell epitope prediction methods. *Immunome. Res.* 6 (Suppl 2): S2.
- Ge R and Shan W (2011). Bacterial phosphoproteomic analysis reveals the correlation between protein phosphorylation and bacterial pathogenicity. *Genomics Proteomics Bioinformatics* 9: 119-127.
- Guruprasad K, Reddy BV and Pandit MW (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Eng.* 4: 155-161.
- Han X, Hu Q, Ding S, Chen W, et al. (2012). Identification and immunological characteristics of chaperonin GroEL in *Riemerella anatipestifer*. *Appl. Microbiol. Biotechnol.* 93: 1197-1205.
- Hu Q, Ding C, Tu J, Wang X, et al. (2012). Immunoproteomics analysis of whole cell bacterial proteins of *Riemerella anatipestifer*. *Vet. Microbiol.* 157: 428-438.
- Hunnicut DW, Kempf MJ and McBride MJ (2002). Mutations in *Flavobacterium johnsoniae* gldF and gldG disrupt gliding motility and interfere with membrane localization of GldA. *J. Bacteriol.* 184: 2370-2378.
- Jers C, Soufi B, Grangeasse C, Deutscher J, et al. (2008). Phosphoproteomics in bacteria: towards a systemic understanding of bacterial phosphorylation networks. *Expert. Rev. Proteomics* 5: 619-627.
- Juncker AS, Willenbrock H, Von Heijne G, Brunak S, et al. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12: 1652-1662.
- Komarova AV, Tchufistova LS, Dreyfus M and Boni IV (2005). AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *J. Bacteriol.* 187: 1344-1349.
- Li JX, Tang Y, Gao JY, Huang CH, et al. (2011). *Riemerella anatipestifer* infection in chickens. *Pak. Vet. J.* 31: 65-69.
- Liu H, Huang L, Luo J, Chen W, et al. (2012). Prediction and identification of B cell epitopes derived from EWS/FLI-1 fusion protein of Ewing's sarcoma. *Med. Oncol.* 29: 3421-3430.
- Mahdavi M, Mohabatkar H, Keyhanfar M, Dehkordi AJ, et al. (2012). Linear and conformational B cell epitope prediction of the HER 2 ECD-subdomain III by *in silico* methods. *Asian Pac. J. Cancer Prev.* 13: 3053-3059.
- Marquardt C, Fang Q, Will E, Peng J, et al. (2003). Posttranslational modification of serine to formylglycine in bacterial sulfatases. Recognition of the modification motif by the iron-sulfur protein AtsB. *J. Biol. Chem.* 278: 2212-2218.
- Mignot T, Shaevitz JW, Hartzell PL and Zusman DR (2007). Evidence that focal adhesion complexes power bacterial gliding motility. *Science* 315: 853-856.
- Miller ML, Soufi B, Jers C, Blom N, et al. (2009). NetPhosBac - a predictor for Ser/Thr phosphorylation sites in bacterial proteins. *Proteomics* 9: 116-125.
- Möller S, Croning MD and Apweiler R (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17: 646-653.
- Nakagawa S, Niimura Y, Miura K and Gojobori T (2010). Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 107: 6382-6387.
- Nelson SS and McBride MJ (2006). Mutations in *Flavobacterium johnsoniae* secDF result in defects in gliding motility and chitin utilization. *J. Bacteriol.* 188: 348-351.
- Pathanasophon P, Phuektes P, Tanticharoenyos T, Narongsak W, et al. (2002). A potential new serotype of *Riemerella*

- anatipestifer* isolated from ducks in Thailand. *Avian Pathol.* 31: 267-270.
- Rubbenstroth D, Ryll M, Behr KP and Rautenschlein S (2009). Pathogenesis of *Riemerella anatipestifer* in turkeys after experimental mono-infection via respiratory routes or dual infection together with the avian metapneumovirus. *Avian Pathol.* 38: 497-507.
- Sarver CF, Morishita TY and Nersessian B (2005). The effect of route of inoculation and challenge dosage on *Riemerella anatipestifer* infection in Pekin ducks (*Anas platyrhynchos*). *Avian Dis.* 49: 104-107.
- Schlotawa L, Ennemann EC, Radhakrishnan K, Schmidt B, et al. (2011). SUMF1 mutations affecting stability and activity of formylglycine generating enzyme predict clinical outcome in multiple sulfatase deficiency. *Eur. J. Hum. Genet.* 19: 253-261.
- Schmidt MA, Riley LW and Benz I (2003). Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol.* 11: 554-561.
- Subramaniam S, Chua KL, Tan HM, Loh H, et al. (1997). Phylogenetic position of *Riemerella anatipestifer* based on 16S rRNA gene sequences. *Int. J. Syst. Bacteriol.* 47: 562-565.
- Subramaniam S, Huang B, Loh H, Kwang J, et al. (2000). Characterization of a predominant immunogenic outer membrane protein of *Riemerella anatipestifer*. *Clin. Diagn. Lab. Immunol.* 7: 168-174.
- Sun N, Liu JH, Yang F, Lin DC, et al. (2012). Molecular characterization of the antimicrobial resistance of *Riemerella anatipestifer* isolated from ducks. *Vet. Microbiol.* 158: 376-383.
- Tian Y, Chen W, Yang Y, Xu X, et al. (2013). Identification of B cell epitopes of dengue virus 2 NS3 protein by monoclonal antibody. *Appl. Microbiol. Biotechnol.* 97: 1553-1560.
- Wang X, Zhu D, Wang M, Cheng A, et al. (2012). Complete genome sequence of *Riemerella anatipestifer* reference strain. *J. Bacteriol.* 194: 3270-3271.
- Whitmore SE and Lamont RJ (2012). Tyrosine phosphorylation and bacterial virulence. *Int. J. Oral Sci.* 4: 1-6.
- Yang C, Bolotin E, Jiang T, Sladek FM, et al. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 389: 52-65.
- Zhong CY, Cheng AC, Wang MS, Zhu dK, et al. (2009). Antibiotic susceptibility of *Riemerella anatipestifer* field isolates. *Avian Dis.* 53: 601-607.
- Zhou JH, Zhang J, Chen HT, Ma LN, et al. (2010). Analysis of synonymous codon usage in foot-and-mouth disease virus. *Vet. Res. Commun.* 34: 393-404.