CLONING AND SEQUENCE ANALYSIS OF ENVELOPE GLYCOPROTEIN E2 OR NUCLEOCAPSID PROTEIN OF THE CLASSICAL SWINE FEVER VIRUS THAT FOUND IN THAILAND

By
Amonrat Dechpan

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree
MASTER OF SCIENCE
Department of Biology
Graduate School
SILPAKORN UNIVERSITY
2011

CLONING AND SEQUENCE ANALYSIS OF ENVELOPE GLYCOPROTEIN E2 OR NUCLEOCAPSID PROTEIN OF THE CLASSICAL SWINE FEVER VIRUS THAT FOUND IN THAILAND

By
Amonrat Dechpan

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree
MASTER OF SCIENCE
Department of Biology
Graduate School
SILPAKORN UNIVERSITY
2011

การโคลนนิ่ง และการวิเคราะห์ลำดับนิวคลีโอไทด์ แอนวิโลปไกลโคลโปรตีนอีสอง หรือ นิวคลีโอแคปสิดโปรตีนของเชื้อไวรัสอหิวาต์สุกรที่พบในประเทศไทย

โดย
นางสาวอมรรัตน์ เดชปาน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาชีววิทยา
ภาควิชาชีววิทยา
บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร
ปีการศึกษา 2554

The Graduate School, Silpakorn University has approved and accredited the Thesis title of "Cloning and Sequence Analysis of Envelope Glycoprotein E2 or Nucleocapsid Protein of the Classical Swine Fever Virus that found in Thailand" submitted by Miss Amonrat Dechpan as a partial fulfillment of the requirements for the degree of Master of Science in Biology

……………………..………………………………
(Assistant Professor Panjai Tantatsanawong, Ph.D.)
Dean of the Graduate School
………/……..…/……..…

The Thesis Advisor

1. Jundee Rabablert, Ph.D.
2. Sittiruk Roytrakul, Ph.D.

The Thesis Examination Committee

……………………………………Chairman
(Assistant Professor Supanyika Sengsai, Ph.D.)
……../……………..../……….

…………………………………Committee
(Wittawat Wiriyarat, D.V.M., Ph.D.)
……../……………..../……….

………………………………Major-advisor
(Jundee Rabablert, Ph.D.)
……../……………..../……….

………………………Co-advisor
(Sittiruk Roytrakul, Ph.D.)
……../……………..../……….

51303207: MAJOR: BIOLOGY

KEY WORD : CLASSICAL SWINE FEVER VIRUS(CSFV)/NUCLEOCAPSID/E2

AMONRAT DECHPAN : CLONING AND SEQUENCE ANALYSIS OF ENVELOPE GLYCOPROTEIN E2 OR NUCLEOCAPSID PROTEIN OF THE CLASSICAL SWINE FEVER VIRUS THAT FOUND IN THAILAND. THESIS ADVISORS : JUNDEE RABABLERT, Ph.D. AND SITTIRUK ROYTRAKUL Ph.D.101 pp.

The CSFV Bangkhen strain has been isolated in Thailand. We aimed to study the differences between Bangkhen strain and the others, the E2 and C genes of this strain were cloned and sequences were determined. Alignment of C sequences of Bangkhen strain showed clear conservation and a homology as high as 93 – 94 % compared with Thiverval and Chinese strain. Furthermore, the similar E2 sequences of these strain were approximate 94- 96 %. In this study, *Eco*RI site in nucleotides of CSFV- Bangkhen E2 gene was found, while other strains had no restriction site for this enzymes. For N-linked glycosylation prediction, the one potential N-linked glycosylation site of nucleocapsid was predicted. The E2 of Bangkhen strain showed 6 N-linked glycosylation sites, while the E2 of Thiverval or Chinese strain showed 7 N-linked glycosylation sites. Moreover, The C gene of Bangkhen strain showed 4 epitope sites whereas that of Thiverval or Chinese strain showed only 3 epitope sites.

Department of Biology Graduate School, Silpakorn University Academic Year 2011

Student's signature…………………….

Thesis Advisor's signature 1………………………… 2………………………………….

51303207 : สาขาวิชาชีววิทยา

คำสำคัญ : ไวรัสอหิวาต์สุกร/ นิวคลีโอแคปสิด / อีสอง

อมรรัตน์ เดชปาน : การโคลนนิ่ง และ การวิเคราะห์ลำดับนิวคลีโอไทด์ แอนวิโลปไกลโคลโปรตีนอีสอง หรือ นิวคลีโอแคปสิดโปรตีน ของเชื้อไวรัสอหิวาต์สุกร ที่พบในประเทศไทย. อาจารย์ที่ปรึกษาวิทยานิพนธ์ : อ. ดร. จันทร์ดี ระแบบเลิศ และ ดร. สิทธิรักษ์ รอยตระกูล. 101 หน้า

เชื้อไวรัสอหิวาต์สุกรสายพันธุ์บางเขนพบในประเทศไทย ผู้วิจัยศึกษาความแตกต่างของสายพันธุ์บางเขนกับสายพันธุ์อื่น ส่วนยีนอีสองและยีนนิวคลีโอแคปสิด จะถูกโคลน และ ลำดับนิวคลีโอไทป์ถูกนำมาวิเคราะห์ ผลการศึกษาครั้งนี้พบว่า ลำดับนิวคลีโอไทป์ของยีนนิวคลีโอแคปสิด และยีนอีสองของสายพันธุ์บางเขนมีค่าประมาณ 93% - 94 % และ 94% - 96 % ตามลำดับ หลังจากเปรียบเทียบกับสายพันธุ์ Thiverval และ สายพันธุ์ Chinese ส่วนยีนอีสองของสายพันธุ์บางเขนพบตำแหน่งจำเพาะคือเอ็นไซม์ *Eco*RI ในการหาตำแหน่ง N-linked glycosylation ยีนนิวคลีโอแคปสิด พบเพียงหนึ่งตำแหน่ง ส่วนยีนอีสองของสายพันธุ์บางเขนพบ 6 ตำแหน่ง ในขณะที่ สายพันธุ์ Thiverval และ สายพันธุ์ Chinese พบ 7 ตำแหน่ง นอกจากนั้นส่วนยีนนิวคลีโอแคปสิดสายพันธุ์บางเขน ยังพบ epitope ถึง 4 ตำแหน่ง ในขณะที่สายพันธุ์ Thiverval และ สายพันธุ์ Chinese พบเพียง 3 ตำแหน่งเท่านั้น

ภาควิชาชีววิทยา บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร ปีการศึกษา 2554

ลายมือชื่อนักศึกษา.......................................

ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์ 1. ……………………….. 2. ………………………….…

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and deep appreciation to my major advisor, Dr. Jundee Rabablert, Department of Biology, Faculty of Science, Silpakorn University and my co-advisors and Dr .Sittiruk Roytrakul , National Center for Genetic Engineering and Biotechnology (BIOTEC) for their kind expression, excellent supervision, scholarship support, advice and suggestion on technical problem and encouragement throughout the period of my graduate study and in preparation of this thesis. Without their thoughts, attenuation and advice, this thesis could not have been completed.

Special thanks to Miss. Jarunee Satra, Department of Livestock Development, Ministry of Agriculture and Cooperatives for Classical swine fever virus wildtype strain in this study.

I would like to express my special thank to Assistant Professor Dr. Supanyika Sengsai and Dr. Wittawat Wiriyarat.

I would like to express my special thank to the Faculty of Graduate Studies, Silpakorn University for supporting in part by the thesis grant.

I would like to express my special thanks to the Development and Promotion of Science and Technology talents project (DPST).

I would like to thank all staff at the laboratory room, Department of Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University.

Finally, I would like to thank my friends and my family for their love and care which are the great support to my success.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

Classical swine fever (CSF), also known as hog cholera, was highly contagious, often fatal domestic and wild swine disease with widespread economic implications (72). It was classified by the Office International des Epizooties (OIE) as a disease under strict surveillance (35, 43, 71). CSF was first reported in 1833, in Ohio, USA (69). In Thailand, the disease was reported in Bangkhen area, Bangkok (22, 112). The clinical signs of CSF may occur in acute, subacute, chronic, late onset, or inapparent course. In the acute form, there was fever, huddling of sick animals, loss of appetite, dullness, weakness, conjunctivitis, constipation followed by diarrhoea, and an unsteady gait. Several days after the onset of clinical signs, the ears, abdomen and inner thighs may show a purple discoloration. Animals with acute disease died within one to two weeks. With low virulence strains, the only expression may be poor reproductive performance and the birth of piglets with neurologic defects such as congenital tremor. CSF was primarily spread by direct contact between swine and infected swine or indirect contact between swine and fomites contaminated with virus, such as vehicles, equipment, bedding, feed, waste, or people and their clothing. CSF virus can survive in meat and pig products for many months and can be passed through this route when infected product was ingested. There was no evidence of its spread through aerosol routes but could spread short distances via mechanical vectors. Swines that were chronic carriers of the disease may show no clinical signs of illness but may shed the virus in their feces. Humans were not affected by this virus. Swine was the only specie known to be susceptible.

The classical swine fever virus (CSFV) was the causative agent of this disease. There was only one serotype of CSFV but it can be classified genetically into three genogroups (80). All genogroups had been isolated in Thailand. It belongs to the *Pestivirus* genus within the *Flaviviridae* family (39, 59, 70). The *Pestivirus* genus also contains two other viruses of economic importance: border disease virus (BDV) and bovine viral diarrhoea virus (BVDV). BVDVs (I and II) were mainly isolated from cattle, and BDV was preferably replicating in ovine species (79, 113). CSFV was restricted to swine and wild boar, while non-CSFV infects many different ruminant species as well as swine and wild boar (127).

The CSFV was an enveloped, non-segmented, positive-strand RNA about 12.3 kb, encoding a polyprotein precursor that was processed to mature proteins by cellular and viral proteases (73). The CSFV genome encodes one nonstructural protein, P7, as well as the structural proteins with the order: core which forms the nucleocapsid protein, and three glycosylated envelope proteins, E0, E1 and E2, followed by NS2, NS3, NS4a, NS4b, NS5a and NS5b nonstructural proteins (68, 93, 107).

The E2 was the major envelope glycoprotein exposed on the outer surface of the virion and represents an important target for induction of the immune response during infection (92). E2 with the presence of a hypervariable region and several amino acid substitution mutants can escape from capture by antibodies (34, 124). In contrast, the nucleocapsid (C) was good conserved between genotypes. It was a small protein and contains many basic amino acids (lysine and arginine) according to the deduced codons from cDNA (67, 121). The C protein of CSFV was also suggested to represent an important target for the host's immune reaction.

In consideration of the epizootiological importance and diagnostic diffculties of CSFV, there had been a need to identify the virus. To study the difference between infectious CSFV found in Thailand and other strains, the E2 and nucleocapsid gene of Bangkhen strain were cloned and sequences were determined.

CHAPTER II
LITERATURE REVIEW

1. Classical swine fever (CSF)

Classical swine fever or hog cholera (also sometimes called swine plague based on the German word *Schweinepest*) is highly contagious, often fatal domestic and swine disease with widespread economic implications (72). It was classified by the Office International des Epizooties (OIE) as a disease under strict surveillance (35, 43, 71). CSF was first reported in 1833, in Ohio, USA (69). In Thailand, the disease was reported in Bangkhen area, Bangkok (22, 112). CSF was for a long time widely distributed in the Americas, Europe and Asia. It was finally eradicated from North America in 1978 (114). In 1997-1998, an outbreak in the Netherlands spread to more than 400 herds and cost $2.3 billion to eradicate. Approximately 12 million swines were killed, some in eradication efforts but most for welfare reasons associated with the epidemic. The United Kingdom experienced a CSF epizootic in 2000, and minor outbreaks were reported in Romania, Slovakia, Spain and Germany in 2001. North America is also at risk for the introduction of this disease, which is still endemic in much of South and Central America, including parts of Mexico. Australia and New Zealand are currently free of the disease (81, 95)

1.1 Transmission

The most common method of transmission is through direct contact between healthy swine and those infected with CSFV. Infected swines are the only reservoir of virus. Blood, secretions and excretions (including oronasal and lacrimal secretions, urine, feces and semen) and tissues contain infectious virus. Virus shedding can begin before the onset of clinical signs, and occurs throughout the course of acute or subclinical disease. Chronically or persistently infected swines can shed virus continuously or intermittently for months. Transmission between swines occurs mainly by the oral or oronasal routes, via direct or indirect contact. It can also cross the placenta and establish a persistent infection in the developing fetus (13). CSFV is often spread by feeding uncooked contaminated garbage. Animals can also be infected through the mucous membranes, conjunctiva and skin abrasions. CSFV can spread by genital transmission or artificial insemination. Infected carrier sows may give birth to persistently infected swines. Virus is also transmitted to swines by fomites contaminated with the virus (vehicles, equipment, clothing), or when they are injected with contaminated needles. Mechanical transmission over short distances by biting insects may be possible (128). Aerosol transmission over distances of less than 500 meters has been demonstrated experimentally, but is not considered important except possibly under unusual circumstances (27). Transmission by other animals such as rodents, birds or pets contaminated by the virus has been shown to be unlikely (28).

CSFV is moderately fragile in the environment; this virus is reported to survive for three days at 50ºC (122 ºF) and 7 to 15 days at 37ºC (98.6ºF). CSFV can remain infectious for nearly three months in refrigerated meat and for more than four years in frozen meat. In this proteinaceous environment, this virus does not appear to be inactivated by smoking or salt curing. Reported virus survival times in cured and smoked meats vary with the technique, and range from 17 to more than 180 days (83). Smuggled meats have furthermore been identified as posing a very high risk for introduction of CSF (137). The major outbreaks that spread through EU countries in 1997 – 1998 were caused by a virus that is believed to have originated in Asia (71) and to have been introduced into European domestic swines via swill illegally fed to swines in a single area in Germany towards the end of 1996 (36). The incubation period can range from 2 to 15 days, depending on the virulence of the strain, the route of inoculation and the dose. Under field conditions, disease may not become evident in a herd for 2 to 4 weeks or longer.

1.2 Clinical signs

The signs of classical swine fever vary with the strain of virus, and the age and susceptibility of the swines. More virulent strains cause acute disease; less virulent strains can result in a high percentage of chronic, mild or asymptomatic infections. Although highly virulent strains were once more prevalent, most epizootics are now caused by moderately virulent strains. Older animals are less likely to show severe symptoms than younger swines. Some breed-specific differences have also been reported (26, 38). In breeding herds affected with CSF, there is often a rise in abortions, still births and weak litters. Affected swines will often go off their food, behave slowly, can be hot to touch and can suffer from weakness of the hindquarters and lameness. The acute form of CSF is highly virulent, causing persistent fevers that can raise body temperatures as high as 107 ˚F. Other signs of the acute form include convulsions and lack of appetite. Affected swines will pile or huddle up together. Signs of CSF may not be apparent for several days following infection. Death usually occurs within 5 to 14 days following the onset of illness. The chronic form of CSF causes similar clinical signs in affected swine, but the signs are less severe than in the acute form. Discoloration of the abdominal skin and red splotches around the ears and extremities often occur. Swines with chronic CSF can live for more than 100 days after the onset of infection. The mild or clinically inapparent form of CSF seldom results in noticeable clinical signs. Affected swines suffer short periods of illness often followed by periods of recovery. Eventually, a terminal relapse occurs. The mild strain may cause small litter size, stillbirths, and other reproductive failures. High mortality during weaning may also indicate the presence of this mild strain of classical swine fever virus. (72).

1.3 Diagnosis

The palatine tonsil is the site of early infection and the tissue most consistently positive for virus. It is the preferred tissue for CSFV immunostaining and virus isolation. Both the direct fluorescent antibody (DFA) test and immunohistochemistry demonstrate CSFV antigen in the tonsillar crypt, surface epithelium, and germinal centers. Other important tissues are the submandibular, gastrohepatic, and mesenteric lymph nodes, thymus, spleen, kidney, lower ileum, cecal tonsil, and brain. Histopathology is most useful as an adjunct for CSF diagnosis if combined with immunohistochemical staining. With relatively virulent CSF, swines that die often have histopathological lesions of marked lymphoid necrosis and atrophy of germinal centers in lymph nodes and spleen. Hemorrhage may be present in many organs, but is a nonspecific finding. Antigen may be demonstrated in vascular endothelium in swines with virulent CSF. There is often nonspecific mild encephalitis with perivascular cuffing of lymphocytes, and microgliosis.

1.4 Field diagnosis

Any investigation of a septicemic disease of swines causing prolonged high fever should include CSF in the differential. A thorough history, searching for evidence of contact with raw garbage, recent addition of swines of unknown origin, or contact with people who have traveled internationally, is recommended. Examine and send for laboratory diagnosis samples from sick, dying, and dead swines. Blood counts may be useful, but are sometimes abnormal only immediately after the initial rise in body temperature. An initial lymphopenia may persist for a few days, but the lymphocyte differential may return to normal or be masked by neutrophilia in a total leukocyte count. Because of the frequency of secondary diseases in a CSF outbreak, a presumptive diagnosis should not be based on only one necropsy. A definitive field diagnosis of CSF is not possible with the current strains of virus. Since there are few pathognomonic lesions of CSF (and sometimes few lesions at all), an investigation of any disease of high morbidity with persistent fever in swines should include a laboratory examination of tissues for CSF.

1.5 Laboratory diagnosis

The use of immunological and molecular genetic technology for laboratory diagnosis has progressed rapidly over the last decades. Using the direct immunofluorescence test to detect antigen in frozen tissue sections, results can be available in 2 hours (128). Rapid results are also obtainable using PCR (polymerase chain reaction) to detect viral genetic material, and these tests can distinguish between CSF virus and its close relatives (45, 65, 130). Recently, a study demonstrated that a PCR could detect CSF-viral RNA in formalin-fixed tissues, a considerable advantage under conditions where the submission of fresh samples is not possible (105). ELISA (enzyme-linked immuno-sorbent assay) technology is also used to detect viral antigen in blood or organ samples and offers a rapid way of screening large numbers of samples, but is less sensitive than the PCR (20). However, isolation of the virus remains an essential element of the laboratory armoury to diagnose CSF, although this is made difficult by the fact that it is a non-cytolytic virus and indirect methods are needed to detect it in cell cultures. ELISA is also used to detect antibodies in sera but only some of the tests available distinguish between antibodies produced in response to CSF virus and those elicited by related viruses (74, 75).

1.6 Vaccination

Vaccination is employed in many countries to control outbreaks and has been used effectively in systematic and consistent programs to reduce disease outbreaks to a point where eradication measures are feasible (11). Currently, nearly all vaccines on the market are modified live virus vaccines rendered non pathogenic for swines over 2 weeks of age and are safe to use in pregnant sows. These include the lapinized Chinese (C) strain, Japanese guinea swine cell-culture-adapted strain, and French Triverval strain (6). Subunit vaccines based on the sequence of the E2 region are currently available and have proven effective in experimental trials. It is possible to serologically distinguish swines that have been vaccinated with a subunit vaccine from swines that have recovered from a field strain of CSFV. Vaccination allowed the United States and most countries of the European Union to successfully reduce CSF to the point that sanitary measures could eradicate the disease. This requires a strong veterinary infrastructure, diagnostic laboratory system, and strong support from the government and swine industry. To prevent new outbreaks, strict importation restrictions must be applied to swines and pork products originating from countries where CSF is present. Garbage feeding must be stopped or carefully regulated to ensure proper cooking. Care must be taken to destroy all port garbage.

2. Classical swine fever virus (CSFV)

Classical swine fever virus is the causative agent of Classical swine fever. It belongs to the *Pestivirus* genus within the *Flaviviridae* family (39, 59, 70). There is only one serotype of CSFV, but it can be genetically classified into three genogroups (80). All the genogroups have been isolated in Thailand. The *Pestivirus* genus also contains two other viruses of economic importance: border disease virus (BDV) and bovine viral diarrhoea virus (BVDV), here defined as non - CSFV. BVDVs (I and II) are mainly isolated from cattle, and BDV is preferably replicating in ovine species (79, 113, 116). CSFV is restricted to swine and wild boar, non-CSFV infects many different ruminant species as well as swine and wild boar (van Rijn, 2007). Humans are not affected by this virus.

CSFV infects swines and replicates predominantly in myeloid cells, including macrophages and dendritic cells (109). The interior of the particle contains the capsid protein (C) and the genomic RNA (115). It is a spherical enveloped particle of about 40–60 nm in diameter with a single stranded RNA genome of about 12,300 bases with positive polarity, non-segmented, encoding a polyprotein precursor that is processed to mature proteins by cellular and viral proteases (71,73). The CSFV genome made up of a single open reading frame (ORF) flanked by a 3′ and 5′ non-coding region (NCR), the latter contains conserved regions, implicated in the translational events.(106). It is no poly-A on the 3′ end of the genome. (This means that virus has no post-transcriptional modifications, and has simple RNA genomes.) It encodes one nonstructural protein, P7, as well as the structural proteins with the order: core which forming the nucleocapsid protein , and three glycosylated envelope proteins, E0, E1 and E2, followed by NS2, NS3, NS4a, NS4b, NS5a and NS5b nonstructural proteins (Figure 1) (68, 107).

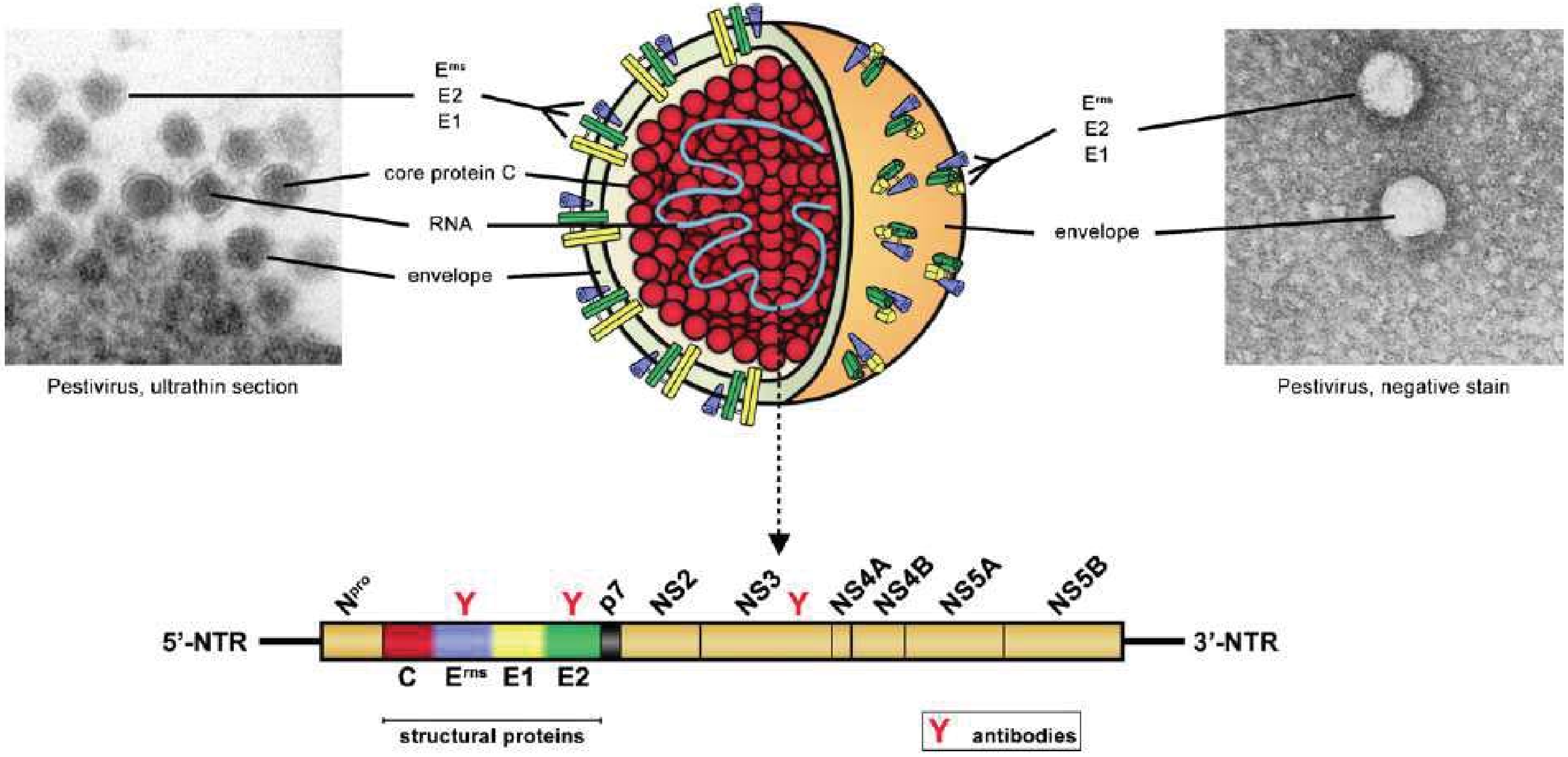

Figure 1 Schematic description of the genome organization and virion structure of CSFV.
(Electron microscopy: Dr. H. Granzow and Dr. F. Weiland; design: M. Joern; Friedrich-Loeffler-Institut Insel Riems.)

Unlike other *Flaviviridae*, the initial translation products of the *Pestivirus* genome are not structural, but nonstructural proteins such as Npro. Npro (168 amino acids, 23KDa) is the first non-structural protein encoded in the ORF protein. Npro (N-terminal auto-proteinase) exerts two known functions: an auto-protease activity for co-translational cleavage from the nascent downstream nucleocapsid protein (94, 136) and as antagonistic effect on the IFN-α/β induction pathway (90, 91, 97). Npro is a cysteine proteinase, that has similarities to subtilisin-like proteinases and which is not found in other viral systems (3). Mutants lacking the Npro gene, as opposed to wild-type CSFV, induced type I interferon in cell culture, which suggests a function in the interference with the primary cellular antiviral defense (111). Previous studies indicated that the N-terminal proteinase Npro is required for virulence of CSFV. Replacement of the Npro gene of CSFV by the murine ubiquitin gene only slightly affects the characteristics of virus replication in the porcine kidney cell line SK-6. Mutant vA187-Npro-Ubi obtained by replacement of the Npro gene by the murine ubiquitin gene in the CSFV genome, did not induce any clinical symptoms in experimentally infected pigs (4). Npro deletion mutants are attenuated and induce protective immunity in SPF (Specific pathogen free) pigs after single oronasal inoculation (64).

A signal sequence in the carboxy-terminus of capsid targets the nascent polyprotein to the membrane, initiating a dual processing event by host enzymes. Signal peptidase cleaves capsid from Erns in the ER lumen (93) and signal peptide peptidase (SPP) performs an additional processing event within the capsid membrane anchor (44). The highly basic nature of the mature capsid protein indicates that it likely functions to bind and package the RNA genome, although little is known about its role in this respect.

For replication, virus attaches to host receptors and is endocytosed into vesicles in the host cell. Fusion of virus membrane with the vesicle membrane occurs; RNA genome is released into the cytoplasm. The positive-sense genomic ssRNA is translated into a polyprotein, which is cleaved into all structural and non structural proteins. Replication takes place at the surface of endoplasmic reticulum. A negative-sense complementary ssRNA is synthesized using the genomic RNA as a template. New genomic RNA is synthesized, using the negative-sense RNA as a template. Virus assembly occurs at the endoplasmic reticulum. The virion buds at the endoplasmic reticulum, is transported to the Golgi apparatus, and then buds from the cell membrane.

2.1 Nucleocapsid (C) protein

The genome of CSFV is translated into a single polyprotein that is processed by cellular and viral proteases into 12 mature proteins. The virion consists of four structural proteins, the core protein and the glycoproteins E0, E1, and E2 (121). The nucleocapsid (core) protein of CSFV is a small protein rich in basic amino acids (lysine and arginine), according to the deduced codons from cDNA (67) and locates at or near the N terminus of the polyprotein (84). In the pestivirus polyprotein, the core protein is located between the N-terminal protease Npro and the glycoprotein Erns. Npro generates the N terminus (Ser169) of the core protein by autocatalytic cleavage of the polyprotein (94, 108). Core protein is followed by Erns, whose N terminus (Asp268) is generated by signal peptidase (signalase, or SP) (44, 93). The nucleocapsid is well conserved between genotypes. It has highly basic and similar in many biological properties found in other Faviviruses (67). In hepatitis C virus infection, it has been shown that epitopes of capsid protein are important for the T-cell and B-cell mediated immune responses (41). The capsid protein of CSFV was also suggested to represent an important target for the host's immune reaction.

2.2 Glycoprotein E2

Envelope glycoprotein E2 is the immunodominant protein of classical swine fever virus (CSFV). It contains sequential neutralizing epitopes to induce virus-neutralizing antibodies and mount protective immunity in the natural host (118). For vaccine development against CSF, E2 (E2-CSFV) is a major target in all kind of (experimental) vaccines (88, 119, 120, 135), as well as for diagnostics (21, 126, 133). Monoclonal antibodies directed against E2 can be used to discriminate between different pestivirus species as well as between strains of the same species (53). E2 is the major envelope glycoprotein exposed on the outer surface of the virion. In infected animals significant levels of antibodies are raised against E2. Entry of pestiviruses into cells is mediated by the interaction of envelope proteins E2 with cellular receptor molecules. Mapping antigenic domains on E2 of CSFV has been extensively reported in the literature (55, 122, 134, 140).

The E2 protein contains four antigenic domains, A to D. Domains A, B and C contain epitopes for neutralizing MAbs (134). These antigenic domains have been mapped to the N-terminal half of E2 and are located on two independent structural units in a proposed model of the antigenic structure of E2 (124). One structural antigenic unit consists of domains B and C (unit B/C, residues 690-800), and the other contains a highly conserved domain A (unit A, residues 766). The latter unit also contains a hydrophobic region, which is highly conserved among all pestiviruses (Figure 2).

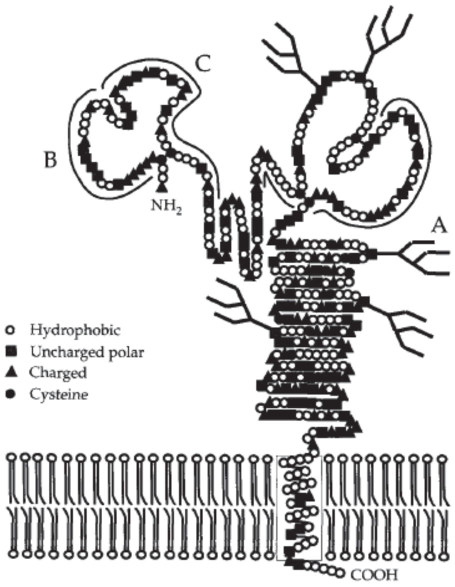

Figure 2 Proposed antigenic structure of CSFV envelope glycoprotein E2. Amino acids were grouped as follows: hydrophobic (A, G, M, I, L, V, F, W, P) ; uncharged polar (N, Q, S, T, Y) ; charged (D, E, K, R, H) and cysteine (C). Proposed disulphide bonds in the N-terminal half of E2, antigenic domains and putative glycosylation groups are indicated (125).

Genetic diversity of E2 among different groups has been extensively studied (14, 15, 58, 78, 80, 103). The N-terminal half of E2 is more variable than the C-terminal half (14). Different patterns of reactivity with mAbs provided clues of antigenic variation of E2 among different CSFV isolates (12, 33, 53, 66, 77,). A study using neutralizing mAbs to select mAb-resistant mutants showed that, in most cases, single point mutations could lead to complete loss of mAbs binding (132). Furthermore, amino acid (aa) substitutions at position 710 on the E2 proteins of different strains affected binding and neutralization by a panel of mAbs (127).

3. RNA isolation

RNA isolation is one of the most often underestimated critical steps when RT-PCR is used for mass sample screening. Two points in the isolation process have to be carefully considered: the treatment and handling of the samples prior to RNA isolation and the storage of the isolated RNA (7). Since the first step in the procedure is lysis of the cell by adding a strong denaturant (such as SDS, phenol) that inactivates the enzymes including the RNases. It is typically added prior to and after the isolation, when RNA integrity is at risk. Pretreatment of the samples and buffers for optimal performance of the commercially available kits are optimized for the kind of samples used (cell culture supernatant, serum, tissue, blood), and efficient RNA isolation can only be achieved if the right kits are used for the right samples. Different kits are recommended by the manufactures for RNA extraction from different samples, such as serum, plasma, and cell culture supernatants, or from blood or from tissues. Two technologies are in use for nucleic acid isolation for RT-PCR: the liquid-based and the silica- or glass-based methods.

The liquid-based methods are enhancements of the original nucleic acid isolation using guanidinium thiocyanate/acid phenol: chloroform (17, 18). Ready-to-use reagents are commercially available under different brand names (TRIZOL®, TRI Reagent®, Stratagene RNA Isolation Kit®, etc.), allowing single-step disruption/separation procedures. The tissue or cell sample is disrupted and homogenized in the ready-to-use reagent, chloroform is mixed with the lysate, and then the mixture is separated into three phases by centrifugation. The RNA is then precipitated from the aqueous phase with isopropanol. This extraction method produces high yields of intact RNA molecules. The drawbacks are that skilled personnel are needed and that the extraction cannot be automated. Therefore, it is mostly used for RNA extraction from a few or single samples and more often from “difficult” sample (mainly rotten or degraded tissue).

Silica- or glass-based matrices or filters selectively absorb nucleic acid in the presence of chaotropic salts (9), which immediately inactivate the RNAses. After washing off the remaining components of the lysate using a high-salt buffer, the bound RNA is eluted with water. In many commercially available kits, the glass or silicon filter or fleece is housed in a spin column or in a 96-well plate. Solution are driven through the filter by centrifugation or under vacuum. Some kits also include a proteinase K or DNAse incubation for removal of protein and DNA, respectively. Although the yield of RNA is lower than that obtained with guanidinium thiocyanate/acid phenol: chloroform extraction, the silica-based kits can be adapted to be used in nucleic acid extraction robots; thus, allowing a full automation of CSFV diagnosis. In addition, no reagent is need.

4. RT-PCR

RT-PCR (reverse transcription-polymerase chain reaction) is a sensitive method for the detection of mRNA expression levels. Traditionally RT-PCR involves two steps: the RT reaction and PCR amplification. RNA is first reverse transcribed into cDNA using a reverse transcriptase, the resulting cDNA is used as templates for subsequent PCR amplification using primers specific for one or more genes. RT-PCR can also be carried out as one-step RT-PCR in which all reaction components are mixed in one tube prior to starting the reactions. Although one-step RT-PCR offers simplicity and convenience and minimizes the possibility for contamination, the resulting cDNA cannot be repeated used as in two step RT-PCR.

One-step RT-PCR combines the first-strand cDNA synthesis (reverse transcription) reaction and PCR reaction in the same tube, simplifying reaction setup and reducing the possibility of contamination. One-step RT-PCR allows easier processing of large numbers of samples, and helps minimize carryover contamination, since tubes are not opened between cDNA synthesis and amplification. By amplifying the entire cDNA sample, one-step RT-PCR can provide greater sensitivity - from as little as 0.01 pg total RNA. One-step reactions allow for the use of sequence-specific primers only.

Two-step PCR begins with the reverse transcription of either total RNA or poly(A)+ RNA into cDNA using a reverse transcriptase. Following first-strand synthesis, the cDNA is transferred to a separate tube for the PCR step. Two-step RT-PCR is useful for detecting multiple messages from a single RNA sample. You'll get greater flexibility when choosing primers and polymerase than with one-step RT-PCR systems. When performing two-step RT-PCR, you have the option of using either oligo(dT), random hexamer, or gene-specific primers, and then PCR is performed with either Taq DNA Polymerase.

After RT-PCR, reaction mixture containing the amplicon is subjected to electrophoresis in agarose gel. Double-stranded DNA is visualized under ultraviolet (UV) light after incubation with a fluorescent intercalating stain, generally with ethidium bromide. By including a DNA marker, the size of the amplicon can be estimated and eventual false amplicons and primer-dimers become visible.

The RT-PCR protocols for diagnosis of CSF had been used in many laboratories such as the protocol amplifies 5′untranslated region (129), E2 glycoprotein gene (49), NS5B non-structural protein gene (29).

5. Plasmid

5.1 The pGEM-T easy vector

The pGEM-T easy vector (Promega, USA) is a commercial plasmid that is linear. It is convenient systems for the cloning of PCR products. The vector is prepared by cutting Promega's pGEM®-T Easy Vectors with EcoRV and adding a 3′ terminal thymidine(T) to both ends. The T-overhangs at the insertion site greatly improve the efficiency of ligation of PCR products by preventing recircularization of the vector and providing a compatible overhang for PCR products generated by certain thermostable polymerases. The pGEM-T easy vector is high copy number vectors containing T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme β-galactosidase.

Insertional inactivation of the α-peptide allows identification of recombinants by blue/white screening on indicator plates. The molecular mechanism for blue/white screening is based on a genetic engineering of the lac operon in the Escherichia coli DH5α. The vector encodes the α subunit of LacZ protein with an internal multiple cloning site (MCS), while the chromosome of the host strain encodes the remaining Ω subunit to form a functional β-galactosidase enzyme. The MCS can be cleaved by different restriction enzymes so that the foreign DNA can be inserted within the lacZα gene, thus disrupting the production of functional β-galactosidase. The chemical required for this screen is X-gal, a colourless modified galactose sugar that is metabolized by β-galactosidase to form 5-bromo-4-chloro-indoxyl which is spontaneously oxidized to the bright blue insoluble pigment 5,5'-dibromo-4,4'-dichloro-indigo, and thus functions as an indicator. Isopropyl β-D-1-thiogalactopyranoside (IPTG), which functions as the inducer of the Lac operon, can be used in some strains to enhance the phenotype, although it is unnecessary with many common laboratory strains. The hydrolysis of colourless X-gal by the β-galactosidase causes the characteristic blue colour in the colonies; it shows that the colonies contain vector without insert. White colonies indicate insertion of foreign DNA and loss of the cells ability to hydrolyse the marker.

The pGEM-T easy vector contains numerous restriction sites within the multiple cloning region. This region is flanked by recognition sites for the restriction enzymes EcoRI, BstZI and NotI, providing three single-enzyme digestions for release of the insert. The pGEM-T easy vector cloning region is flanked by recognition sites for the enzyme BstZI. Alternatively, a double-digestion may be used to release the insert from either vector. Additionally, the plasmid also contained ampicillin resistance gene (*Amp*^{r}) that will allowed growing of the recombinant bacteria in selection media containing ampicillin.

The pGEM-T easy vector had been used in many laboratories such as the Recombination of CSFV - E2 gene in *Chlamydomonas reinhardtii* chroloplasts (32), Cloning of antifungal gene from *Bacillus licheniformis* (61).

5.2 The pTZ57R/T vector

The InsTAclone™ PCR Cloning Kit is a TA system for direct one-step cloning of PCR products with 3'-dA overhangs. Overhangs of this type are generated during PCR with *Taq* DNA polymerase and other thermostable DNA polymerases which lack proofreading activity. The high quality TA cloning vector pTZ57R/T is ready to use for efficient ligation with PCR products providing high cloning yields and low background. To increase the speed, convenience and efficiency of cloning, the InsTAclone™ PCR Cloning Kit has been combined with the TransformAid™ Bacterial Transformation Kit. According to protocol, ligation and preparation of competent cells is performed in parallel. Therefore, it only takes approximately one hour from the completion of PCR to plating of transformed cells. Our transformation protocol is often faster than transformation of commercially available competent cells. The DNA insert can be readily excised from the versatile polylinker of pTZ57R/T, sequenced using standard M13/pUC primers or *in vitro* transcribed with T7 RNA polymerase. The InsTAclone™ PCR Cloning Kit takes advantage of the terminal transferase activity of *Taq* DNA polymerase and other non-proofreading thermostable DNA polymerases. Such enzymes add a single 3'-A overhang to both ends of the PCR product. The structure of these PCR products favors direct cloning into a linearized cloning vector with single 3'-ddT overhangs. Such overhangs at the vector cloning site not only facilitate cloning, but also prevent the recircularization of the vector. As a result, more than 90% of recombinant clones contain the vector with an insert.

The pTZ57R/T vector had been used in many laboratories such as the cloning of fusion protein and nucleoprotein gene segments of PPRV (50), E1 gene of chikungunya virus (76).

5.3 The pET-17b vector

The pET-17b vector includes an N-terminal 11aa T7•Tag sequence and a region of useful cloning sites. Included in the multiple cloning region are two BstX I sites, which allow efficient cloning using an asymmetric linker. Unique sites are shown on the vector map except for the two BstX I sites. The pET System is the most powerful system yet developed for the cloning and expression of recombinant proteins in *E. coli*. Target genes are cloned in pET plasmids under control of strong bacteriophage T7 transcription and (optionally) translation signals; expression is induced by providing a source of T7 RNA polymerase in the host cell. T7 RNA polymerase is so selective and active that almost all of the cell's resources are converted to target gene expression; the desired product can comprise more than 50% of the total cell protein a few hours after induction. Another important benefit of this system is its ability to maintain target genes transcriptionally silent in the uninduced state. Target genes are initially cloned using hosts that do not contain the T7 RNA polymerase gene, thus eliminating plasmid instability due to the production of proteins potentially toxic to the host cell. Once established in a nonexpression host, plasmids are then transferred into expression hosts containing a chromosomal copy of the T7 RNA polymerase gene under *lacUV5* control, and expression is induced by the addition of IPTG.

The pET-17b vector had been used in many laboratories such as the expression of P2 protein *Haemophilus influenzae* (48), glycine N-methyltransferases (60).

6. 3D structure protein prediction

Homology models of proteins are of great interest for planning and analyzing biological experiments when no experimental three dimensional structures are available. Building homology models requires specialized programs and up-to-date sequence and structural databases. SWISS-MODEL workspace (http://swissmodel. expasy.org/workspace/) is a web-based integrated service dedicated to protein structure homology modeling. It assists and guides the user in building protein homology models at different levels of complexity. Protein sequence and structure databases necessary for modeling are accessible from the workspace and are updated in regular intervals. Tools for template selection, model building and structure quality evaluation can be invoked from within the workspace. (2). SWISS-MODEL had been used in many laboratories such as the exploring the sub site specificity of *Schistosoma mansoni* (104), Quasi-atomic model of bacteriophage T7 Procapsid Shell (138).

7. Glycosylation

Most proteins do not perform their function without undergoing some form of post translational modification. The addition of a carbohydrate moiety to a protein molecule is referred to as protein glycosylation. Glycosylation (42, 89, 982), a common post translational modification, plays a role in protein folding, transport and half-life, as well as being involved in cell-cell interactions and antigenicity. Glycosylation is an enzymatic process, with the exception of glycation, and involves the addition of sugars to the protein to build up glycan chains. During this process, the linking of monosaccharide units to the amino acid chains sets up the stage for a series of enzymatic reactions that lead to the formation of glycoproteins (n and o linked oligosaccharides that are found to a protein entity). In all 16 known enzymes are supposed to mediate this reaction. A typical glycoprotein has at least 41 bonds which involve 8 amino acids and 13 different monosaccharide units and includes the glycophosphatidylinositol (GPI) and phosphoglycosyl linkages. The major sites of protein glycosylation in the body are ER, Golgi body, nucleus and the cell fluid. There are four types of glycosylation: N-linked, O-linked, C-mannosylation and GPI (glycophosphatidyl-inositol) anchor attachment. C-mannosylation involves the addition of α-mannopyranosyl to the indole of tryptophan. GPI anchors concern membrane anchoring of a protein by the addition of GPI near the C terminus. O-linked and N-linked glycosylation are the most common.

O-linked glycosylation consists of the stepwise build-up of various sugars on Ser or Thr residues. O-glycosylation has no known consensus sequence. However, Pro is often present around O-glycosylation sites and O-glycosylation occurs at a later stage in protein processing (19). O-linked glycosylation of secreted and membrane bound proteins is a post-translational event that takes place in the cis-Golgi compartment after N-glycosylation and folding of the protein. It refers to the attachment of glycans to serine and threonine, and, to a lesser extent, to hydroxyproline and hydroxylysine. O-linked glycans play important roles in protein localization and trafficking, protein solubility, antigenicity and cell-cell interactions. O-linked glycans are built up in a stepwise fashion with sugars added incrementally. The most common type of O-glycosylation in secreted and membrane-bound mammalian proteins seems to be the addition of reducing terminal Nacetylgalactosamine (GalNAc). This type of O-linked glycan is also referred to as ‘mucin-type’ glycan. The reducing terminal GalNAc residue can be further extended with galactose (Gal), N-acetylglucosamine (GlcNAc) or GlcNAc and Gal resulting in 8 common core structures, which are often further decorated with the addition of up to three sialic acid residues. In addition to the mucin-type O-linked glycans, a variety of mammalian proteins are known to have mannose , fucose , glucose , Gal or xylose as reducing terminal linkages. Some cytoplasmic and nuclear proteins have simple O-linked glycans in which a single N-acetylglucosamine residue is linked to a serine or a threonine. This modification has been identified in a number of eukaryotes including plants and filamentous fungi, although its presence in S. cerevisiae is disputed. This type of O-linked glycosylation plays an important role in the modulation of the biological activity of intracellular proteins; in some proteins the same residue may be subject to competing phosphorylation and O-linked glycosylation.

N-linked glycosylation begins with the addition of a 14-sugar precursor to an asparagine amino acid. It contains glucose, mannose and n-acetylglucosamine molecules. This entity is then transferred to the ER lumen. The oligosaccharyl transferase enzyme attaches the oligosaccharide chain to asparagine that occurs in the tripeptide sequence, Asn-X-Ser or Asn-X-Thr (8) or in some rare cases Asn-Xxx-Cys. X can be any amino acid other than Proline, is required for N-glycosylation, although not sufficient on its own. The oligosaccharide attached protein sequence now folds correctly and is now translocated to the Golgi body where the mannose residue is removed.

The ‘NetNGlyc’ 1.0 server is a N-linked glycosylation site predictor (http://www.cbs.dtu.dk/services/ NetNGlyc). It can predicts N-Glycosylation sites using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequences. Sequences having N-glycosylation potential >0.5 are considered as cut-off value. The method is described in detail in the following article "Prediction of N-glycosylation sites in human proteins" The NetNGlyc’ server had been used in many laboratories such as the prediction of N-Linked Glycosylation of CSFV Strain Brescia E2 Glycoprotein (86), HA and NA gene of Human Influenza H1N1 Virus (110), AfOch1 Protein of *Aspergillus fumigatus* (54).

8. Epitope

An epitope, also known as antigenic determinant, is the part of an antigen that is recognized by the immune system, specifically by antibodies, B cells, or T cells. The part of an antibody that recognizes the epitope is called a paratope. Although epitopes are usually thought to be derived from non-self proteins, sequences derived from the host that can be recognized are also classified as epitopes. The epitopes of protein antigens are divided into two categories, conformational epitopes and linear epitopes, based on their structure and interaction with the paratope. A conformational epitope is composed of discontinuous sections of the antigen's amino acid sequence. These epitopes interact with the paratope based on the 3-D surface features and shape or tertiary structure of the antigen. Most epitopes are conformational. By contrast, linear epitopes interact with the paratope based on their primary structure. A linear epitope is formed by a continuous sequence of amino acids from the antigen.

For epitope prediction, Emini surface accessibility scale method was developed for comparing hepatitis A virus （HAV） sequence and poliovirus type1 sequence by predicted surface features, based upon indices of surface probability (37). This method assumes the absence of significant internal deletions or insertions. For a given amino acid sequence, a point for sequence number n is a normalized product of the surface probabilities of amino acids in positions n - 2 to n + 3, using the empirical amino acid accessible surface probabilities of Janin (47) which are fractional probabilities (0.26 to 0.97) determined for an amino acid found on the surface of a protein, based upon structural data from 28 proteins. A surface residue is defined as one with >20 A° (2.0 nm) of water-accessible surface. With these fractional surface probabilities for amino acids, a surface probability (S) at sequence position n can be defined as follows:

$$Sn = \left(\prod_{i=1}^{6} \delta_{n+4+i}\right)(0.37)^{-6}$$

S_x is the fractional surface probability for the amino acid at position x. The Sn for a random hexapeptide sequence = 1.0, with probabilities greater than 1.0 indicating an increased probability for being found on the surface.

The Emini surface accessibility scale method had been used in many laboratories such as the prediction of epitopes in Nanoviridae nano-organisms (23), apxIVA in *Actinobacillus pleuropneumoniae* (102).

CHAPTER III

MATERIALS AND METHODS

1. Viruses

The CSFV Bangkhen strain was obtained from Department of Livestock Development, Ministry of Agriculture and Cooperatives, Thailand. This virus was used as the RNA template for one step RT PCR in this study.

2. RNA extractions and one step RT-PCR

Total cellular RNA and the CSFV RNA from spleen suspension in PBS pH 7.0 was isolated by using RNA purification Mini kit (QIAgen, Germany), according to the manufacture's instructions, to give rapid extraction and high quality extracted RNA. The RNA pellet was resuspended in RNase-free water and stored at – 80ºC until use.

The extracted RNA was amplified by one step RT-PCR with E2 or C primer pairs. One step RT-PCR reaction mixture of 50 μl was prepared with the following reaction component: 20 μl of template RNA(10 ng), 10 μl of 5x QIAGEN OneStep RT-PCR Buffer containing 12.5 mM $MgCl_2$, 2 μl of 2.5 mM dNTPmix, 2 μl of Enzyme mix and 1 μl of of each 20 μM primer. The amplifications were carried out using a thermocycler as follows: A reverse transcription reaction at 50ºC for 30 minutes and initial activation at 95ºC for 15 minutes followed by 35 cycles consisting of denaturation at 94ºC for 1 minute, primer annealing at 53ºC for 1 minutes and extension at 72ºC for 2 minutes; followed by a final extension step at 72ºC for 10 minutes. All reactions ended with a final 4ºC hold step.

3. Cloning and construction of recombinant plasmid

The 1035 E2 gene fragment was amplified by using two oligonucleotides from the CSFV C/HVRI strain (Genbank: AY805221). The forward primer CSE23 contained a sequence for an *Eco*RI restriction site, while the reverse primer CSE22 contained a *Xho*I restriction site. To generate the 403 C gene, primers CSC4 and

CSC3 were used. The forward primer CSC4 contained a sequence for an *Eco*RI restriction site, while the reverse primer CSC3 contained a *Hind*III restriction site (Table 1).

The blunt E2 or C gene fragment were ligated into multi-cloning site of pGEM – T Easy vector (Promega, USA), according to instructions from the supplier (Figure 3). The ligation mixtures was transformed with heat-shock method into competent *Escherichia coli* DH5α cells for 45–50 seconds in a water bath at 42ºC and returned the tubes to ice for 2 minutes. The cells were added with 1 ml of LB medium (1 % bacto-tryptone, 0.5% bacto-yeast extract and 0.5% NaCl, pH 7.0) and incubated for 1.5 hours at 37ºC with shaking (~150 rpm). Aliquots were spread on LB/ampicillin/IPTG/X-Gal plates and incubated overnight at 37ºC. Clones were picked out randomly through blue/white screening. White colonies generally contained the insert gene.

The present of E2 or C gene were checked. The positive single clone described above was inoculated with 2 ml LB medium with 100 µl / ml ampicillin and was incubated overnight at 37ºC with 200 rpm shaking. The recombinant plasmid DNA was purified, using the QIAprep mini kit (QIAGEN, Germany). The DNA was double digested with restriction enzymes: *Eco*RI – *Xho*I for E2 gene and *Eco*RI – *Hind*III for C gene. The products from restriction enzyme digestion were visualized by 1.0 % agarose gel electrophoresis after stained with ethidium bromide.

Table 1 Primers used for one step RT-PCR amplification and sequencing.

Region	**Primer name**	**Primer sequence(5´-3´)**	**Primer length**	**Tm (50 mM Na^+)**	**% GC**
Nucleocapsid	CSC4	TGGGAATTCAAGGTAGTGACGGAAAGCTGTACC	33	67.03	48.48
	CSC3	AGAAAGCTTGGCTTTTTCTAATTTCTTCCTAG	32	60.77	34.38
E2	CSE23	TCGGAATTCAAGCACAAGGCCGGCTAGCCTG	31	69.98	58.06
	CSE22	CAGTACTCGAGAAATTCTGCGAAGTAATCTGAG	33	64.54	42.42

4. DNA sequencing and sequence analysis

The purified recombinant plasmid was sent to First BASE Laboratories (Selangor, Malasia) for DNA sequencing service.

Clustal W (117) was used to align all nucleotide sequences. The E2 and C sequence of Bangkhen strain were compared with Thiveval (Genbank: EU490425) and Chinese strain (Genbank: AY805221). Since both strains were used for vaccine production in Thailand.

5. Prediction of 3D structure protein for the CSFV- E2 and CSFV - C

The SWISS-MODEL Workspace was used to find 3D structure protein of E2 and C of CSFV. (http://swissmodel.expasy.org/workspace/) (2, 40, 96)

6. Prediction of N-Linked Glycosylation profile for the CSFV- E2 and CSFV - C

NetNGlyc 1.0 server was used to find the N-glycosylated proteins of E2 and C. NetNGlyc 1.0 server (http://www.cbs.dtu.dk/ services/NetNGlyc/) predicts N-Glycosylation sites using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequences. Sequences having N-glycosylation potential >0.5 are considered as cut-off value (101).

7. Prediction of epitope for the CSFV- E2 and CSFV - C

Antibody Epitope Prediction progame : IEDB Analysis Resource was used to find epitope sequence of E2 and C. Emini surface accessibility scale method (http://tools.immuneepitope.org/tools/bcell/iedb_input) predicts epitope site by using formulae

$$Sn = \left(\prod_{i=1}^{6} \delta_{n+4+i} \right) (0.37)^{-6}$$

Where Sn is the surface probability, dn is the fractional surface probability value, and i vary from 1 to 6. A hexapeptide sequence with Sn greater than 1.0 indicates an increased probability for being found on the surface (37).

8. Construction of recombinant plasmid for protein expression.

pTZ57R/T vector (Figure 4) was used for cloning vector and pET – 17b vector was used for expression of E2 and C. Start codon and stop codon were added at early part and the end of genes by primers. New forward primers contained a sequence for a start codon (ATG) and a *Nde*I restriction site while the new reverse primers contained a stop codon (TAG) and a *Xho*I restriction site. The PCR product of CSFV- E2 fragment or CSFV- nucleocapsid protein gene fragment was ligated into multi-cloning site of pTZ57R/T cloning vector (Fermentas, USA) ratio 1:3. Insert genes in recombint plasmid were digested with *Nde*I and *Xho*I enzymes and were ligated into pET – 17b vector for expresstion. (Figure 5)

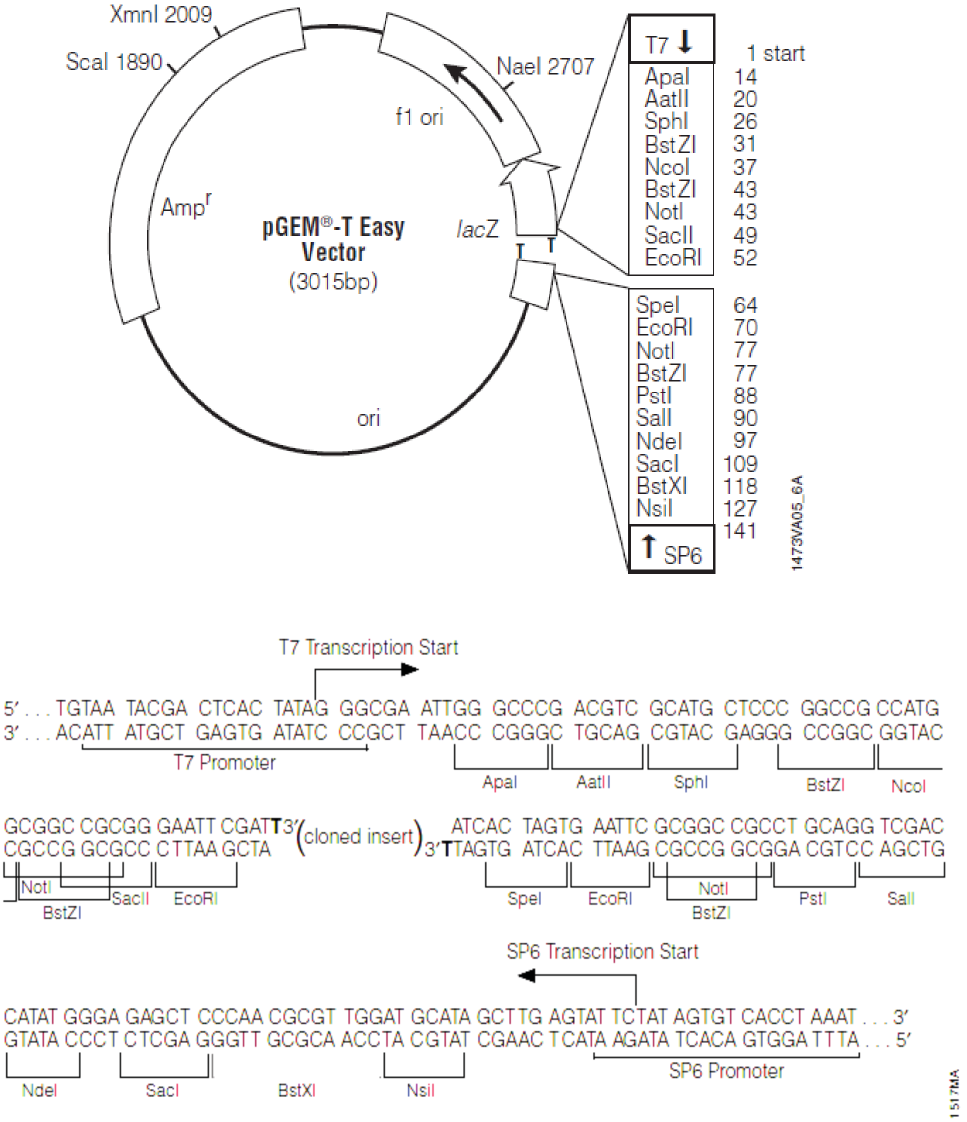

Figure 3 The pGEM-T Easy Vector Map and multiple cloning sequences. The vector has been linearized at base 60 with *Eco*RV and a T added to both 3′ends. The top strand shown corresponds to the RNA synthesized by T7 RNA polymerase. The bottom strand corresponds to the RNA synthesized by SP6 RNA polymerase.

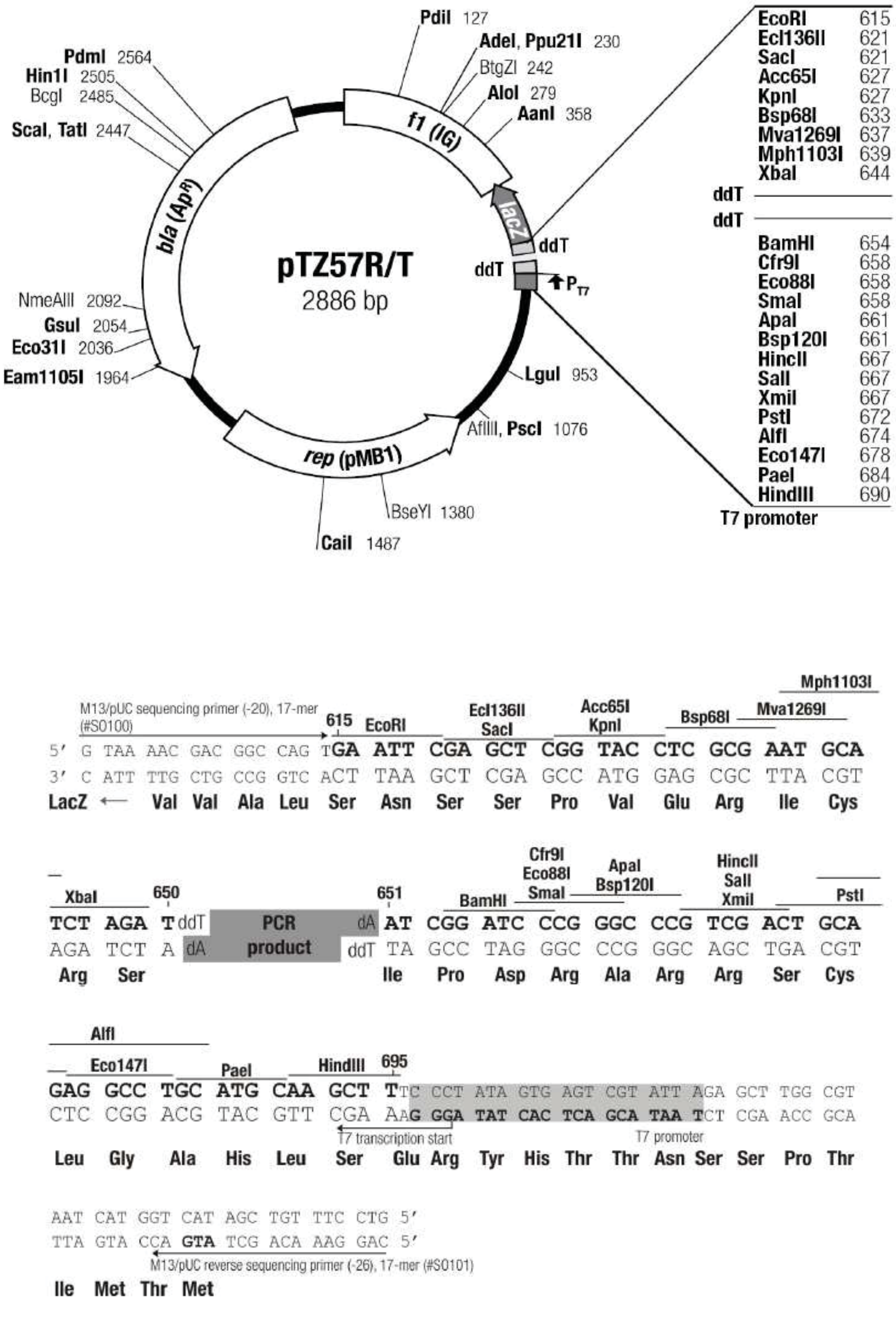

Figure 4 The pTZ57R/T cloning vector Map and multiple cloning sequences. Unique restriction sites were indicated.

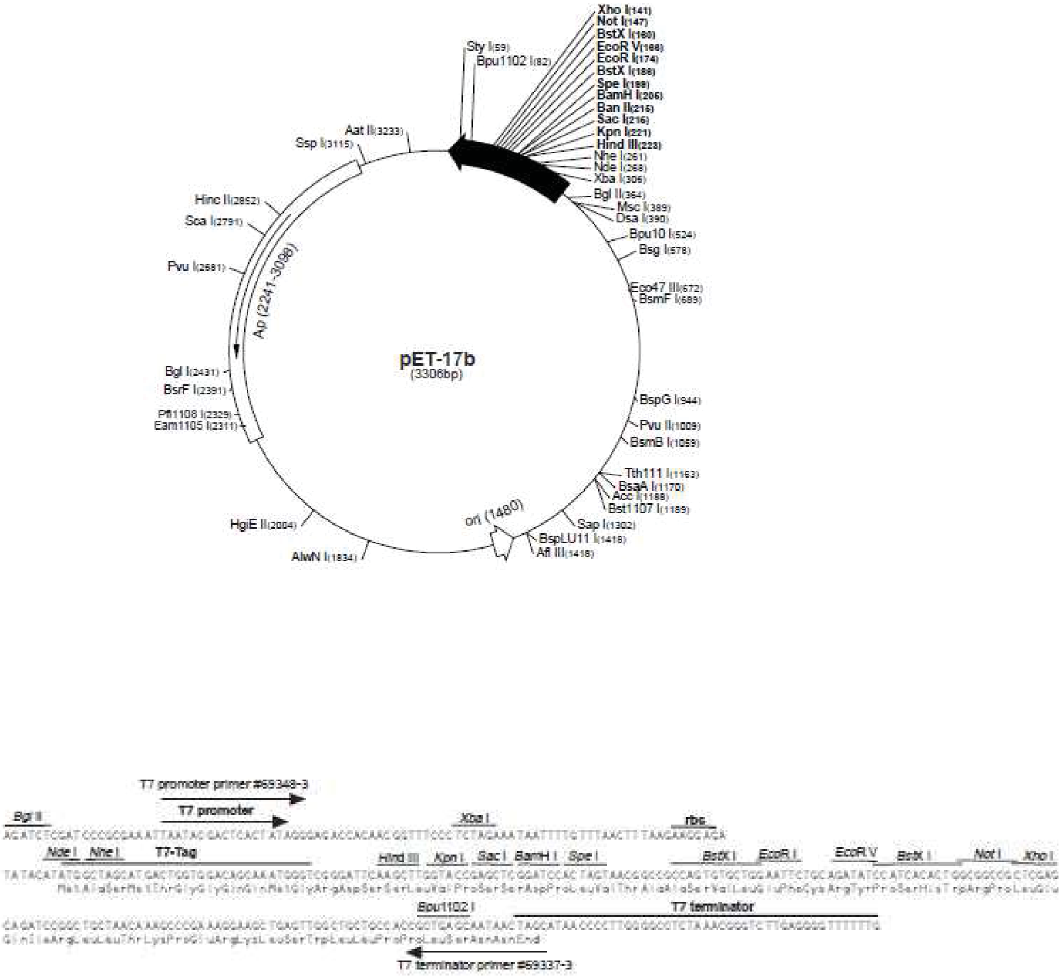

Figure 5 The pET – 17b vector Map and cloning/expression region.

CHAPTER IV

RESULTS

1. One step RT-PCR amplification of E2 or C encoding sequences

The infectious RNA of Classical swine fever virus Bangkhen strain that found in Thailand was used as a template. DNA of E2 or C genes was amplified with one step RT-PCR. The results showed a specific band of about 400 bp for C gene and about 1000 bp for E2 gene. The sizes of amplified DNAs were shown as expected (Figure 6).

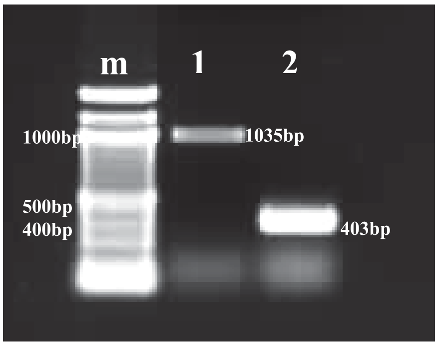

Figure 6 One step RT-PCR products of E2 or C from infectious CSFV Bangkhen strain found in Thailand, lane m: DNA markers (3,000, 1,500, 1,000, 900, 800, 700, 600, 500, 400, 300, 200, 100), lane 1 : E2 gene (1035 bp) and lane 2: C gene (403 bp).

2. Cloning of the one step RT-PCR products

DNA products were electrophoresed on 1.0 % agarose gel. Then, the bands of interest were extracted from the gel using QIAquick Gel Extraction Kit (QIAGEN, Germany). The purified products were clone into pGEM® – T Easy vector directly. The recombinant plasmid DNA extracted from the positive clones was used for restriction endonuclease analysis. Recombinant plasmid DNA containing E2 or C gene were double digested with restriction enzymes: *Eco*RI – *Xho*I for E2 gene and *Eco*RI – *Hind*III for C gene (Figure 7). From analyzing the gel picture (Figure 7), it can be assumed that all molecules were recombined, resulting a fragment of 3015 bp related to the liniarized vector and another fragment, corresponding to E2 and C gene.

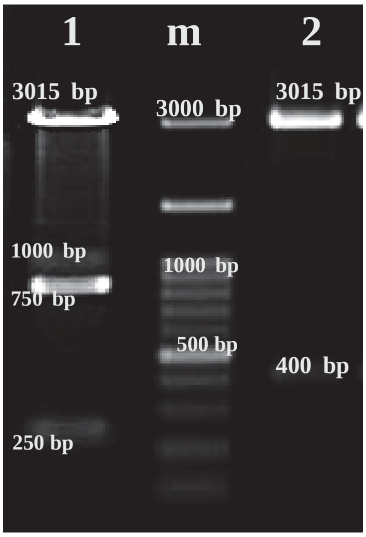

Figure 7 Digestion of recombinant plasmids, lane m: DNA markers (3,000, 1,500, 1,000, 900, 800, 700, 600, 500, 400, 300, 200, 100), lane 1: recombinant CSFV Bangkhen–E2 was digested with *Eco*RI – *Xho*I and lane2: recombinant CSFV Bangkhen–C was digested with *Eco*RI – *Hind*III.

As shown in figure 7 lane 2, a fragment of about 400 bp in size that corresponds of the C gene as expected. Thus, the nucleotides of CSFV-Bangkhen C gene were not cut by *Eco*RI and *Hind*III. Three bands at about 1000 bp, 750 bp and 250 bp were seen in lane 1 when the recombinant plasmid DNA with E2 was digested. The result suggested that the restriction endonuclease cutting site *Eco*RI – *Xho*I occurred in CSFV Bangkhen E2 gene.

3. Sequence analysis

CSFV-Bangkhen C gene and E2 were chosen for nucleotides sequencing. The recombinant plasmid DNA encoding C gene or E2 inserted in pGEM ®– T Easy vector were sequenced by First BASE Laboratories (Selangor, Malasia). The nucleotides sequence were compared with that of Thiverval strain (Genbank: EU490425) and Chinese strain (Genbank: AY805221). The results of C gene showed clear conservation and their homology as high as 93% - 94% was found among these stains (Table 2).

Table 2 Alignment of C gene of Bangkhen strain compared with Thiverval and Chinese strain.

```
             ……………..…|…………20……………..|…………40……………|………… 60……………..|…………...80
1 Chinese    GGTAGCGACG  GAAAGCTTTA  CCATACATAT  GTGTGCATCG  ATGGCTGCAT  ACTGCTGAAG  CTGGCCAAGA  GGGGTGAGCC
2 Thiverval  GGTAGTGACG  GAAGGCTTTA  CCATATATAT   GTGTGCATCG  ATGGTTGCAT  ACTGCTGAAG  CTAGCCAAAA  GGGGCGAGCC
3 Bangkhen   GGTAGTGACG  GAAAGCTGTA CCATATATAT   GTGTGCATCG  ATGGCTGCAT ACTGCTGAAG  CTAGCCAAGA  GGGACGAGCC

             ……………..…|…………100……………..|……… 120……………|…………140……………..|…………160
1 Chinese    AAGAACCCTG  AAGTGGATTA  GAAATTTCAC  CGACTGTCCA   TTGTGGGTTA  CCAGTTGCTC  CGATGATGGC  GCAAGTGGGA
2 Thiverval  AAGAACCCTG  AAGTGGATTA  GAAATTTCAC  CGACTGTCCA   TTGTGGGTTA  CCAGTTGCTC  TGATGATGGC  GCAAGTGGAA
3 Bangkhen   AAGAACCCTG  AAGTGGATTA  GAAATCTTAC  CGACTGCCCA   CTGTGGGTTA  CTAGTTGCTC  TGATGATGGC  GCAAGTGGAA

             ……………..…|…………180……………..|…………200……………|…………..220……………..|…………240
1 Chinese    GTAAAGAGAA  GAAGCCAGAT  AGGATCAACA  AAGGCAAATT  AAAAATAGCC  CCAAAAGAGC  ATGAGAAGGA  CAGCAGAACT
2 Thiverval  GTAAAGAGAA  GAAGCCAGAT  AGGATCAACA  AGGGTAAATT  AAAAATAGCC  CCGAAAGAGC  ATGAGAAGGA  CAGCAGAACT
3 Bangkhen   GTAAAGAGAA  GAAGCCAGAT  AGGATCAGCA  AGGGAAAATT  AAAAATAGCC  CCAAAAGAGC  ATGAGAAGGA
             CAGCAGGACC

1 Chinese    ……………..…|…………260……………..|…………280……………|…………300……………..|…………320
2 Thiverval  AGGCCACCTG  ACGCTACGAT  CGTGGTGGAA  GGAGTAAAAT  ACCAGGTCAA  AAAGAAAGGT  AAAGTTAAAG  GAAAGAATAC
3 Bangkhen   AAGCCACCTG  ACGCTACGAT  CGTAGTGGAA  GGAGTAAAAT  ACCAGGTCAG  AAAGAAAGGT  AAAGTTAAAG  GAAAGAATAC
             AAGCCACCTG  ATGCTACGAT  TGTAGTGGAA  GGAGTAAAGT  ACCAGGTCAA  AAAGAAAGGG  AAAGTTAAGG  GAAAGAATAC

1 Chinese    ……………..…|…………340……………..|…………360……………|…………..380……………..|…………400
2 Thiverval  CCAAGACGGC  CTGTACCACA  ACAAGAATAA  ACCACCAGAA  TCCAGGAAGA  AATTAGAAAA  AGCC
3 Bangkhen   CCAAGACGGC  CTGTACCACA  ATAAGAATAA  ACCACCAGAA  TCTAGGAAGA  AATTAGAAAA  AACC
             CCAAGACGGC  CTGTACCACA  ACAAGAATAA  ACCACCAGAA  TCTAGGAAGA  AATTAGAAAA  AGCC

             Sequences (Chinese: Thiverval)      Aligned. Score:  95
             Sequences (Chinese: Bangkhen)     Aligned. Score:  93
             Sequences (Thiverval: Bangkhen) Aligned. Score:  94
```

Table 3 Alignment of C amino acid sequence of Bangkhen strain compared with Thiverval and Chinese strain.

	……………10…………..20…………...30……………40……………50
Chinese_aa	GSDGKLYHTY VCIDGCILLK LAKRGEPRTL KWIRNFTDCP LWVTSCSDDG
Thiverval _aa	……...R ……I……………………………………………………………………..
Bangkhen_aa	…………….. I………………………....D…………………L……………………….
	…………..60…………..70…………..80…………90…………..100
Chinese_aa	ASGSKEKKPD RINKGKLKIA PKEHEKDSRT RPPDATIVVE GVKYQVKKKG
Thiverval _aa	………………………………………………………K…………………………R……
Bangkhen_aa	……………………S……………………………..K………………………………
	…………110…………..120…………..130
Chinese_aa	KVKGKNTQDG LYHNKNKPPE SRKKLEKA
Thiverval _aa	…………………………………………… T
Bangkhen_aa	………………………………………………

The amino acid sequences of C were shown in Table 3. The 128 residues molecules were translated from nucleotide sequences in Table 2. The results showed good conservation and high homology. The amino acid sequence of C differ in 5 residues between Bangkhen and Chinese strain, and 6 residues between Bangkhen and Thiverval strain.

As shown in Table 4, nucleotide sequences of Bangkhen E2 gene were compared with Thiverval (Genbank: EU490425) and Chinese strain (Genbank: AY805221). The results of E2 gene showed their homology as high as 94% - 96% was found among these stains. In addition, we found *Eco*RI restriction site at residues 244 – 249 (GAATTC) of Bangkhen E2 gene.

Table 4 Alignment of E2 gene of Bangkhen strain compared with Thiverval and Chinese strain.

```
             ...............|............20................|............40...............|...........  60................|..............80
1 Chinese    GCACAAGGCC GGCTAGCCTG CAAGGAAGAT TACAGGTACG CAATATCGTC AACCGATGAG ATAGGGCTAC TTGGGGCCGG
2 Thiverval  GCACAAGGCC AGCTAGCCTG CAAGGAAGAT TACAGGTACG CAATATCATC AACCAATGAG ATAGGGCTAC TCGGGGCCGG
3 Bangkhen   GCACAAGGCC GGCTAGCCTG TAAGGAAGAT TACAGGTATG CAATATCATC AACTAATGAG ATAGGGCTAC TCGGGGCCGG

             ...............|............100................|.........  120...............|............140................|............160
1 Chinese    AGGTCTCACC ACCACCTGGA AGGAATACAA CCACGATTTG CAACTGAATG ACGGGACCGT CAAGGCCAGT TGCGTGGCAG
2 Thiverval  AGGTCTCACC ACCACCTGGA AAGAATACAA CCACGATTTG CAACTGAATG ACGGGACCGT TAAGGCCATT TGCGTGGCAG
3 Bangkhen   GGGTCTCACT ACCACCTGGA AAGAACACAG ACACGATTTG CAACTGAATG ACGGGACCGT TAAGGCCATT TGCGCGGCAG

             ...............|............180................|..........200...............|...........220................|............240
1 Chinese    GTTCCTTTAA AGTCACAGCA CTTAATGTGG TCAGTAGGAG GTATTTGGCG TCATTGCATA AGAAGGCTTT ACCCACTTCC
2 Thiverval  GTTCCTTTAA AGTCACAGCA CTTAATGTGG TCAGTAGGAG GTATTTGGCA TCATTGCATA AGGAGGCTTT ACCCACTTCC
3 Bangkhen   GTTCCTTTAA AGTCACAGCA CTTAATGTGG TCAGTAGGAG GTATTTGGCG TCATTGCATA AAAAGGCTTT ACCCACTTCA

             ...............|............260................|............280...............|.............300................|............320
1 Chinese    GTGACATTCG AGCTCCTGTT CGACGGGACC AACCCATCAA CTGAGGAAAT GGGAGATGAC TTCAGGTCCG GGCTGTGCCC
2 Thiverval  GTGGCATTCG AGCTCCTGTT CGACGGGACC AACCCATCAA CTGAGGAAAT GGGAGATGAC TTCGGGTTCG GGCAGTGCCC
3 Bangkhen   GTGGAATTCG AGCTCCTGTT CGACGGGACC AACCCATCAA CTGAGGAAAT GGGAGATGAC TTCGGGTTCG GGCTGTGCCC

             ...............|...........340................|............360...............|...........380..............|............400
1 Chinese    GTTTGATACG AGTCCTGTTG TTAAGGGAAA GTACAATACG ACCTTGTTGA ACGGTAGTGC TTTCTATCTT GTCTGCCCAA
2 Thiverval  GTTTGATACG AGTCCTGTTG TCAAGGGAAA GTACAATACA ACCTTGTTGA ACGGTAGTGC TTTCTATCTT GTCTGTCCAA
3 Bangkhen   GTTTGATACG AGTCCTGTCG TCAAGGGAAA GTACAATACA ACCTTGTTGA ACGGTAGTGC TTTCTATCTT GTCTGCCCAA

             ...............|............420................|............440...............|.............460................|............480
1 Chinese    TAGGGTGGAC GGGTGTCATA GAGTGCACAG CAGTGAGCCC AACAACTCTG AGGACAGAAG TGGTAAAGAC CTTCAGGAGA
2 Thiverval  TAGGGTGGAC GGGTGTTATA GAGTGCACAG CAGTGAGCCC AACAACTCTG AGAACAGAAG TGGTAAAGAC CTTCAGGAGG
3 Bangkhen   TAGGGTGGAC GGGTGTTATA GAGTGCACGG CAGTGAGCCC AACAACTCTG AGAACAGAAG TGGTAAAGAC CTTCAGGAGA

             ...............|............500................|............520...............|.............540................|............560
1 Chinese    GACAAGCCCT TTCCGCACAG AATGGATTGT GTGACCACCA CAGTGGAAAA TGAAGATTTA TTCTATTGTA AGTTGGGGGG
2 Thiverval  GACAAGCCCT TTCCGCACAG AATGTTTTGT GTGACCACAA CAGTGGAAAA TGAAGATTTA TTCTACTGTA AGTTGGGGGG
3 Bangkhen   GAGAAGCCCT TTCCGCACAG AATGGATTGT GTGACCACAA CAGTGGAAAA TGAAGATTTA TTCTACTGTA AGTTGGGGGG

             ...............|............580................|............600...............|.............620................|............640
1 Chinese    CAACTGGACA TGTGTGAAAG GCGAGCCAGT GGTCTACACA GGGGGGGTAG TAAAACAATG TAGATGGTGT GGCTTCGACT
2 Thiverval  CAACTGGACA TGTGTGAAAG GTGAACCAGT GGTCTACACG GGGGGGCTAG TAAAACAATG CAGATGGTGT GGCTTTGACT
3 Bangkhen   CAACTGGACA TGTGTGAAAG GTGAACCAGT GGTTTACACG GGGGGGCTGG TAAAACAATG CAGATGGTGT GGCTTTGACT

             ...............|............660................|............680...............|.............700................|............720
1 Chinese    TCGATGGGCC TGACGGACTC CCGCATTACC CCATAGGTAA GTGCATTTTG GCAAATGAGA CAGGTTACAG AATAGTAGAT
2 Thiverval  TCAATGAGCC TGACGGACTC CCACACTACC CCATAGGTAA GTGCATTTTG GCAAATGAGA CAGGTTACAG AATAGTGGAT
3 Bangkhen   TCAAGGAGCC TGACGGACTC CCGCACTACC CTATAGGTAA GTGCATTTTG GCAAATGAGA CAGGTTACAG AATAGTAGAT

             ...............|............740................|............760...............|.............780................|............800
1 Chinese    TCAACGGACT GTAACAGAGA TGGCGTTGTA ATCAGCACAG AGGGGAGTCA TGAGTGCTTG ATCGGTAACA CGACTGTCAA
2 Thiverval  TCAACAGACT GTAACAGAGA TGGTGTTGTA ATCAGCACAG AGGGGAGTCA TGAGTGCTTG ATCGGTAACA CAACTGTCAA
3 Bangkhen   TCAACGGACT GCAACAGGGA TGGTGTTATA ATCAGCACAG ATGGGAGTCA TGAGTGCTTG ATCGGTAACA CCACTGTCAA

             ...............|............820................|............840...............|.............860................|............880
1 Chinese    GGTGCATGCA TCAGATGAAA GACTGGGCCC TATGCCATGC AGACCTAAAG AGATTGTCTC TAGTGCTGGT CCTGTAATGA
2 Thiverval  GGTGCATGCA TCAGATGAAA GACTGGGCCC CATGCCATGC AGACCTAAAG AGATCGTCTC TAGTGCAGGA CCTGTAAGGA
3 Bangkhen   GGTGCATGCA TCAGATGAAA GACTGGGCCC TATGCCGTGC AGACCTAAAG AGATTGTCTC TAGTGCTGGT CCTGTAAGGA

             ...............|............900................|............920...............|.............940................|............960
1 Chinese    AAACCTCCTG TACATTCAAC TACACAAAAA CTTTGAAGAA CAGGTACTAT GAGCCCAGGG ACAGCTACTT CCAGCAATAT
2 Thiverval  AAACTTCCTG TACATTCAAC TACGCAAAAA CTTTGAAGAA CAAGTACTAT GAGCCCAGGG ACAGCTACTT CCAGCAATAT
3 Bangkhen   AAACTTCCTG TACATTCAAC TACGCAAAAA CTTTGAAGAA CAAGTACTAT GAGCCCAGGG ACAGCTACTT CCAGCAATAT

             ...............|............980................|...........1000...............|............1020................|...........1040
1 Chinese    ATGCTTAAGG GTGAGTATCA GTACTGGTTT GACCTGGATG CGACTGACCG CCACTCAGAT TACTTCGCAG AATTT
2 Thiverval  ATGCTTAAGG GCGAGTATCA GTACTGGTTT GACCTGGACG TGACTGACCG CCACTCAGAT TACTTCGCAG AATTT
3 Bangkhen   ATGCTTAAGG GTGAGTATCA GTACTGGTTT GACCTGGACG TGACTGACCG CCACTCAGAT TACTTCGCAG AATTT

             Sequences (Chinese: Thiverval)      Aligned. Score:  95
             Sequences (Chinese: Bangkhen)      Aligned. Score:  94
             Sequences (Thiverval: Bangkhen) Aligned. Score:  96
```

Table 5 Alignment of E2 amino acid sequence of Bangkhen strain compared with Thiverval and Chinese strain.

	……………10………..20…………...30……………40……………50
Chinese_aa	AQGRLACKE DYRYAISSTD EIGLLGAGGL TTTWKEYNHD LQLNDGTVKAS
Thiverval _aa	……Q………………………N…………………………………….……..…………… I
Bangkhen_aa	…………………………....……………………….……HR…………………… I
	…………..60…………...70…………..80….……….90…………..100
Chinese_aa	CVAGSFKVTA LNVVSRRYLA SLHKKALPTS VTFELLFDGT NPSTEEMGDD
Thiverval _aa	……………………………… ……........E…………A…………………………..
Bangkhen_aa	..……………………… ..………...…….E……....….E………………………………
	………..110…………..120……..…..130…………..140…………..150
Chinese_aa	FRSGLCPFDT SPVVKGKYNT TLLNGSAFYL VCPIGWTGVI ECTAVSPTTL
Thiverval _aa	..GF..Q…………………………………………………………………………
Bangkhen_aa	..GF.. L……………………………………………………………………………
	…………160…………..170…………..180…………..190…………..200
Chinese_aa	RTEVVKTFRR DKPFPHRMDC VTTTVENEDL FYCKLGGNWT CVKGEPVVYT
Thiverval _aa	……………………..………F……………………………………….………….
Bangkhen_aa	………………...E……………………………………………………….………………..
	…………..210……..…..220………...230…………..240…………..250
Chinese_aa	GGVVKQCRWC GFDFDGPDGL PHYPIGKCIL ANETGYRIVD STDCNRDGVV
Thiverval _aa	…..L……………………...E………………………………………………………..
Bangkhen_aa	…..L……………………...E………………………………………………………I
	………260…………..270…………..280……...….290…………..300
Chinese_aa	ISTEGSHECL IGNTTVKVHA SDERLGPMPC RPKEIVSSAG PVMKTSCTFN
Thiverval _aa	……………………………………………………………………R…………
Bangkhen_aa	…..D……………………………………………...……………R…………
	…………310….……..320……..……330…………..340…………..350
Chinese_aa	YTKTLKNRYY EPRDSYFQQY MLKGEYQYWF DLDATDRHSD YFAEF
Thiverval _aa	...A……….K…………………………………………………………………
Bangkhen_aa	...A……….K…………………………………………………………………

The 345 residues molecules of amino acid sequence of E2 were translated from nucleotide sequence in Table 4. The amino acid sequence of E2 gene differs in 16 residues between Bangkhen and Chinese strain, and 10 residues between Bangkhen and Thiverval strain.

4. Prediction of 3D structure protein

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web. The purpose of this server is to make protein modeling accessible to all biochemists and molecular biologists. In this work it was used to find 3D structure protein. The results of prediction showed no suitable templates found. The server was unable to create 3D structure model of CSFV- C and CSFV-E2 protein.

5. Prediction of N-Linked Glycosylation

N-linked glycosylation was a type of linkage that is mainly formed between sugar molecules in biochemistry. This type of linkage was shared between sugar molecules and a specific sequence of protein (amino acids). It was important for the folding of some proteins. The amino acid sequences of CSFV-C and CSFV-E2 were translated. N-Glycosylation sited of both proteins were predicted using NetNGlyc 1.0 server. Sequence having N- glycosylation potential > 0.5 was considered as cut-off value. Figure 8-10 showed the prediction of N-Linked Glycosylation of CSFV - C. We found N- glycosylation site at position 35-NFTD in Chinese and thiverval strain. In Bangkhen strain N- glycosylation site at position 35-NLTD was predicted.

For CSFV-E2 protein, there were 7 positions of the N-linked glycosylation in Chinese and thiverval strain. However, there were only 6 positions that predicted in Bangkhen strain. N-linked glycosylation site at position 37-NHDL can not be found in this strain.

Name: C of Chinese strain **Length: 128**

GSDGKLYHTYVCIDGCILLKLAKRGEPRTLKWIRNFTDCP 40
LWVTSCSDDGASGSKEKKPDRINKGKLKIAPKEHEKDSRT 80
RPPDATIVVEGVKYQVKKKGKVKGKNTQDGLYHNKNKPPE 120
SRKKLEKA

(Threshold=0.5)

SeqName	**Position**	**Potential**	**Jury agreement**	**N-Glyc result**
Sequence	35 NFTD	0.5758	(8/9)	+

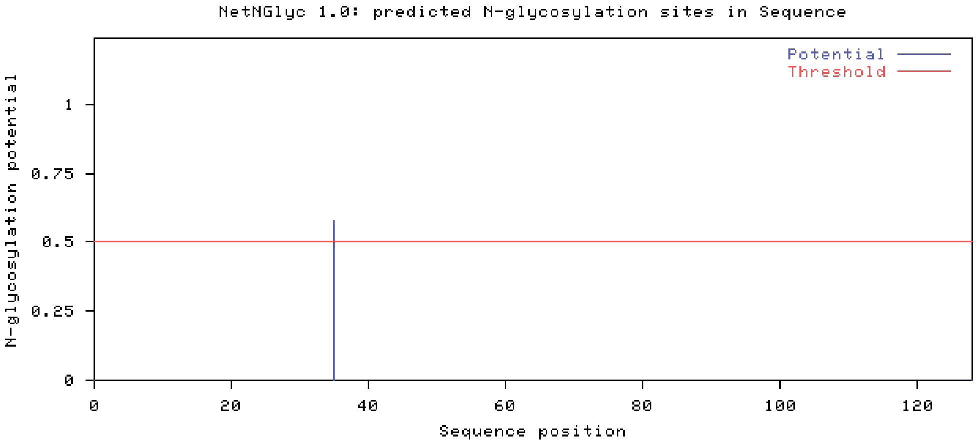

Figure 8 N-glycosylation prediction of CSFV - C of Chinese strain.

Name: C of Thiverval strain **Length: 128**

GSDGRLYHIYVCIDGCILLKLAKRGEPRTLKWIRNFTDCP 40
LWVTSCSDDGASGSKEKKPDRINKGKLKIAPKEHEKDSRT 80
KPPDATIVVEGVKYQVRKKGKVKGKNTQDGLYHNKNKPPE 120
SRKKLEKT

(Threshold=0.5)

SeqName	**Position**	**Potential**	**Jury agreement**	**N-Glyc result**
Sequence	35 NFTD	0.5760	(8/9)	+

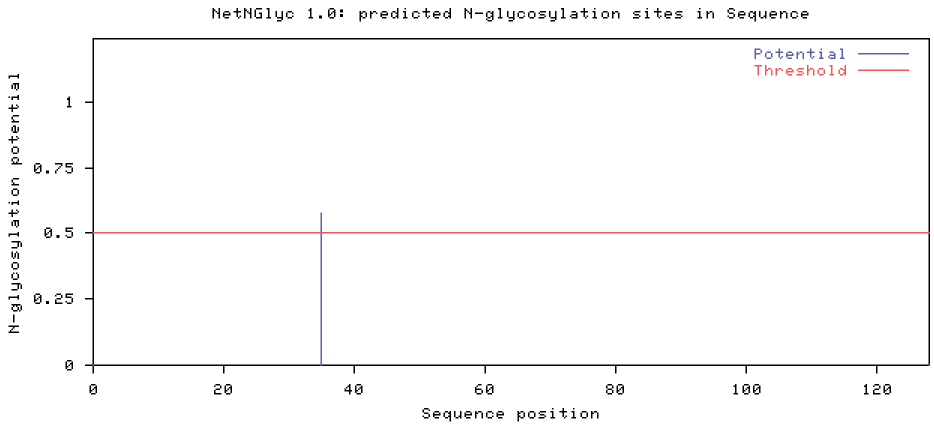

Figure 9 N-glycosylation prediction of CSFV – C of Thiverval strain.

Name: C of Bangkhen strain Length: 128

GSDGKLYHIYVCIDGCILLKLAKRDEPRTLKWIRNLTDCP 40
LWVTSCSDDGASGSKEKKPDRISKGKLKIAPKEHEKDSRT 80
KPPDATIVVEGVKYQVKKKGKVKGKNTQDGLYHNKNKPPE 120
SRKKLEKA

(Threshold=0.5)

SeqName	**Position**	**Potential**	**Jury agreement**	**N-Glyc result**
Sequence	35 NLTD	0.6408	(9/9)	++

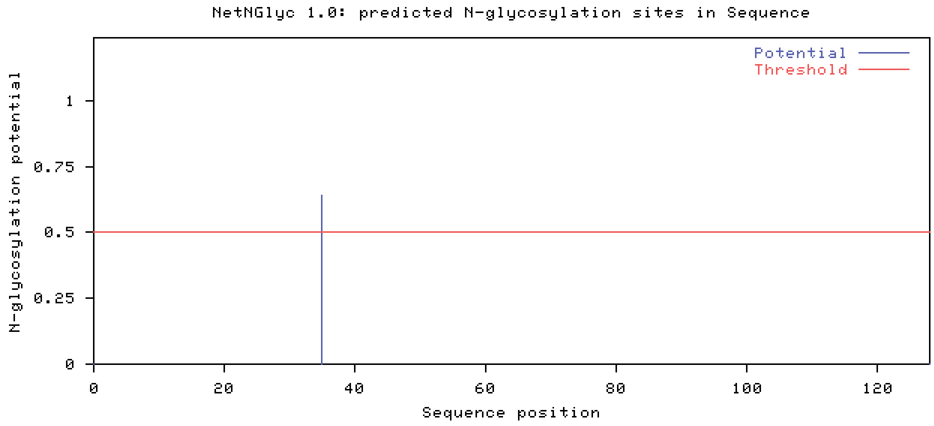

Figure 10 N-glycosylation prediction of CSFV - C of Bangkhen strain.

Name E2 of Chinese **Length: 345**

AQGRLACKEDYRYAISSTDEIGLLGAGGLTTTWKEYNHDL 40
QLNDGTVKASCVAGSFKVTALNVVSRRYLASLHKKALPTS 80
VTFELLFDGTNPSTEEMGDDFRSGLCPFDTSPVVKGKYNT 120
TLLNGSAFYLVCPIGWTGVIECTAVSPTTLRTEVVKTFRR 160
DKPFPHRMDCVTTTVENEDLFYCKLGGNWTCVKGEPVVYT 200
GGVVKQCRWCGFDFDGPDGLPHYPIGKCILANETGYRIVD 240
STDCNRDGVVISTEGSHECLIGNTTVKVHASDERLGPMPC 280
RPKEIVSSAGPVMKTSCTFNYTKTLKNRYYEPRDSYFQQY 320
MLKGEYQYWFDLDATDRHSDYFAEF 360
(Threshold=0.5)

SeqName	Position	Potential	Jury agreement	N-Glyc result
Sequence	37 NHDL	0.7121	(9/9)	++
Sequence	43 NDGT	0.6112	(9/9)	++
Sequence	62 NVVS	0.8001	(9/9)	+++
Sequence	91 NPST	0.5069	(3/9)	+ WARNING: PRO-X1. SEQUON ASN-XAA-SER/THR.
Sequence	119 NTTL	0.6991	(9/9)	++ SEQUON ASN-XAA-SER/THR.
Sequence	124 NGSA	0.7157	(9/9)	++ SEQUON ASN-XAA-SER/THR.
Sequence	177 NEDL	0.5866	(7/9)	+
Sequence	188 NWTC	0.6298	(9/9)	++ SEQUON ASN-XAA-SER/THR.
Sequence	232 NETG	0.6383	(9/9)	++ SEQUON ASN-XAA-SER/THR.
Sequence	245 NRDG	0.4961	(5/9)	-
Sequence	263 NTTV	0.6244	(7/9)	+ SEQUON ASN-XAA-SER/THR.
Sequence	300 NYTK	0.5428	(6/9)	+ SEQUON ASN-XAA-SER/THR.
Sequence	307 NRYY	0.4460	(7/9)	-

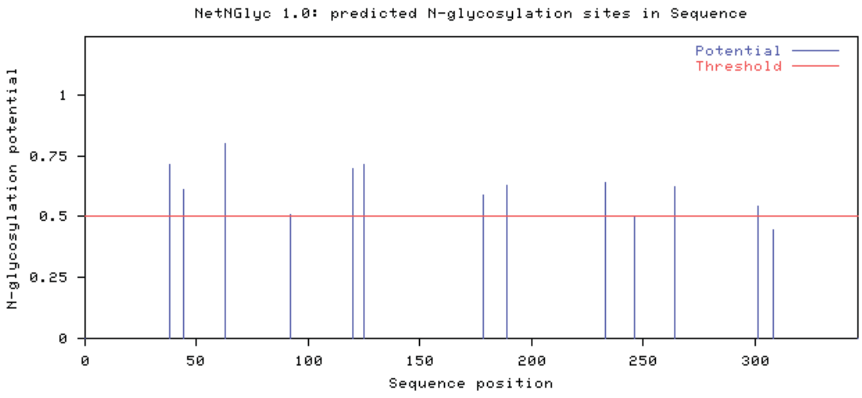

Figure 11 N-glycosylation prediction of CSFV – E2 of Chinese strain.

Name: E2 of Thiverval Length: 345

AQGQLACKEDYRYAISSTNEIGLLGAGGLTTTWKEYNHDL 40
QLNDGTVKAICVAGSFKVTALNVVSRRYLASLHKEALPTS 80
VAFELLFDGTNPSTEEMGDDFGFGQCPFDTSPVVKGKYNT 120
TLLNGSAFYLVCPIGWTGVIECTAVSPTTLRTEVVKTFRR 160
DKPFPHRMFCVTTTVENEDLFYCKLGGNWTCVKGEPVVYT 200
GGLVKQCRWCGFDFNEPDGLPHYPIGKCILANETGYRIVD 240
STDCNRDGVVISTEGSHECLIGNTTVKVHASDERLGPMPC 280
RPKEIVSSAGPVRKTSCTFNYAKTLKNKYYEPRDSYFQQY 320
MLKGEYQYWFDLDVTDRHSDYFAEF 360
(Threshold=0.5)

SeqName	**Position**	**Potential**	**Jury agreement**	**N-Glyc result**
Sequence	19 NEIG	0.4948	(5/9)	-
Sequence	37 NHDL	0.7121	(9/9)	++
Sequence	43 NDGT	0.6071	(9/9)	++
Sequence	62 NVVS	0.8001	(9/9)	+++
Sequence	91 NPST	0.5184	(4/9)	+ WARNING: PRO-X1. SEQUON ASN-XAA-SER/THR.
Sequence	119 NTTL	0.6992	(9/9)	++ SEQUON ASN-XAA-SER/THR.
Sequence	124 NGSA	0.7159	(9/9)	++ SEQUON ASN-XAA-SER/THR.
Sequence	177 NEDL	0.5846	(7/9)	+
Sequence	188 NWTC	0.6297	(9/9)	++
Sequence	215 NEPD	0.4580	(6/9)	-
Sequence	232 NETG	0.6382	(9/9)	++ SEQUON ASN-XAA-SER/THR.
Sequence	245 NRDG	0.4961	(5/9)	-
Sequence	263 NTTV	0.6244	(7/9)	++ SEQUON ASN-XAA-SER/THR.
Sequence	300 NYAK	0.5343	(6/9)	+
Sequence	307 NKYY	0.4774	(6/9)	-

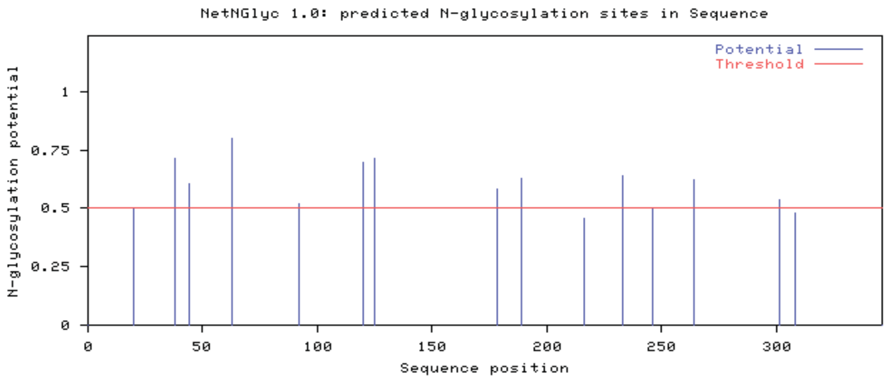

Figure12 N-glycosylation prediction of CSFV – E2 of Thiverval strain.

```
Name:  E2 of Bangkhen                    Length:  345
AQGRLACKEDYRYAISSTNEIGLLGAGGLTTTWKEHRHDL        40
QLNDGTVKAICAAGSFKVTALNVVSRRYLASLHKKALPTS        80
VEFELLFDGTNPSTEEMGDDFGFGLCPFDTSPVVKGKYNT       120
TLLNGSAFYLVCPIGWTGVIECTAVSPTTLRTEVVKTFRR       160
EKPFPHRMDCVTTTVENEDLFYCKLGGNWTCVKGEPVVYT       200
GGLVKQCRWCGFDFKEPDGLPHYPIGKCILANETGYRIVD       240
STDCNRDGVIISTDGSHECLIGNTTVKVHASDERLGPMPC       280
RPKEIVSSAGPVRKTSCTFNYAKTLKNKYYEPRDSYFQQY       320
MLKGEYQYWFDLDVTDRHSDYFAEF                      360
 (Threshold=0.5)
------------------------------------------------------------------------------------
SeqName     Position    Potential       Jury             N-Glyc
                                      agreement          result
------------------------------------------------------------------------------------
Sequence    19 NEIG     0.4946          (5/9)              -
Sequence    43 NDGT     0.6370          (9/9)              ++
Sequence    62 NVVS     0.8004          (9/9)              +++
Sequence    91 NPST     0.5113          (3/9)              +     WARNING:
                                             PRO-X1. SEQUON ASN-XAA-SER/THR.
Sequence    119 NTTL    0.6992          (9/9)              ++
                                               SEQUON ASN-XAA-SER/THR.
Sequence    124 NGSA    0.7157          (9/9)              ++
                                               SEQUON ASN-XAA-SER/THR.
Sequence    177 NEDL    0.5868          (7/9)              +
Sequence    188 NWTC    0.6297          (9/9)              ++
                                               SEQUON ASN-XAA-SER/THR.
Sequence    232 NETG    0.6381          (9/9)              ++
                                               SEQUON ASN-XAA-SER/THR.
Sequence    245 NRDG    0.4646          (5/9)              -
Sequence    263 NTTV    0.6244          (7/9)              +
                                               SEQUON ASN-XAA-SER/THR.
Sequence    300 NYAK    0.5343          (6/9)              +
Sequence    307 NKYY    0.4773          (6/9)              -
```

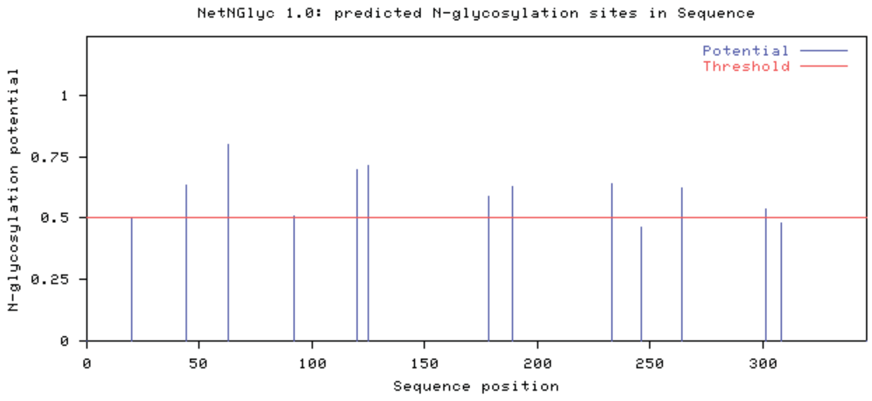

Figure13 N-glycosylation prediction of CSFV – E2 of Bangkhen strain.

6. Prediction of epitope

The prediction of epitopes has been done with IEDB Analysis Resource Server. Emini Surface accessibility scale method predicts epitope site (Figure14 and 15) The epitope predictions for CSFV - C were given in Table 6. The similar epitope peptide regions CSFV - C were 54-SKEKKPDR-61, 72-KEHEKDSRTRPP-83 and 95-QVKKKG-100, however we found epitope site 22- AKRDEPR-28 in Bangkhen strain.

Table 6 Predicted epitope peptides of CSFV – C.

CSFV - C	Start Position	End Position	Peptide	Peptide Length
Chinese	54	61	SKEKKPDR	8
	72	83	KEHEKDSRTRPP	12
	95	100	QVKKKG	6
Thiverval	54	61	SKEKKPDR	8
	72	83	KEHEKDSRTKPP	12
	95	100	QVRKKG	6
Bangkhen	22	28	AKRDEPR	7
	54	61	SKEKKPDR	8
	72	83	KEHEKDSRTKPP	12
	95	100	QVKKKG	6

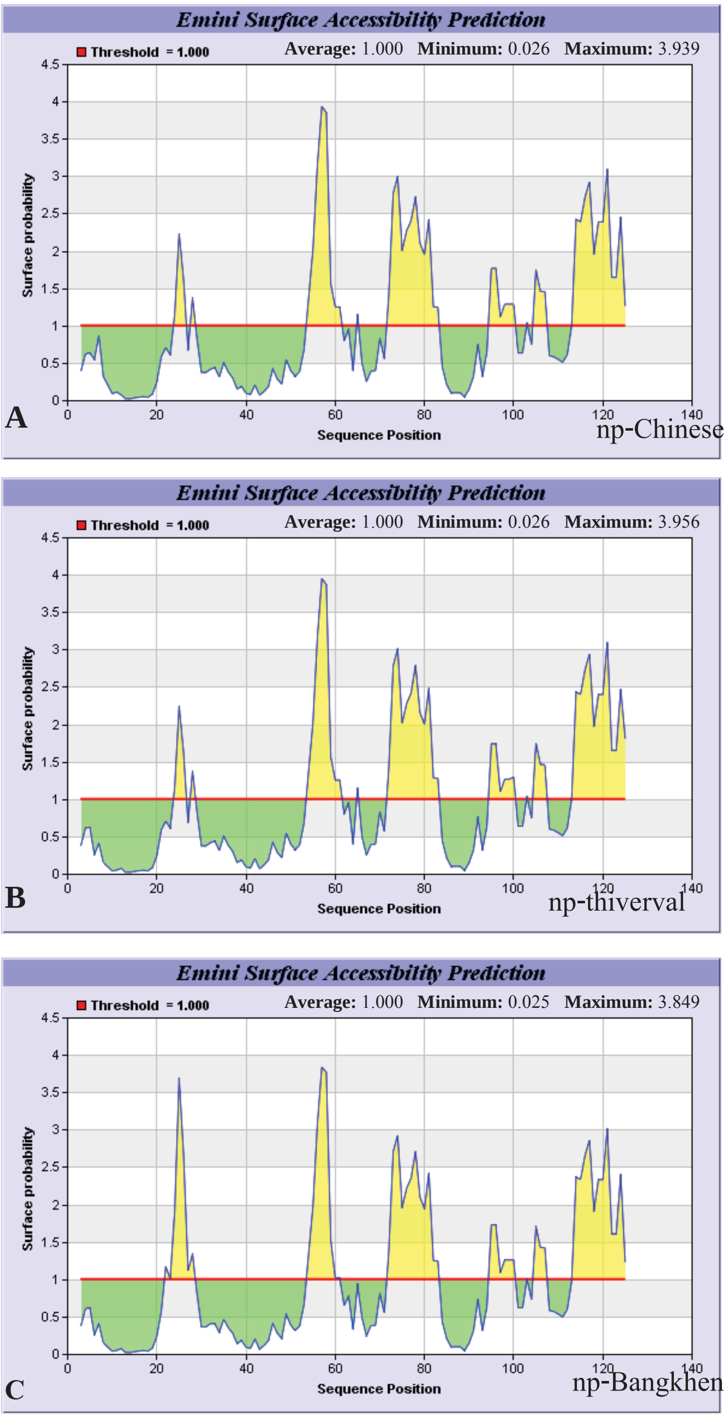

Figure14 Emini surface accessibility prediction plots of CSFV- C.

As shown in Table 7, there were 6 epitope peptide regions of CSFV-E2 proteins in each strain. The similar peptide regions were 31- TTWKEYNH-38, 89-GTNPSTEEM-97, 157-TFRRDKPFPH-166, 174-TVENED-179, 300- NYTKTLKN RYYEPRDSYFQ-318 and 333- DATDRHS-339. These epitopes can be a milestone for vaccine and antidotes design against their respective viruses.

Table 7 Predicted epitope peptides of CSFV-E2.

CSFV-E2	**Start Position**	**End Position**	**Peptide**	**Peptide Length**
Chinese	31	38	TTWKEYNH	8
	89	97	GTNPSTEEM	9
	157	166	TFRRDKPFPH	10
	174	179	TVENED	6
	300	318	NYTKTLKNRYYEPRDSYFQ	19
	333	339	DATDRHS	7
Thiverval	31	38	TTWKEYNH	8
	89	97	GTNPSTEEM	9
	157	165	TFRRDKPFP	9
	174	179	TVENED	6
	300	318	NYAKTLKNKYYEPRDSYFQ	19
	334	339	VTDRHS	6
Bangkhen	31	39	TTWKEHRHD	9
	89	97	GTNPSTEEM	9
	157	166	TFRREKPFPH	10
	174	179	TVENED	6
	300	318	NYAKTLKNKYYEPRDSYFQ	19
	334	339	VTDRHS	6

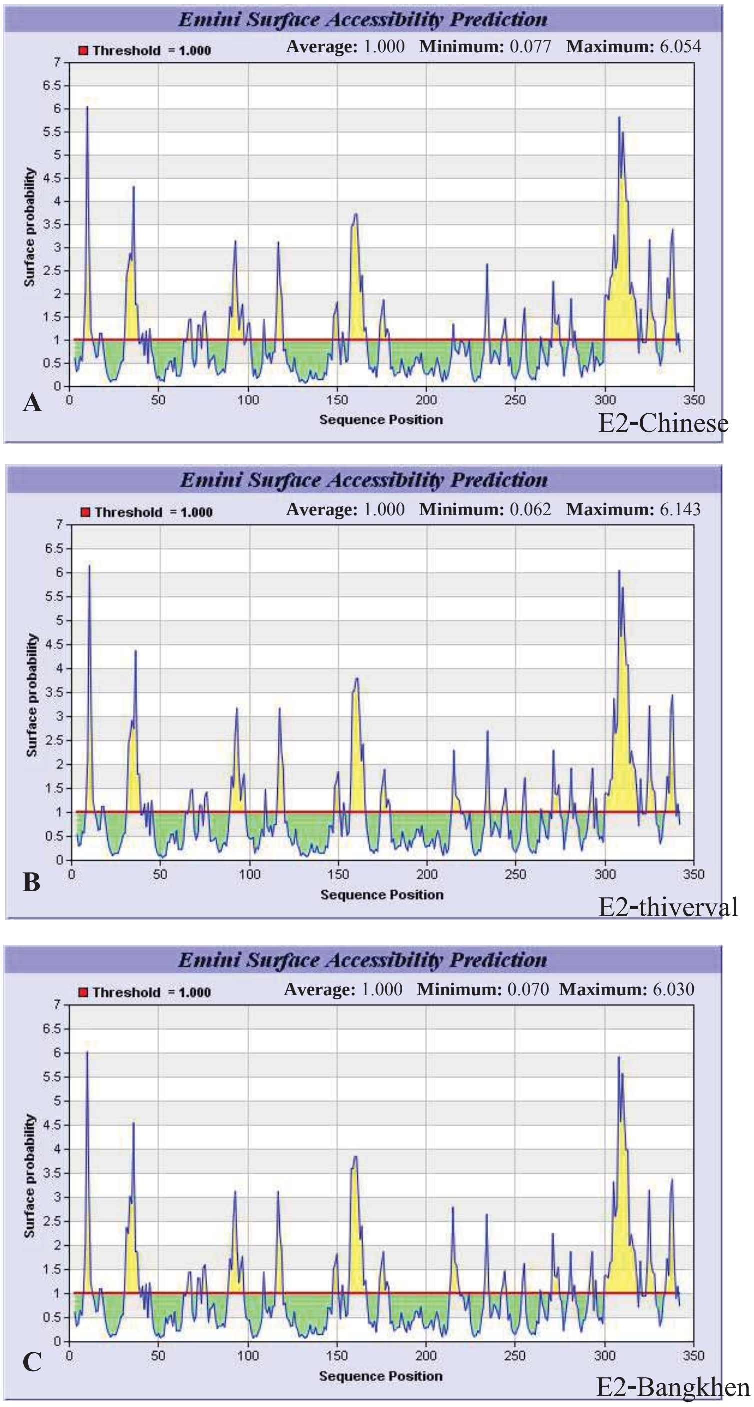

Figure15 Emini surface accessibility prediction plots of CSFV-E2.

7. Construction of recombinant plasmid for protein expression.

For protein expression, nucleic acid sequence of CSFV - C gene and CSFV-E2 gene were considered. Early part of genes without start codon(ATG) and the end of genes without stop codon (UAA, UGA, UAG).(Figure 16, Figure 17)

```
GGTAGTGACG  GAAGGCTTTA  CCATATATAT  GTGTGCATCG
ATGGTTGCAT  ACTGCTGAAG  CTAGCCAAAA  GGGGCGAGCC
AAGAACCCTG  AAGTGGATTA  GAAATTTCAC  CGACTGTCCA
TTGTGGGTTA  CCAGTTGCTC  TGATGATGGC  GCAAGTGGAA
GTAAAGAGAA  GAAGCCAGAT  AGGATCAACA  AGGGTAAATT
AAAAATAGCC  CCGAAAGAGC  ATGAGAAGGA  CAGCAGAACT
AAGCCACCTG  ACGCTACGAT  CGTAGTGGAA  GGAGTAAAAT
ACCAGGTCAG  AAAGAAAGGT  AAAGTTAAAG  GAAAGAATAC
CCAAGACGGC  CTGTACCACA  ATAAGAATAA  ACCACCAGAA
TCTAGGAAGA  AATTAGAAAA  AGCC
```

Figure 16 The 384 bp of CSFV - C gene. (The genome position 740 – 1123 were obtained from GenBank, National Center of Biotechnology Information (NCBI), U.S.A: EU490425).

```
GCACAAGGCC  AGCTAGCCTG  CAAGGAAGAT  TACAGGTACG
CAATATCATC  AACCAATGAG  ATAGGGCTAC  TCGGGGCCGG
AGGTCTCACC  ACCACCTGGA  AAGAATACAA  CCACGATTTG
CAACTGAATG  ACGGGACCGT  TAAGGCCATT  TGCGTGGCAG
GTTCCTTTAA  AGTCACAGCA  CTTAATGTGG  TCAGTAGGAG
GTATTTGGCA  TCATTGCATA  AGGAGGCTTT  ACCCACTTCC
GTGGCATTCG  AGCTCCTGTT  CGACGGGACC  AACCCATCAA
CTGAGGAAAT  GGGAGATGAC  TTCGGGTTCG  GGCAGTGCCC
GTTTGATACG  AGTCCTGTTG  TCAAGGGAAA  GTACAATACA
ACCTTGTTGA  ACGGTAGTGC  TTTCTATCTT  GTCTGTCCAA
TAGGGTGGAC  GGGTGTTATA  GAGTGCACAG  CAGTGAGCCC
AACAACTCTG  AGAACAGAAG  TGGTAAAGAC  CTTCAGGAGG
GACAAGCCCT  TTCCGCACAG  AATGTTTTGT  GTGACCACAA
CAGTGGAAAA  TGAAGATTTA  TTCTACTGTA  AGTTGGGGGG
CAACTGGACA  TGTGTGAAAG  GTGAACCAGT  GGTCTACACG
GGGGGGCTAG  TAAAACAATG  CAGATGGTGT  GGCTTTGACT
TCAATGAGCC  TGACGGACTC  CCACACTACC  CCATAGGTAA
GTGCATTTTG  GCAAATGAGA  CAGGTTACAG  AATAGTGGAT
TCAACAGACT  GTAACAGAGA  TGGTGTTGTA  ATCAGCACAG
AGGGGAGTCA  TGAGTGCTTG  ATCGGTAACA  CAACTGTCAA
GGTGCATGCA  TCAGATGAAA  GACTGGGCCC  CATGCCATGC
AGACCTAAAG  AGATCGTCTC  TAGTGCAGGA  CCTGTAAGGA
AAACTTCCTG  TACATTCAAC  TACGCAAAAA  CTTTGAAGAA
CAAGTACTAT  GAGCCCAGGG  ACAGCTACTT  CCAGCAATAT
ATGCTTAAGG  GCGA
```

Figure 17 The 1035 bp of CSFV –E2 gene. (The genome position 2432 – 3466 were obtained from GenBank, National Center of Biotechnology Information (NCBI), U.S.A: EU490425).

In this study, the primers for protein expression were designed again on the basis of highly conserved region of early parts and the end of genes. Primer CAP1 and CAP2 were used to amplify CSFV - C gene and primer EPR1 and EPR2 were used to amplify CSFV - E2 gene. The forward primer CAP1 and EPR1 contained a sequence for start codon (ATG) and *Nde*I restriction site, while the reverse primer CAP2 and EPR2 contained a sequence for stop codon (TAG) and *Xho*I restriction site. (Table 8)

Table 8 CSFV-C and CSFV-E2 primers designed for protein expresstion.

Primer name	Sequence (5′ - 3′)	Length	Tm(50mM Na^{+})	%GC
CAP 1	ATACATATGGGTAGTGACGGAAAGCTGTACC	31	64.69	45
CAP 2	CTGCTCGAGCTAGGCTTTTTCTAATTTCTTCCTAG	35	65.59	43
EPR 1	ATACATATGGCACAAGGCCAGCTAGCCTG	29	66.27	52
EPR 2	CTGCTCGAGCTAAAATTCTGCGAAGTAATCTG	32	64.61	44

The nucleocapsid and E2 gene of CSFV were amplified by PCR with new primers (E2 gene: EPR1, EPR2 primer and nucleocapsid gene : CAP1 , CAP2 primer). PCR reaction mixture was prepared with 2x PCR master mix (Promega , USA) , each primer pairs , DNA template and water. These primers were annealing at 53˚ C for 1 minute in PCR thermal cycle step. The results showed about 1000 bp for E2 gene in lane 2 but can not found about 400 bp for nucleocapsid gene in lane 1. In addition, it showed non - specific bands at below 250 bp of both PCR products. (Figure 18) It was indicated that, temperature of 53˚ C should not be used for annealing step of these genes amplification.

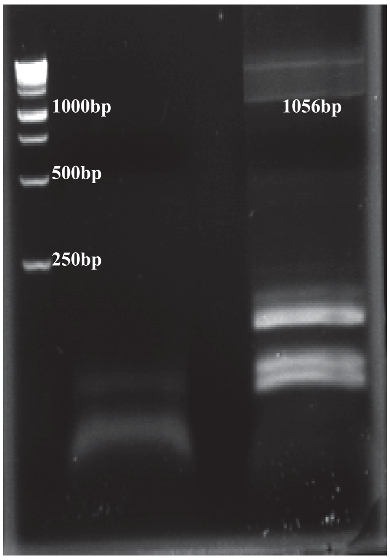

Figure 18 PCR amplification of C or E2 gene with new primers (annealing step at 53˚ C), lane 1 : CSFV – C and lane 2: CSFV – E2 gene.

Thus, the temperature of annealing step was changed. The PCR products were amplified by other temperature in this step (60˚C or 65˚ C). In figure 19 (annealing step at 60˚ C), the results showed that poor quality band at 400 bp of CSFV – C gene and 1000 bp of CSFV – E2 gene , while figure 20 (annealing step at 65˚ C), non - specific bands at below 250 bp were found.

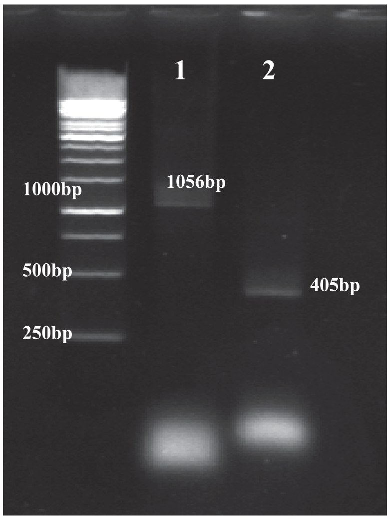

Figure 19 PCR amplification of C or E2 gene with new primers (annealing step at 60˚ C), lane 1 : CSFV – E2 gene and lane 2: CSFV – C gene.

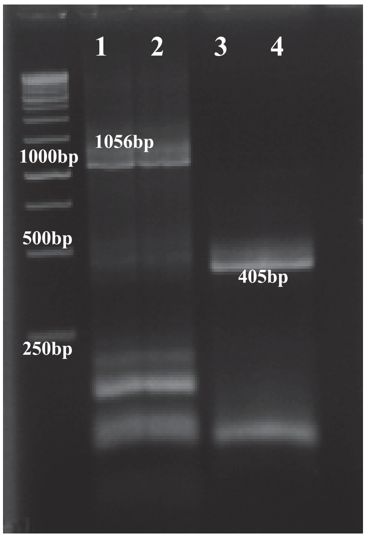

Figure 20 PCR amplification of C or E2 gene with new primers (annealing step at 65˚ C), lane 1-2: CSFV – E2 gene and lane 3-4: CSFV – C gene.

The expected DNA fragments were extracted from agarose gel. The blunt C and E2 gene fragment were ligated into multi-cloning site of pTZ57R/T vector at ratio vector: gene was 1:3. The ligation mixture was transformed into *E. coli* DH5α. The cells were spread on LB/ ampicillin plates and incubated overnight at 37˚ C. The 10 – 30 single colony were picked out randomly and were tested with PCR for detect recombinant plasmid. The resulted suggested that random colony had no E2 - pTZ57R/t and C – pTZ57R/t recombinant plasmid in cells (Figure 21).

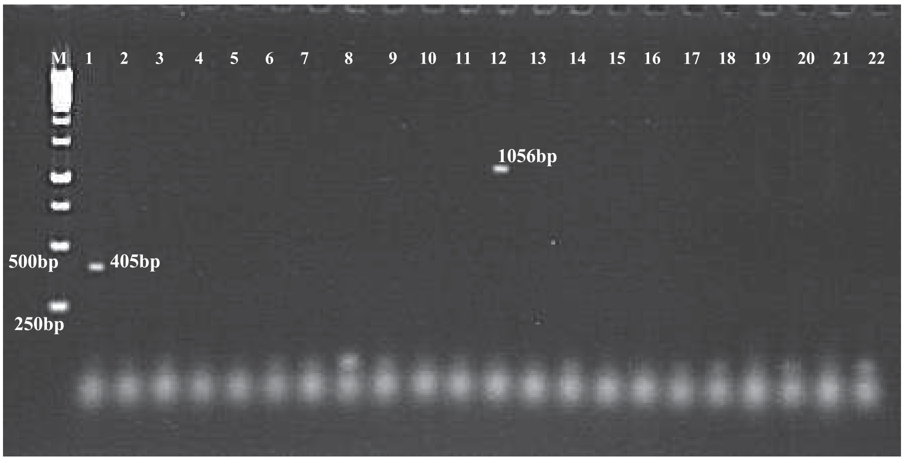

Figure 21 PCR amplification from random colony , lane 1 : E2 gene (+control), lane2-11 E2 colony PCR, lane 12 : C gene (+control) and lane 13-22 C gene colony PCR.

CHAPTER V

DISCUSSION

Gene cloning and sequencing were considered to be the foundation of molecular biology and genetic engineering. In order to study the infectious CSFV found in Thailand, it is necessary to analyze E2 and C genes of CSFV. Envelope glycoprotein E2 is the most immunogenic and most variable protein of pestivirus. It represents an important target for induction of the immune response during infection. E2 is the major envelope glycoprotein exposed on the surface of the virion. It is essential for virus attachment and entry into the host cells as well as cell tropism (88, 131). This glycoprotein has been implicated as one of the virulent determinants (85, 87). In addition, it can induce neutralizing antibodies and confer protective immunity in pigs (5, 10, 25, 46, 125, 132).

In many viral diseases, those C proteins of RNA or DNA viruses are important targets of humoral and cellular immune response to virus infection. The primary function of viral C proteins is to provide a protective shell for the viral genome inside the virion (63). In addition, the C protein may also play an important role in the pathogenesis of viral infection (62).

In consideration of the differences between infectious CSFV Bangkhen strain found in Thailand and other strains. E2 and C of Bangkhen strain had been amplified using one step RT-PCR. RT-PCR amplification products were eletrophoresed in 1% agarose gel (Figure 4). A good quality amplification without any non – specific products was obtained in both cases. The amplified products corresponded to the gene length, the gel indicated the approximate presence of 400 and 1000 bp product for C and E2 gene, respectively.

Both genes were cloned into a cloning vector pGEM T-Easy vector. The pGEM T-Easy vector is a PCR product direct cloning small vector of 3,015 bp. It contains the linearized vector with a free Thymine at the 5′ end. It can be used to obtain a large amount of plasmid DNA generating many copies when cloning. The cloning site of this vector is integrated inside the *lacZ* gene allowing selection of recombined molecules by IPTG (Isopropyl β-D-1- thiogalactopyranoside) and X-Gal (bromo-chloro-indolyl-galactopyranoside) which contained ampicilin resistance to the transformed cells.

The recombinant of CSFV-Bangkhen E2 was digested by *Eco*RI -*Xho*I, while other strains were not digested. The *Eco*RI restriction site at residues 244 – 249 (GAATTC) of CSFV-Bangkhen E2 gene was found. The results revealed that the E2 gene of Bangkhen strain had been different due to a presence of a variable region which was not found in other strains. The nucleotide sequences of C gene similarity were 93% between the Bangkhen strain and the Chinese vaccine strain. The similarity between the Bangkhen and the Thiverval strain was 94%. These results were in concordance with previous study by Liu *et al* (1998). The identity of the C sequences in many strains varied from 84 to 95 %. The amino acid sequences of C translated from the nucleotide sequences in Table 2 showed high conservation and only 5-6 residues difference in amino acid among these strains. The amino acid sequences of C similarity were 97% between Bangkhen strain and other strains. From the results in this study, C genes were clear conservation. In addition, alignment of nucleotide sequence of E2 gene similarity was 94 - 96% and the amino acid sequence of E2 gene similarity was 95 - 97% between the Bangkhen and other strain.

N-linked glycosylation refers to the attachment of oligosaccharides to a nitrogen atom, usually the asparagine residues. N-glycosylation occurs on secreted or membrane bound proteins. The consensus sequence for N-glycosylation was Asn-Xaa-Ser/Thr (52). The sites of N-glycosylation was identified by the 'NetNGlyc' predictor (http://www.cbs.dtu.dk/services/ NetNGlyc) using artificial neural networks that examined the sequence context of Asn-Xaa-Ser/Thr sequences. Sequences having N-glycosylation potential >0.5 are considered as cut-off value. The method is described in detail in the following article "Prediction of N-glycosylation sites in human proteins" (110). In general, glycosylation of enveloped virus structural proteins has been shown to be important for receptor binding, membrane fusion, virus budding, infectivity, cell-cell interactions and antigenicity (1, 30). Classical swine fever virus outer surface E2 glycoprotein represents an important target to induce protective immunization during infection, although the influence of N-glycosylation pattern in antigenicity is yet unclear (82). Predicted E2 glycosylation sites are highly conserved among CSFV isolates (99, 100). E2 glycoproteins had been associated with CSFV virulence (71, 85). The removal of putative glycosylation sites in the E2 glycoprotein would affect viral virulence/pathogenesis in swine. Glycosylation of E2 could be modified for development of live attenuated classical swine fever vaccines (87). N-glycosylation sites within CSFV E2 had been predicted previously (73). According to a glycosylation analysis algorithm (http://www.cbs.dtu.dk/services/), E2 of the CSFV strain Brescia has five putative N-linked sites but not confirmed by experimental evidence. A sixth N-linked glycosylation site was presented in several strains. In this study, the server can predict only one N-linked glycosylation site in C protein and many N-linked glycosylation site in E2 protein. The E2 protein of Bangkhen strain, the N-glycosylation site 37-NHDL was missing.

An epitope, also known as antigenic determinant, is the part of an antigen that is recognized by the immune system. The part of an antibody that recognizes the epitope is called a paratope. Epitope prediction such as the prediction of protein regions recognized specific antibody is a topical problem (24). To improve the quality of protein alignment, Emilio *et al* (1985) developed a method for comparing two sequences on the basis of the specific features of their surface using the scale that characterized the probability of amino acid surface accessibility (37). The neutralizing epitopes corresponding to different regions of E2 were proposed and used as vaccines against CSFV in either mono- or multi-peptide formulations (31, 57,140). In hepatitis C virus infection, it had been shown that epitopes of C protein had been important for the T- and B-cell mediated immune responses. (41). Xin *et al* (2011) identified a linear epitope recognized by the mAb, corresponding to amino acids 107-TQDGLYHNKN-116 of the C protein (139). This results were in concordance with Figure 14 in this study. In different strains, the epitope prediction by IEDB Analysis server was similar. Six epitope peptide regions in E2 and three in C proteins were found. However, CSFV-C protein Bangkhen strain was found 4 epitope peptide regions. The peptide region of 22-AKRDEPR-28 was predicted.

For protein expression, E2 - pTZ5RT or C - pTZ5RT recombinant plasmid were not successed because the quality of PCR products amplified by new primers were poor. It was contaminanted with non – specific amplification products or primer

– dimmer. The E2 fragment was longer than 1 kb, thus the efficiency of DNA fragment cloning was poor.

CHAPTER VI
CONCLUSION

Classical swine fever is a highly contagious viral disease of swine in many areas of the world including Thailand. In this study, E2 and C gene of CSFV Bangkhen strain were amplified by using one step RT-PCR and ligated into cloning vector pGEM® T-Easy vector. The results showed that E2 gene of Bangkhen strain was different from other strains because it could be found *Eco*RI restriction site, while other strains had no restriction site for this enzymes. Additionally, the nucleotides sequence of Bangkhen C gene showed clear conservation and high homology when compared with Thiverval and Chinese strain. For N-linked glycosylation prediction, the one potential N-linked glycosylation site of nucleocapsid was predicted. The E2 of Bangkhen strain showed 6 N-linked glycosylation sites, while the E2 of Thiverval or Chinese strain showed 7 N-linked glycosylation sites. Moreover, The C gene of Bangkhen strain showed 4 epitope sites whereas that of Thiverval or Chinese strain showed only 3 epitope sites.

REFERENCES

1. Abe YE, Takashita K, Sugawara Y, Matsuzaki Y, Muraki, Hongo S. Effect of the addition of oligosaccharides on the biological activities and antigenicity of influenza A/H3N2 virus hemagglutinin. J Virol 2004; 78: 9605– 9611.
2. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics 2006; 22: 195-201.
3. Bauhofer O, Summerfield A, McCullough KC, Ruggli N. Role of double-stranded RNA and N-pro of classical swine fever virus in the activation of monocyte-derived dendritic cells. Virology 2005; 343: 93-105.
4. Bauhofer O, Summerfield A, Sakoda Y, Tratschin JD, Hofmann MA, Ruggli N. Classical swine fever virus N-pro interacts with interferon regulatory factor 3 and induces its proteasomal degradation. J Virol 2007; 81: 3087-3096.
5. Beer M, Reimann I, Hoffmann B, Depner K. Novel marker vaccines against classical swine fever. Vaccine 2007; 25: 5665-5670.
6. Biront P, Leunen. Classical swine fever and related viral infections. Vaccine 1988; 181–197.
7. Blacksell SD, Khounsy S, Westbury HA. The effect of sample degradation and RNA stabilization on classical swine fever virus RT-PCR and ELISA methods. J Virol Method. 2004; 118: 33-37.
8. Blom N, Sicheritz PT, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 2004; 4: 1633-1649.
9. Boom R, Sol CJ, Salimans MM, Jansen CL, Werth-van Dillen PM, van der Noordaa J. Rapid and sample method for purification of nucleic acids. J Clin Microbiol 1990; 28: 495-503.

10. Bouma A, de Smit AJ, de Kluijver EP, Terpstra C, Moormann RJ. Efficacy and stability of a subunit vaccine based on glycoprotein E2 of classical swine fever virus. Vet Microbiol 1999; 66: 101-114.

11. Bouma A, de Smit AJ, de Jong MC, De Kluijver EP, Moormann RJ. Determination of the onset of the herd-immunity induced by the E2 subunit vaccine against classical swine fever virus. Vaccine 2000; 18: 1374–1381.

12. Chang CY, Huang CC, Lin YJ, Deng MC, Chen HC, Tsai CH, Chang WM, Wang FI. Antigenic domains analysis of classical swine fever virus E2 glycoprotein by mutagenesis and conformation-dependent monoclonal antibodies. Virus Res 2010; 149: 183-189.

13. Charleston B, Fray MD, Baigent S, Carr BV, Morrison WI. Establishment of persistent infection with non-cytopathic bovine viral diarrhoea virus in cattle is associated with a failure to induce type I interferon. J Gen Virol 2001; 82: 1893–1897.

14. Chen N, Hu H, Zhang Z, Shuai J, Jiang L, Fang W. Genetic diversity of the envelope glycoprotein E2 of classical swine fever virus: recent isolates branched away from historical and vaccine strains. Vet Microbiol 2008; 127: 286-299.

15. Chen N, Li D, Yuan X, Li X, Hu H, Zhu B, Wan X, Fang W. Genetic characterization of E2 gene of classical swine fever virus by restriction fragment length polymorphism and phylogenetic analysis. Virus Genes 2010a; 40: 389-396.

16. Chen N, Tong C, Li D, Wan J, Yuan X, Li X, Peng J, Fang W. Antigenic analysis of classical swine fever virus E2 glycoprotein using pig antibodies identifies residues contributing to antigenic variation of the vaccine C-strain and group 2 strains circulating in China. J Virol 2010b; 7: 378-391.

17. Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 1979; 18: 5294-5299.

18. Chomczynski P, Sacchi N. Single - step method of RNA isolation by acid guanidinium thicyanate-phenol-chloroform extraction. Anal Biochem 1987; 162: 156-159.

19. Christlet THT, Veluraja K. Database analysis of O-glycosylation sites in proteins. Biophys J 2001; 80: 952-960.

20, Clavijo A, Zhou EM, Vydelingum S, Heckert R. Development and evaluation of a novel antigen capture assay for the detection of classical swine fever virus antigens. Vet Microbiol 1998; 60: 155-168.

21. Clavijo A, Lin M, Riva J, Mallory M, Lin F, Zhou EM. Development of a competitive ELISA using a truncated E2 recombinant protein as antigen for detection of antibodies to classical swine fever virus. Res Vet Sci 2001; 70: 1–7.

22. Damrongwatanapokin S, Pinyochon W, Parchariyanon S, Inui K. Classical swine fever in Thailand. Vientiane, Lao PDR 1999; 109-110.

23. Dangre DM, Deshmukh SR, Rathod DP, Umare VD, Ullah I. Prediction and comparative analysis of MHC binding peptides and epitopes in Nanoviridae nano-organisms. J Proteomics Bioinform 2010; 3: 155-172.

24. Davydov YI, Tonevitsky AG. Prediction of linear B-cell epitopes. Biology 2009; 43: 150–158.

25. de Smit AJ, Bouma A, de Kluijver EP, Terpstra C, Moormann RJ. Duration of the protection of an E2 subunit marker vaccine against classical swine fever after a single vaccination. Vet Microbiol 2001; 78: 307-317.

26. Depner KR, Hinrich U, Bickhardt K, Greiser WI, Pohlenz, Moenning V, Liess B. Influence of breed-related factor on the course of classical swine fever virus infection. Vet Rec 1997; 140: 506-507.

27. Dewulf J, Laevens H, Koenen F, Mintiens K, de Kruif A. Airborne transmission of classical swine fever virus under experimental conditions. Vet Rec 2000; 147:735-738.

28. Dewulf J, Laevens H, Koenen F, Mintiens K, de Kruif A. Evaluation of the potential of dogs, cats and rats to spread classical swine fever virus. Vet Rec 2001; 149: 212-213.

29. Diaz de Arce H, Nunez JI, Ganges L, Barreras M, Frias MT, Sobrino F. An RT-PCR assay for the specific detection of classical swine fever virus in clinical samples. Vet Res 1998; 29: 431-440.

30. Doms RW, Lamb RA, Rose JK, Helenius A. Folding and assembly of viral membrane proteins. Virology 1993; 193: 545–562.

31. Dong XN, Wei K, Liu ZQ, Chen YH. Candidate peptide vaccine induced protection against classical swine fever virus. Vaccine 2002; 21**:** 167-173.

32. Dong MH, Kai XQ, Shen GF, Zhang ZF , Yi NL, Su ZL, Shao HB. Recombination and expression of classical swine fever virus (CSFV) structural protein E2 gene in *Chlamydomonas reinhardtii* chroloplasts. Biointerfaces 2007; 55: 26–30.

33. Edwards S, Sands JJ. Antigenic comparisons of hog cholera virus isolates from Europe, America and Asia using monoclonal antibodies. Dtsch Tierarztl Wochenschr 1990; 97: 79-81.

34. Edwards S, Hoennig V, Wensvoort G. The development of an international reference panel of monoclonal antibodies for the differentiation of hog cholera virus from other pestiviruses. Vet Microbiol 1991; 29: 101-8.

35. Edwards S, Fukusho A, Lefèvre PC, Lipowski A, Pejsak Z, Roehe P, Westergaard J. Classical swine fever: the global situation. Vet Microbiol 2000; 73: 103-19.

36. Elbers ARW, Stegeman A, Moser H, Ekker HM, Smak JA, Pluimers FH. The classical swine fever epidemic 1997 – 1998 in the Netherlands: descriptive epidemiology. Preventive Veterinary Medicine 1999; 42: 157-184.

37. Emilio AE, Joseph VH, Debra SP, Joshua B. Induction of Hepatitis A virus-neutralizing antibody by a virus- specific synthetic peptide.Virology 1985; 55: 836-839

38. Floegel NG, Bunzenthal C, Fischer S, Moennig V. Virulence of recent and former classical swine fever virus isolates evaluated by their clinical

and pathological signs. J Vet Med B Infect Dis Vet Public Health 2003; 50: 214-220.

39. Francki RIB, Faquet CM, Knudson DL, Brown F. Flaviviridae. Arch Virol Suppl 1991; 2: 223-233.

40. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis 1997; 18: 2714-2723.

41. Handchuh G, Caselman WH. Bacterial expression and purification of hepatitis C virus capsid protein of different size. J Hepatol 1995; 22: 143-150.

42. Hart GW. Glycosylation. Cell Biol 1992; 4: 1017-1023.

43. He DM, Qian KX, Shen GF, Zhang ZF, LI YN, Su ZL, et al. Recombinant and expression of classical swine fever virus (CSFV) structural protein E2 gene in *Chlamydomonas reinhardtii* chroloplasts. Colloids and Surfaces B 2007; 55: 26-30.

44. Heimann M, Roman SG, Martoglio B, Thiel HJ, Rümenapf T. Core protein of pestiviruses is processed at the C terminus by signal peptide peptidase. J Virol 2006; 80: 1915–1921.

45. Hoffmann B, Beer M, Schelp C, Schirrmeier H, Depner K. Validation of a real-time RT-PCR assay for sensitive and specific detection of classical swine fever. J Virol 2005; 130: 36-44.

46. Hulst MM, Westra DF, Wensvoort G, Moormann RJ. Glycoprotein E1 of hog cholera virus expressed in insect cells protects swine from hog cholera. J Virol 1993;67: 5435–5442.

47. Janin J, Wodak S, Levitt M, Maigret B. Conformation of amino acid side-chains in proteins. J Mol Biol 1978; 125: 357-386.

48. Jeffrey KP, Liang SM, Milan SB, Sharon M, Joseph YT. Production of *Haemophilus influenzae* type-b porin in *Escherichia coli* and its folding into the trimeric form. Gene 1995; 152: 85-88.

49. Katz JB, Ridpath JF, Bolin SR. Presumptive diagnostic differentiation of hog cholera virus from bovine viral diarrhea and border disease viruses by

using a cDNA nested-amplification approach. J Clin Microbiol 1993; 31: 565-568.

50. Kerur N, Jhala MK, Joshi CG. Genetic characterization of Indian peste des petits ruminants virus (PPRV) by sequencing and phylogenetic analysis of fusion protein and nucleoprotein gene segments. Res Vet Sci 2008; 85: 176–183.

51. Konig M, Lengsfeld T, Pauly T, Stark R, Thiel HJ. Classical swine fever virus: independent induction of protective immunity by two structural glycoproteins. J Virol 1995; 69: 6479–6486.

52. Kornfeld R, Kornfeld S. Assembly of asparagine-linked oligosaccharides. Biochem 1985; 54: 631–664.

53. Kosmidou A, Ahl R, Thiel HJ, Weiland E. Differentiation of classical swine fever virus (CSFV) strains using monoclonal antibodies against structural glycoproteins. Vet Microbiol 1995; 47: 111–118.

54. Kotz A, Wagener J, Engel J, Routier FH, Echtenacher B. Approaching the secrets of N-glycosylation in *Aspergillus fumigatus*: characterization of the AfOch1 protein. PLoS ONE 2010; 5: e15729.

55. Lin M, Lin F, Mallory M, Clavijo. A deletions of structural glycoprotein E2 of classical swine fever virus strain Alfort/187 resolve a linear epitope of monoclonal antibody WH303 and the minimal N-terminal domain essential for binding immunoglobulin G antibodies of a pig hyperimmune serum. J Virol 2000; 74: 11619–11625.

56. Liu JJ, Wong ML, Chen PF, Chang TJ. Cloning, expression and sequence analysis of the classical swine fever virus nucleocapsid protein. Virus Genes 1998; 16: 225-34.

57. Liu SG, Tu CC, Wang CL, Yu XL, Wu JM, Guo SP, Shao ML, Gong Q, Zhu QH, Kong XG. The protective immune response induced by B cell epitope of classical swine fever virus glycoprotein E2. J Virol Meth 2006; 134: 125-129.

58. Lowings JP, Paton DJ, Sands JJ, De Mia GM, Rutili D. Classical swine fever: genetic detection and analysis of differences between virus isolates. J Gen Virol 1994; 75: 3461-3468.

59. Lowings JP, Ibata G, Needham J, Paton D. Classical swine fever virus diversity and evolution. J Gen Virol 1996; 77: 1311–21.

60. Luka Z, Wagner C. Expression and purification of glycine N-methyltransferases in *Escherichia coli*. Protein Expression and Purification 2003; 28: 280–286.

61. Mahadtanapuk S, Sanguansermsri M, Yu LD, Vilaithong T, Anuntalabhochai S. Cloning of antifungal gene from *Bacillus licheniformis* by application of low energy ion beam bombardment. Surface & Coatings Tech 2009; 203: 2546–2549

62. Major ME, Vitvitski L, Mink MA, Schleef M, Whalen RG, Trépo C, Inchauspé G. DNA-based immunization with chimeric vectors for the induction of immune responses against the hepatitis C virus nucleocapsid. J Virol 1995; 69:5798–5805.

63. Matsumoto M, Hwang SB, Jeng KS, Zhu N, Lai MM. Homotypic interaction and multimerization of hepatitis C virus core protein. Virology. 1996; 218: 43–51.

64. Mayer D, Hofmann MA, Tratschin JD. Attenuation of classical swine fever virus by deletion of the viral N-pro gene. Vaccine 2004; 22: 317-328.

65. Mcgoldrick A, Lowings JP, Ibata G, Sands JJ, Belak S, Paton DJ. A novel approach to the detection of classical swine fever virus by RT-PCR with a fluorogenic probe (TaqMan). J Virol 1998; 72: 125-135.

66. Mendoza S, Correa GP, Aguilera E, Colmenares G, Torres O, Cruz T, Romero A, Hernandez BE, Ciprian A. Antigenic differentiation of classical swine fever vaccinal strain PAV-250 from other strains, including field strains from Mexico. Vaccine 2007; 25: 7120-7124.

67. Meyers G, Rumenapf T, Thiel HJ. Molecular cloning and nucleotide sequence of the genome of hog cholera virus. Virology 1989; 171: 555–567.

68. Meyers G, Theil HJ, Rumenapf T. Classical swine fever virus: Recovery of infectious viruses from cDNA constructs and generation of recombinant cytopathogenic defective interfering particles. J Virol 1996a; 50: 1588-1595.

69. Meyers G, Thiel HJ. Molecular characterization of pestiviruses. Adv Virus Res 1996b; 47: 53-118.

70. Moenning V, Plagemann PGW. The pestiviruses. Adv Virus res 1992; 41: 53-98.

71. Moennig V. Introduction to classical swine fever virus, disease and control policy. Vet Microbiol 2000; 73: 93-102.

72. Moennig V, Floegel NG, Greiser WI. Clinical signs and epidemiology of classical swine fever: a review of new knowledge. Vet J 2003; 165: 11-20.

73. Moormann RJM, Warmerdam PA, van der Meer B, Hulst MM. Nucleotide sequence of hog cholera virus RNA: properties of the polyprotein encoded by the open reading frame spanning the viral genomic RNA. Vet Microbiol 1990; 23: 185-191.

74. Moser C, Ruggli N, Duri TJ, Hofmann MA. Detection of antibodies against classical swine fever virus in swine sera by indirect ELISA using recombinant envelope glycoprotein E2. Vet Microbiol 1996; 51: 41-53.

75. Muller A, Depner KR, Liess B. Evolution of a gp 55 (E2) recombinant-based ELISA for the detection of antibodies induced by classical swine fever virus. Deutsche Tierartzliche Wochenschrift 1996; 103: 451-453.

76. Naresh KCVM, Sangamithra P, Rajasekhar M, Saigopal DVR. Surveillance of chikungunya virus in Andhra Pradesh, Southern India. Asian Pacific Journal of Tropical Medicine 2010; 860-865

77. Nishimori T, Yamada S, Shimizu M. Production of monoclonal antibodies against classical swine fever virus and their use for antigenic characterization of Japanese isolates. J Vet Med Sci 1996; 58: 707-710.

78. Parchariyanon S, Inui K, Damrongwatanapokin S, Pinyochon W, Lowings P, Paton D. Sequence analysis of E2 glycoprotein genes of classical swine fever viruses: identification of a novel genogroup in Thailand. Dtsch Tierarztl Wochenschr 2000; 107: 236-238.

79. Paton DJ, Simpson V, Done SH. Infection of pigs and cattle with bovine diarrhea virus on a farm in England. Vet Rec 1992; 131: 185-8.

80. Paton DJ, McGoldrick A, Greiser-Wilke I, Parchariyanon S, Song JY, Liou PP, Stadejek T, Lowings JP, Björklund H, Belák S. Genetic typing of classical swine fever virus. Vet Microbiol 2000; 73: 137-57.

81. Pereda AJ, Greiser WI, Schmitt B, Rincon MA, Mogollon JD, Sabogal ZY, Lora AM, Sanguinetti H, Piccone ME. Phylogenetic analysis of classical swine fever virus (CSFV) field isolates from outbreak in South and Central America. Virus Res 2005; 110: 111-118.

82. Raquel M, Jeovanis G, Luis JG, Yasser Z, Louise R, Pauline MR, Raymond AD, David JH, Jose AC. The N-glycosylation of classical swine fever virus E2 glycoprotein extracellular domain expressed in the milk of goat. Arch Biochem Biophys 2010; 500: 169-180.

83. Ribbens S, Dewulf J, Koenen F, Laevens H, de Kruif A. Transmission of classical swine fever. A review. Vet Q 2004; 26: 146-155.

84. Rice CM. *Flaviviridae*: the viruses and their replication. Fields virol 1996; 931- 959.

85. Risatti GR, Borca MV, Kutish GF, Lu Z, Holinka LG, French RA, Tulman ER, Rock DL. The E2 glycoprotein of classical swine fever virus is a virulence determinant in swine. J Virol 2005; 79: 3787-3796.

86. Risatti GR, Holinka LG, Fernandez SI, Carrillo C, Lu Z, Borca MV. N-linked glycosylation status of classical swine fever virus strain Brescia E2 glycoprotein influences virulence in swine. J Virol 2007a; 81: 924–933.

87. Risatti GR, Holinka LG, Fernandez Sainz I, Carrillo C, Kutish GF, Lu Z, Zhu J, Rock DL, Borca MV. Mutations in the carboxyl terminal region of E2 glycoprotein of classical swine fever virus are responsible for viral attenuation in swine. Virol 2007b; 364: 371-382.

88. Reimann I, Depner K, Trapp S, Beer M. An avirulent chimeric Pestivirus with altered cell tropism protects pigs against lethal infection with classical swine fever virus. Virol 2004; 322: 143–157.

89. Robert SG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology 2002; 12: 43-56.

90. Ruggli N, Bird BH, Liu L, Bauhofer O, Tratschin JD, Hofmann MA. N-pro of classical swine fever virus is an antagonist of double-stranded RNA-mediated apoptosis and IFN-alpha/beta induction. Virology 2005; 340: 265-276.

91. Ruggli N, Summerfield A, Fiebach AR, Guzylack PL, Bauhofer O, Lamm CG, Waltersperger S, Matsuno K, Liu L, Gerber M, Choi KH, Hofmann MA, Sakoda Y, Tratschin JD. Classical swine fever virus can remain virulent after specific elimination of the interferon regulatory factor 3-degrading function of N-pro. J Virol 2009; 83: 817-829.

92. Rümenapf T, Stark R, Meyers G, Thiel HJ. Structural proteins of hog cholera virus expressed by vaccinia virus: further characterization and induction of protective immunity. J Virol 1991; 65: 589-97.

93. Rümenapf T, Unger G, Strauss JH, Thiel HJ. Processing of the envelope glycoproteins of pestiviruses. J Virol 1993; 67: 3288-94.

94. Rümenapf T, Stark R, Heimann M, Thiel HJ. N-terminal protease of pestiviruses: identification of putative catalytic residues by sitedirected mutagenesis. J Virol 1998; 72: 2544–2547.

95. Saatkamp HW, Berentsen PBM, Horst HS. Economic aspects of the control of classical swine fever outbreaks in the Europran Union. Vet Microbiol 2000; 73: 221-237.

96. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research 2003; 31: 3381-3385.

97. Seago J, Hilton L, Reid E, Doceul V, Jeyatheesan J, Moganeradj K, McCauley J, Charleston B, Goodbourn S. The Npro product of classical swine fever virus and bovine viral diarrhea virus uses a conserved mechanism to target interferon regulatory factor-3. J Gen Virol 2007; 88: 3002-3006.

98. Seitz O. Synthesis and the effects of glycosylation on protein structure and activity. BioChem 2000; 1: 214-246.

99. Shi X, Elliott RM. Analysis of N-linked glycosylation of Hantaan virus glycoproteins and the role of oligosaccharide side chains in protein folding and intracellular trafficking. J Virol 2004; 78: 5414–5422.

100. Shi X, Brauburger K, Elliott RM. Role of N-linked glycans on Bunyamwera virus glycoproteins in intracellular trafficking, protein folding and virus infectivity. J Virol 2005; 79: 13725–13734.

101. Shiladitya C, Parikshit B, Dipanjan D, Anupam M, Nobumichi K, Mamta C. Computational identification of the post-translational modification sites and the functional family prediction reveals possible moonlighting role of rotaviral proteins. Bioinformation 2010; 10: 448-451.

102. Shin MK, Cha SB, Lee WJ, Yoo HS. Predicting genetic traits and epitope analysis of apxIVA in *Actinobacillus pleuropneumoniae*. J Microbiol 2011; 49: 462-468.

103. Shiu JS, Chang MH, Liu ST, Ho WC, Lai SS, Chang TJ, Chang YS. Molecular cloning and nucleotide sequence determination of three envelope genes of classical swine fever virus Taiwan isolate p97. Virus Res 1996; 41: 173-178.

104. Silva Jr FP, Ribeirob F, Katz N, Giovanni SS. Exploring the subsite specificity of *Schistosoma mansoni* aspartyl hemoglobinase through comparative molecular modeling. FEBS Letters 2002; 514: 141-148

105. Singh VK, Kumar GS, Paliwal OP. Detection of classical swine fever virus in archival formalin-fixed tissues by reverse transcription-polymerase chain reaction. Veterinary Science 2005; 79: 81-84.

106. Stadejek T, Vilcek S, Lowings JP, Pordany AB, Paton DJ, Belak S. Genetic heterogeneity of classical swine fever virus in Central Europe. Virus Res 1997; 52: 195–204.

107. Stark R, Rümenapf T, Meyers G, Thiel HJ. Genomic localization of hog cholera virus glycoproteins. Virology 1990; 174: 286-289.

108. Stark R, Meyers G, Rümenapf T, Thiel HJ. Processing of pestivirus polyprotein: cleavage site between autoprotease and nucleocapsid protein of classical swine fever virus. J Virol 1993; 67: 7088-7095.

109. Summerfield A, Zingle K, Inumaru S, McCullough KC. Induction of apoptosis in bone marrow neutrophil-lineage cells by classical swine fever virus. J Gen Virol 2001; 82: 1309–1318.

110. Sun S, Wang Q, Zhao F, Chen W, Li Z. Prediction of Biological Functions on Glycosylation Site Migrations in Human Influenza H1N1 Viruses. PLoS ONE 2012; 7: e32119.

111. Suradhat S, Intrakamhaeng M, Damrongwatanapokin S. The correlation of virus-specific interferon-gamma production and protection against classical swine fever virus infection. Vet Immunol 2001; 83: 177-189.

112. Suradhat S, Damrongwatanapokin S. The influence of maternal immunity on the efficacy of a classical swine fever vaccine against classical swine fever virus, genogroup 2.2, infection. Vet Microbiol 2003; 92: 187–94.

113. Terpstra C, Wensvoort G. Natural infection of pig with bovine viral diarrhea virus associated with signs resembling swine fever. Res Vet Sci 1988; 45: 137-42

114. Terpstra C. Infectious diseases of livestock, with special reference to southern Africa. Oxford University Press 1994; 1: 654-657.

115. Thiel HJ, Stark R, Weiland E, Rümenapf T, Meyers G. Hog cholera virus: molecular composition of virions from a pestivirus. J Virol1991; 65: 4705-4712.

116. Thiel HJ, Collett MS, Gould EA, Heinz FX, Houghton M, Meyers G, Purcell RH, Rice CM. Family Flaviviridae In: Virus Taxonomy, eighth ed. Elsevier Academic Press 2005; 981–998.

117. Thompson JD, Higgins DG, Gibson TJ. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.". Nucleic Acids Res 1994; 22: 4673-4680.

118. Tian H, Liu XT, Wu JY, Shang YJ, Jiang T, Zheng HX, Xie QG. Expression of major antigen domains of E2 gene of CSFV and analysisof its immunological activity. Vilologica Sinica 2008; 23: 247-254.

119. van Gennip HG, van Rijn PA, Widjojoatmodjo MN, de Smit AJ, Moormann RJ. Chimeric classical swine fever viruses containing envelope protein E(RNS) or E2 of bovine viral diarrhoea virus protect pigs against challenge with CSFV and induce a distinguishable antibody response. Vaccine 2000; 19: 447–459.

120. van Gennip HG, Bouma A, van Rijn PA, Widjojoatmodjo MN, Moormann RJ. Experimental non-transmissible marker vaccines for classical swine fever (CSF) by trans-complementation of E(rns) or E2 of CSFV. Vaccine 2002; 20: 1544–1556.

121. van Rijn PA, Thiel HJ, Stark R, Weiland E, Rumenapf T, Meyers G. Hog cholera virus: molecular composition of virions from a pestivirus. J Virol 1991; 65: 4705-4721.

122. van Rijn PA, van Gennip HGP, de Meijer EJ, Moormann RJM. A preliminary map of epitopes on envelope glycoprotein E1 of HCV strain Brescia. Vet Microbiol 1992; 33: 221–230.

123. van Rijn PA, van Gennip HGP, de Meijer EJ, Moormann RJM. Epitope mapping of envelope glycoprotein E1 of hog cholera virus strain Brescia. J GenVirol 1993; 74: 2053–2060.

124. van Rijn PA, Miedema GKW, Wensvoort G, van Gennip HGP, Moormann RJ. Antigenic structure of envelope glycoprotein E1 of hog cholera virus. J Virol 1994; 68: 3934-42.

125. van Rijn PA, Bossers A, Wensvoort G, Moormann RJ. Classical swine fever virus (CSFV) envelope glycoprotein E2 containing one structural antigenic unit protects pigs from lethal CSFV challenge. J Gen Virol 1996; 77: 2737–2745.

126. van Rijn PA, van Gennip HG, Moormann RJ. An experimental marker vaccine and accompanying serological diagnostic test both based on envelope glycoprotein E2 of classical swine fever virus (CSFV). Vaccine 1999; 17: 433–440.

127. van Rijn PA. A common neutralizing epitope on envelope glycoprotein E2 of different pestiviruses: Implications for improvement of vaccines and diagnostics for classical swine fever (CSF). Vet Microbiol 2007;125:150–156.

128. van Oirschot J. Infectious diseases of livestock. Oxford University Press 2004; 2: 975-986.

129. Vilcek S, Herring AJ, Herring JA, Nettleton PF, Lowings JP, Paton DJ. Pestiviruses isolated from pigs, cattle and sheep can be allocated into at least three genogroups using polymerase chain reaction and restriction endonuclease analysis. Arch Virol 1994; 136: 309-323.

130. Vilcek S, Belak S. Genetic identification of pestivirus strain Frijters as a border disease virus from pigs. J Virol 1996; 60: 103-108.

131. Wang Z, Nie Y, Wang P, Ding M, Deng H. Characterization of classical swine fever virus entry by using pseudotyped viruses: E1 and E2 are sufficient to mediate viral entry. Virology 2004; 330: 332-341.

132. Weiland E, Stark R, Haas B, Riimenapf T, Thiel HJ. Pestivirus glycoprotein which induces neutralizing antibodies forms part of a disulphide-linked heterodimer. J Virol 1990; 64: 3563-9.

133. Wensvoort G, Bloemraad M, Terpstra C. An enzyme immunoassay employing monoclonal antibodies and detecting specifically antibodies to classical swine fever virus. Vet Microbiol 1988; 17: 129–140.

134. Wensvoort G. Topographical and functional mapping of epitopes on hog cholera virus with monoclonal antibodies. J Gen Virol 1989; 70: 2865–2876.

135. Widjojoatmodjo MN, van Gennip HG, Bouma A, van Rijn PA, Moormann RJ. Classical swine fever virus E(rns) deletion mutants: trans-complementation and potential use as nontransmissible, modified, live-attenuated marker vaccines. J Virol 2000; 74: 2973–2980.

136. Wiskerchen M, Belzer SK, Collett MS. Pestivirus gene expression: The first protein product of the bovine viral diarrhea virus large open reading frame, p20, possesses proteolytic activity. J Virol 1991; 65: 4508-4514.

137. Wooldridge M, Hartnett E, Cox A, Seaman M. Quantitative risk assessment case study: smuggled meats as disease vectors. Revue scientifique et technique Office International des Épizooties 2006; 25: 105-117.

138. Xabier A, Javier AVM, Paulino GP, Sjors HWS, Jose´ MC, Jose´ LC. Quasi-atomic model of bacteriophage T7 procapsid shell: insights into the structure and evolution of a basic fold. Structure 2007; 15: 461–472.

139. Xin Z, Jia X, Yuan S, Su L, Na L, Shuo Y, Fan H, Huang JH, Ling LJ, Qiu HJ. Identification of a linear epitope on the capsid protein of classical swine fever virus. Virus Res 2011; 156; 134-140.

140. Yu M, Wang LF, Shiell BJ, Morrissy CJ, Westbury HA. Fine mapping of a C-terminal linear epitope highly conserved among the major envelope glycoprotein E2 (gp51 to gp54) of different pestiviruses. Virology 1996; 222: 289–292.

141. Zhu Y, Shi Z, Drew TW, Wang Q, Qiu H, Guo H, Tu C. Antigenic differentiation of classical swine fever viruses in China by monoclonal antibodies. Virus Res 2009; 142: 169-174.

APPENDIX

APPENDIX A

LIST OF ABBREVIATIONS

ºC	degree Celsius
ºF	degree Fahrenheit
Asp	aspartic acid
bp	base pair
BDV	border disease virus
BVDV	bovine viral diarrhoea virus
cDNA	complementary deoxyribonucleic acid
CSF	classical swine fever
CSFV	classical swine fever virus
DNA	deoxyribonucleic acid
E2	envelope glycoprotein E2
ELISA	enzyme linked immunosorbent assay
ER	endoplasmic reticulum
EU	European Union
kb	kilobase
mAb	monoclonal antibody
NCR	non - coding region
NS	non - structural
OIE	the office International des Epizooties
PCR	polymerase chain reaction
RNA	ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
Ser	serine
SPP	signal peptide peptidase
ssRNA	single stranded RNA
USA	United States of America

Amino acid	**RNA codon**
Alanine (Ala/A)	GCU, GCC, GCA, GCG
Arginine (Arg/R)	CGU, CGC, CGA, CGG, AGA, AGG
Asparagine (Asn/N)	AAU, AAC
Aspartic acid (Asp/D)	GAU, GAC
Cysteine (Cys/C)	UGU, UGC
Glutamine (Gln/Q)	CAA, CAG
Glutamic acid (Glu/E)	GAA, GAG
Glycine (Gly/G)	GGU, GGC, GGA, GGG
Histidine (His/H)	CAU, CAC
Isoleucine (Ile/I)	AUU, AUC, AUA
START	AUG
Leucine (Leu/L)	UUA, UUG, CUU, CUC, CUA, CUG
Lysine (Lys/K)	AAA, AAG
Methionine (Met/M)	AUG
Phenylalanine (Phe/F)	UUU, UUC
Proline (Pro/P)	CCU, CCC, CCA, CCG
Serine (Ser/S)	UCU, UCC, UCA, UCG, AGU, AGC
Threonine (Thr/T)	ACU, ACC, ACA, ACG
Tryptophan (Trp/W)	UGG
Tyrosine (Tyr/Y)	UAU, UAC
Valine (Val/V)	GUU, GUC, GUA, GUG
STOP	UAA, UGA, UAG

APPENDIX B

Materials

1. Instruments

Centrifuge

Incubator

Vortex

Water bath

2. Chemicals

One step RT- PCR kit	Qiagen, Germany
pGEM-T easy vector system I	Promega, USA
pTZ57R/T vector	Fermentas, USA
QIAamp viral RNA mini kit	Qiagen, Germany
QIAprep spin miniprep kit	Qiagen, Germany
QIAquick gel extraction kit	Qiagen, Germany
100 bp DNA ladder marker	Geneaid, USA
Lamda DNA marker cut with *EcoRI* and *HindIII*	Bio Labs, Inc

3. Enzymes and accessory buffers

EcoRI-HF restriction enzyme	Bio Labs, Inc
XhoI restriction enzyme	Bio Labs, Inc
HindIII restriction enzyme	Bio Labs, Inc
*Nde*I restriction enzyme	Bio Labs, Inc
100X BSA	Bio Labs, Inc
10X buffer NEB 4	Bio Labs, Inc
10X buffer NEB 2	Bio Labs, Inc
T4 DNA ligase	Bio Labs, Inc
Taq DNA polymerase	Invitrogen

Methods

1. RNA extraction (QIAamp® Viral RNA Mini Kit, QIAGEN Germany)

1. Pipet 560 µl of prepared Buffer AVL containing Carrier RNA into a 1.5 ml microcentrifuge tube.
2. Add 140 µl spleen suspension in PBS pH 7.0 of CSFV wild type – infected swine to the Buffer AVL/Carrier RNA in the microcentrifuge tube. Mix by pulse-vortexing for 15 sec.
3. Incubate at room temperature (15–25°C) for 10 min.
4. Briefly centrifuge the 1.5 ml microcentrifuge tube to remove drops from the inside of the lid.
5. Add 560 µl of ethanol (96–100%) to the sample, and mix by pulse-vortexing for 15 sec. After mixing, briefly centrifuge the 1.5 ml microcentrifuge tube to remove drops from inside the lid.
6. Carefully apply 630 µl of the solution from step 5 to the QIAamp spin column (in a 2-ml collection tube) without wetting the rim. Close the cap, and centrifuge at 6000 x g (8000 rpm) for 1 min. Place the QIAamp spin column into a clean 2-ml collection tube, and discard the tube containing the filtrate.
7. Carefully open the QIAamp spin column, and repeat step 6.
8. Carefully open the QIAamp spin column, and add 500 µl of Buffer AW1. Close the cap, and centrifuge at 6000 x g (8000 rpm) for 1 min. Place the QIAamp spin column in a clean 2-ml collection tube (provided), and discard the tube containing the filtrate.
9. Carefully open the QIAamp spin column, and add 500 µl of Buffer AW2. Close the cap and centrifuge at full speed (20,000 x g; 14,000 rpm) for 3 min.
10. Place the QIAamp spin column in a clean 1.5-ml microcentrifuge tube (not provided). Discard the old collection tube containing the filtrate. Carefully open the QIAamp spin column and add 60 µl of Buffer AVE equilibrated to room temperature. Close the cap, and incubate at room temperature for 1 min. Centrifuge at 6000 x g (8000 rpm) for 1 min.
11. RNA samples were stored at -80°C until required.

2. PCR Purification (QIAquick PCR Purification Kit, QIAGEN Germany)

1. Add 500 µl of Buffer PB to 100 µl one step RT-PCR sample and mix.
2. Place a QIAquick spin column in a provided 2 ml collection tube.
3. To bind DNA, apply the sample to the QIAquick column and centrifuge for 30–60 s.
4. Discard flow-through. Place the QIAquick column back into the same tube. Collection tubes are re-used to reduce plastic waste.
5. To wash, add 0.75 ml Buffer PE to the QIAquick column and centrifuge for 30–60 s.
6. Discard flow-through and place the QIAquick column back in the same tube. Centrifuge the column for an additional 1 min.
7. Place QIAquick column in a clean 1.5 ml microcentrifuge tube.
8. To elute DNA, add 50 µl H_2O to the center of the QIAquick membrane, let the column stand for 1 min, and then centrifuge the column 1 min.
9. The eluted DNA was stored at –20 ºC until used.

3. Gel Extraction (QIAquick gel extraction kit, QIAGEN Germany)

1. Excise the DNA fragment from the agarose gel with a clean, sharp scalpel.
2. Weigh the gel slice in a colorless tube. Add 300 µl of Buffer QG to each 100 mg of gel. (100 mg ~ 100 µl).
3. Incubate at 50°C for 10 min (or until the gel slice has completely dissolved). To help dissolve gel, mix by vortexing the tube every 2–3 min during the incubation.
4. After the gel slice has dissolved completely, check that the color of the mixture is yellow (similar to Buffer QG without dissolved agarose).
5. Add 100 µl isopropanol to the sample and mix.
6. Place a QIAquick spin column in a provided 2 ml collection tube.
7. To bind DNA, apply the sample to the QIAquick column, and centrifuge for 1 min.
8. Discard flow-through and place QIAquick column back in the same collection tube. Collection tubes are re-used to reduce plastic waste.
9. Add 0.5 ml of Buffer QG to QIAquick column and centrifuge for 1 min.

10. To wash, add 0.75 ml of Buffer PE to QIAquick column and centrifuge for 1 min.
11. Discard the flow-through and centrifuge the QIAquick column for an additional 1 min at 13,000 rpm (~17,900 x *g*).
12. Place QIAquick column into a clean 1.5 ml microcentrifuge tube.
13. To elute DNA, add 50 µl H_2O to the center of the QIAquick membrane, let the column stand for 1 min, and then centrifuge the column 1 min.
14. The eluted DNA was stored at –20 ºC until used.

4. Preparation of competent E. coli cells for transformation

1. A single colony of *E. coli* strain DH5α was inoculated into 5 ml of LB broth and incubated overnight at 37°C with shaking.
2. 100 µl of culture was inoculated into 10 ml of LB broth and incubated with shaking at 37°C for 2-3 h or until an OD value approximately 0.5 at wavelength of 600 nm was obtained.
3. Aliquot the culture into 1.5 ml microcentrifuge tube and placed on ice for 5 min. The tube was centrifuged at 3,000 x g at 4°C for 15 min.
4. The supernatant was discarded and suspended the pellet in 4 ml of cold transformation buffer I (TFB I). The tube was centrifuged at 3,000 x g at 4°C for 15 min.
5. The supernatant was discarded and suspended the pellet in 4 ml of cold transformation buffer II (TFB II).
6. The cells suspension was cooled on ice for 15 min. The 200 µl of cells suspension was aliquot into 1.5 ml microcentrifuge tube and immediately transferred to -80°C for storage.

5. Ligation of DNA fragment

The ligation reaction was set up which consisted of 10 µl of 2X rapid ligation buffer, 1 µl of T4 DNA ligase. 0.5 µl of the vector, 5 µl of DNA inserts and sterile deionized water to make a total volume of 20 µl. Reaction was mixed by pipetting and was incubated at 4°C overnight.

6. Bacterial transformation

1. Remove the competent bacterial cells from – 80°C and thawed on ice.
2. Transfer bacterial cells to sterile ice-cold 50 ml polypropylene tube.
3. Add 20 µl of ligation mixture to polypropylene tube. The mixture was incubated on ice for 30 minutes.
4. The tube was incubated at 42°C in water bath for 90 sec and immediately placed on ice for 2 min (heat shock process).
5. Added 800 µl of LB broth and incubated overnight at 37°C with shaking.
6. Centrifuged at 3,000 x g for 5 minutes, and then the supernatant was discarded approximately 900 µl.
7. The residual supernatant was resuspended the cells and was spread on LB agar plates contain ampicillin, X-Gal and IPTG.
8. The plates were incubated overnight at 37°C. Recombinant clones were identified by color screening on indicator plates (blue/white screening). The white colonies were selected.

7. Plasmid purification (QIAprep spin miniprep kit ,Qiagen Germany)

1. A single white colony of bacterial was picked from LB agar plate into 2 ml of LB broth contained ampicillin and incubated overnight at 37°C with shaking.
2. Transfered bacterial culture into 1.5 ml microcentrifuge tube, centrifuged for 5 minutes at full speed, in order to collect the cells and discarded supernatant.
3. Resuspend pelleted bacterial cells in 250 µl Buffer P1 and transfer to a microcentrifugetube.
4. Add 250 µl Buffer P2 and mix thoroughly by inverting the tube gently 4–6 times.
5. Add 350 µl Buffer N3 and mix immediately and thoroughly by inverting the tube 4–6 times. Centrifuge for 10 min at 13,000 rpm (~17,900 x g).
6. Apply the supernatant to the QIAprep spin column by decanting or pipetting. Centrifuge for 30–60 s. Discard the flow-through.
7. Wash the QIAprep spin column by adding 0.5 ml Buffer PB and centrifuging for 30–60 s. Discard the flow-through.

8. Wash QIAprep spin column by adding 0.75 ml Buffer PE and centrifuging for 30–60 s.
9. Discard the flow-through, and centrifuge for an additional 1 min to remove residual wash buffer.
10. Place the QIAprep column in a clean 1.5 ml microcentrifuge tube. To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water to the center of each QIAprep spin column, let stand for 1 min, and centrifuge for 1 min.
11. The eluted DNA was stored at –20 ºC until used.

Preparation of Reagent and Media

1. Reagents for bacterial cultivation

1.1 Luria-Bertani (LB) broth medium

Bacto-Tryptone	1.0	g
NaCl	0.5	g
Bacto-Yeast extract	0.5	g
Distilled water	100	ml

Autoclave at 121°C with pressure 15 lbs/in2 for 15 minutes.

1.2. Luria-Bertani (LB) agar medium

Bacto-Tryptone	1.0	g
NaCl	0.5	g
Bacto-Yeast extract	0.5	g
Agarose	2.0	g
Distilled water	100	ml

Autoclave at 121°C with pressure 15 lbs/in2 for 15 minutes.

For blue/white screening were used LB plates with ampicillin/IPTG/X-Gal: After autoclave, cool to 55 °C. Added 100 mg/ml of ampicillin, 1 M/ml of IPTG and 50 mg/ml of X-Gal. Poured into petridishes.

Let harden, then inverted and stored at 4 °C.

2. Reagents for Blue/White Screening

2.1. Ampicillin (Amp): Stock of 100 mg/ml

Ampicillin	1.0	g
Sterile deionized water	10	ml

and stored at -20 °C.

2.2. IPTG (isopropyl-b-D-thiogalactopyranoside): Stock of 1 M/ml

Isopropyl-b-D-thiogalactopyranoside	2.38	g
Sterile deionized water	10	ml

and stored at 4 °C.

2.3. X-gal (5-bromo-4-chloro-3-indolyl b-D-galactopyranoside): Stock of 50 mg/ml was purchased from Promega, USA and stored at -20 °C.

3. Reagents for preparation of competent cells

3.1. Transformation buffer I (TFB I)

CH_3COOK	0.29	g
RuCl	1.21	g
$CaCl_2 . 6H_2O$	0.22	g
$MgCl_2 . 4H_2O$	0.99	g
Glycerol	15	ml

Adjust to pH = 5.8 with 0.2 M CH_3COOH

Adjusted the volume to 100 ml with deionized water

Sterilised by filtration

3.2.Transformation buffer II (TFB II)

MOPS	0.21	g
RuCl	0.12	g
$CaCl_2 . 6H_2O$	1.64	g
Glycerol	15	ml

Adjust to pH = 6.5 with KOH

Adjusted the volume to 100 ml with deionized water

Sterilised by filtration

4. Reagents for electrophoresis assay

4.1. 5X TAE buffer (stock solution)

Tris-base	24.2	g
Glacial acetic	5.71	ml
0.5 M EDTA (pH 8.0)	10.0	ml

Adjusted the volume to 1000 ml with deionized water

4.2. 1X TAE buffer (working solution)

5X TAE buffer	200	ml
Deionized water	800	ml

4.3. 1.0 % agarose gel

Agarose type II	1.0	g
1X TAE buffer	100	ml

4.4. Ethidium bromide solution: Stock of 10 mg/ml

Ethidium bromide	1.0	g
Deionized water	100	ml

4.5 Loading dye composition(6X)

0.25% Bromophenol blue (BPB)

0.25% xylene cyanol (optional)

30% glycerol in water

Classical swine fever virus strain Thiverval, complete genome

GenBank: EU490425 12321 bp RNA

Country: France

374..12070 /codon_start=1

Product: "polyprotein"

Translation :

"MELNHFELLYKTNEQKPMGVEEPVYDATGKPLFGDPSEVHPQST

KLPHDRGRGNIKTTLKNLPRKGDCRSGNHLGPVSGIYVKPGPVFYQDYMGPVYHRAP

LEFFSEAQFCEVTKRIGRVTGSDGRLYHIYVCIDGCILLKLAKRGEPRTLKWIRNFTD

CPLWVTSCSDDGASGSKEKKPDRINKGKLKIAPKEHEKDSRTKPPDATIVVEGVKYQV

RKKGKVKGKNTQDGLYHNKNKPPESRKKLEKTLLAWAVIAIMLYQPVEAENITQWNLS

DNGTNGIQHAMYLRGISRSLHGIWPEKICKGVPTYLATDTELKEIQGMMDASEGTNYT

CCKLQRHEWNKHGWCNWYNIDPWIQLMNRTQANLAEGPPAKECAVTCRYDKDADVNVV

TQARNRPTTLTGCKKGKNFSFAGTVIEGPCNFNVSVEDILYGDHECGSLLQDTALYLV

DGMTNTIENARQGAARVTSWLGRQLRIAGRRLEGRSKTWFGAYALSPYCNVTSKIGYI

WYTNNCTPACLPKNTKIIGPGKFDTNAEDGKILHEMGGHLSEFLLLSLVVLSDFAPET

ASALYLIFHYVIPQSHEELEGCDTNQLNLTVELRTEDVIPSSVWNVGKYVCVRPDWWP

YETKVALLFEEAGQVVKLAVRALRDLTRVWNSASTTAFLICLIKVLRGQIVQGVIWLL

LVTGAQGQLACKEDYRYAISSTNEIGLLGAGGLTTTWKEYNHDLQLNDGTVKAICVAG

SFKVTALNVVSRRYLASLHKEALPTSVAFELLFDGTNPSTEEMGDDFGFGQCPFDTSP

VVKGKYNTTLLNGSAFYLVCPIGWTGVIECTAVSPTTLRTEVVKTFRRDKPFPHRMFC

VTTTVENEDLFYCKLGGNWTCVKGEPVVYTGGLVKQCRWCGFDFNEPDGLPHYPIGKC

ILANETGYRIVDSTDCNRDGVVISTEGSHECLIGNTTVKVHASDERLGPMPCRPKEIV

SSAGPVRKTSCTFNYAKTLKNKYYEPRDSYFQQYMLKGEYQYWFDLDVTDRHSDYFAE

FVVLVVVALLGGRYILWLIVTYIVLTEQPAAGLPLGQGEVVLIGNLITHTDIEVVVYF

LLLYLVMRDEPIKKWILLLFHAITNNPVKTITVALLMVSGVARGGKIDGGWQRLPETS

FDIQLALTVIVVAVMLLAKRDPTTVPLVVTVATLRTAKMTNGLSTDIAIATVSTALLT

WTYISDYYRYKTWLQYLISTVTGIFLIRVLKGIGELDLHTPTLPSYRPLFFILVYLIS

TAVVTRWNLDIAGLLLQCAPTLLMVFTMWADILTLILILPTYELTKLYYLKEVKIGAE

RGWLWKTNFKRVNDIYEVDQAGEGVYLFPSKQKTSSITGTMLPLIKAILISCISNKWQ

FIYLLLYLIFEVSYYLHKKIIDEIAGGTNFISRLVAALIEANWAFDNEEVRGLKKFFLL

SSRVKELIIKHKVRNEVMVHWFGDEEVYGMPKLVGLVKAATLSKNKHCILCTVCEDRE

WRGETCPKCGRFGPPMTCGMTLADFEEKHYKRIFFREDQSEGPVREEYAGYLQYRARG

QLFLRNLPVLATKVKMLLVGNLGTEVGDLEHLGWVLRGPAVCKKVTEHEKCTTSIMDK

LTAFFGVMPRGTTPRAPVRFPTSLLKIRRGLETGWAYTHQGGISSVDHVTCGKDLLVC

DTMGRTRVVCQSNNKMTDESEYGVKTDSGCPEGARCYVFNPEAVNISGTKGAMVHLQK
TGGEFTCVTASGTPAFFDLKNLKGWSGLPIFEASSGRVVGRVKVGKNEDSKPTKLMSG
IQTVSKSTTDLTEMVKKITTMNRGEFRQITLATGAGKTTELPRSVIEEIGRHKRVLVL
IPLRAAAESVYQYMRQKHPSIAFNLRIGEMKEGDMATGITYASYGYFCQMPQPKLRAA
MVEYSFIFLDEYHCATPEQLAIMGKIHRFSENLRVVAMTATPAGTVTTTGQKHPIEEF
IAPEVMKGEDLGSEYLDIAGLKIPVEEMKSNMLVFVPTRNMAVETAKKLKAKGYNSGY
YYSGEDPSNLRVVTSQSPYVVVATNAIESGVTLPDLDVVVDTGLKCEKRIRLSPKMPF
IVTGLKRMAVTIGEQAQRRGRVGRVKPGRYYRSQETPVGSKDYHYDLLQAQRYGIEDG
INITKSFREMNYDWSLYEEDSLMITQLEILNNLLISEELPMAVKNIMARTDHPEPIQL
AYNSYETQVPVLFPKIKNGEVTDSYDNYTFLNARKLGDDVPPYVYATEDEDLAVELLG
LDWPDPGNQGTVEVGRALKQVVGLSTAENALLVALFGYVGYQALSKRHIPVVTDIYSI
EDHRLEDTTHLQYAPNAIKTEGKETELKELAQGDVQRCVGAMTNYAREGIQFMKSQAL
KVKETPTYKETMDTVTDYVKKFMEALTDSKEDIIKYGLWGTHTALYKSICARLGSETA
FATLVVKWLAFGGESIADHVKQAATDLVVYYIINRPQFPGDTETQQEGRKYVASLLVS
ALVTYTYKSWNYNNLSKIVEPALATLPYAATALKLFAPTRLESVVILSTAIYKTYLSI
RRGKSDGLLGTGVSAAMEIMSQNPVSVGIAVMLGVGAVAAHNAIEASEQKRTLLMKVF
VKNFLDQAATDELVKESPEKIIMALFEAVQTVGNPLRLVYHLYGVFYKGWEAKELAQR
TAGRNLFTLIMFEAVELLGVDSEGKIRQLSSNYILELLYKFRDSIKSSVRDMAISWAP
APFSCDWTPTDDRIGLPQDNFLQVETKCPCGYKMKAVKNCAGELRLLEEEGSFLCRNK
FGRGSRNYRVTKYYDDNLSEIKPVIRMEGHVELYYKGATIKLDFNNSKTILATDKWEV
DHSTLVRVLKRHTGAGYHGAYLGEKPNHKHLIERDCATITKDKVCFLKMKRGCAFTYD
LSLHNLTRLIELVHKNNLEDKEIPAVTVTTWLAYTFVNEDIGTIKPAFGEKVTPEMQE
EITLQPAVVVDTTDVTVTVVGEAPTMTTGETPTAFTSSGSDPKGQQVLKLGVGEGQYP
GTNPQRASLHEAIQGADERPSVLILGSDKATSNRVKTAKNVKVYRGRDPLEVRDMMRR
GKILVIALSRVDNALLKFVDYKGTFLTRETLEALSLGRPKKKNITKAEAQWLLCLEDQ
MEELPDWFAAGEPIFLEANIKHDRYHLVGDIATIKEKAKQLGATDSTKISKEVGAKVY
SMKLSNWVMQEENKQGNLTPLFEELLQQCPPGGQNKTAHMVSAYQLAQGNWMPTSCHV
FMGTISARRTKTHPYEAYVKLRELVEEHKMKTLCPGSSLGKHNEWIIGKIKYQGNLRT
KHMLNPGKVAEQLCREGHRRNVYNKTIGSVMTATGIRLEKLPVVRAQTDTTNFHQAIR
DKIDKEENLQTPGLHKKLMEVFNALKRPELESSYDAVEWEELERGINRKGAAGFFERK
NIGEILDSEKNKVEEIIDNLKKGRNIKYYETAIPKNEKRDVNDDWTSGDFVDEKKPRV
IQYPEAKTRLAITKVMYKWVKQKPVVIPGYEGKTPLFQIFDKVKKEWDQFQNPVAVSF
DTKAWDTQVTTKDLELIKDIQKYYFKKKWHKFIDTLTMHMSEVPVISADGEVYIRKGQ
RGSGQPDTSAGNSMLNVLTMVYAFCEATGVPYKSFDRVAKIHVCGDDGFLITERALGE
KFASKGVQILYEAGKPQKITEGDKMKVAYQFDDIEFCSHTPIQVRWSDNTSSYMPGRN

```
TTTILAKMATRLDSSGERGTIAYEKAVAFSFLLMYSWNPLIRRICLLVLSTELQVKPG
KSTTYYYEGDPTSAYKEVIGHNLFDLKRTSFEKLAKLNLSMSVLGAWTRHTSKRLLQD
VNMGVKEGNWLVNADRLVSSKTGNRYIPGEGHTLQGRHYEELALARKQINNFQGTDR
YNLGPIVNMVLRRLRVMMMTLIGRGV"
ORIGIN
        1 gtatacgagg ttagttcatt ctcgtatgca tgattggaca aattaaaatt tcaatttgga
       61 tcagggcctc cctccagcga cggccgaact gggctagcca tgcccacagt aggactagca
      121 aacggaggga ctagccgtag tggcgagctc cctgggtggt ctaagtcctg agtacaggac
      181 agtcgtcagt agttcgacgt gagcagaagc ccacctcgag atgctatgtg gacgagggca
      241 tgcccaagac acaccttaac cctagcgggg gtcgctaggg tgaaatcaca ccacgtgatg
      301 ggagtacgac ctgatagggt gctgcagagg cccactatta ggctagtata aaaatctctg
      361 ctgtacatgg cacatggagt tgaatcattt tgaactttta tacaaaacaa acgaacaaaa
      421 accaatggga gtggaggaac cggtatacga tgccacgggg aaaccattgt ttggagaccc
      481 gagtgaggta cacccacaat caacactgaa gctaccacat gataggggga gaggtaacat
      541 caaaacaaca ctgaagaacc tacctaggaa aggcgactgc aggagtggca accatctagg
      601 cccggttagc gggatatatg taaagcccgg ccctgtcttt tatcaggact acatgggccc
      661 ggtctaccat agagcccctc tagagttttt tagcgaagcg cagttttgtg aggtgaccaa
      721 aaggataggt agggtgacag gtagtgacgg aaggctttac catatatatg tgtgcatcga
      781 tggttgcata ctgctgaagc tagccaaaag ggcgagccaa agaaccctga agtggattag
      841 aaatttcacc gactgtccat tgtgggttac cagttgctct gatgatggcg caagtggaag
      901 taaagagaag aagccagata ggatcaacaa gggtaaatta aaaatagccc cgaaagagca
      961 tgagaaggac agcagaacta agccacctga cgctacgatc gtagtggaag gagtaaaata
     1021 ccaggtcaga aagaaaggta aagttaaagg aaagaatacc caagacggcc tgtaccacaa
     1081 taagaataaa ccaccagaat ctaggaagaa attagaaaaa accctattgg catgggcggt
     1141 aatagcaatt atgttgtacc aaccagttga agccgaaaat ataactcaat ggaacctgag
     1201 tgacaacggc actaatggta tccagcatgc tatgtacctt agagggatta gcagaagctt
     1261 gcatgggatc tggccggaaa aaatatgcaa aggagtcccc acctacctgg ccacagacac
     1321 ggaactgaaa gaaatacagg gaatgatgga tgccagcgag gggacaaact atacgtgctg
     1381 taagttacag agacatgaat ggaacaaaca tggatggtgt aactggtaca atatagaccc
     1441 ctggatacag ttgatgaata gaacccaagc aaacttggca gaaggccctc cggccaagga
     1501 gtgcgctgtg acttgtaggt acgataaaga tgctgacgtc aacgtggtca cccaggccag
     1561 aaacaggcca acaaccctga ccggctgcaa gaaaggaaaa aatttttctt ttgcaggtac
     1621 agttatagag ggcccatgta atttcaatgt ttccgtggag gatatcttgt atggggatca
     1681 tgagtgcggc agtttgctcc aggacacggc tctgtaccta gtagatggaa tgaccaacac
     1741 tatagagaat gccagacagg gagcagcgag ggtaacatct tggctcggga ggcaactcag
     1801 aattgccggg aggaggttgg agggtagaag caaaacctgg ttcggtgcct atgccctatc
     1861 accttactgt aacgtaacaa gcaaaatagg gtacatatgg tacactaaca actgcacccc
     1921 ggcttgcctc cccaagaata caaagataat aggccccggt aaatttgaca ctaatgcgga
     1981 ggacggaaag attctccatg agatgggcgg ccacctatca gaatttctgc tgctctctct
     2041 ggttgttctg tctgacttcg cccctgaaac agccagcgcg ttatacctca tttttcacta
     2101 cgtgattcct caatcccatg aagaacttga aggctgtgac acaaaccagc tgaatctaac
     2161 agtggaactc aggactgaag acgtaatacc gtcatcagtc tggaatgttg gcaaatatgt
```

2221 gtgtgttaga ccagactggt ggccatatga aaccaggtg gctctgttat ttgaagaggc

2281 aggacaggtc gtaaagttag ccgtacgggc gctgagggat ttaaccaggg tctggaatag

2341 cgcatcaacc acggcattcc tcatctgctt gataaaagta ttaagaggac agatcgtgca

2401 aggtgtgata tggctgctac tagtaactgg ggcacaaggc cagctagcct gcaaggaaga

2461 ttacaggtac gcaatatcat caaccaatga gatagggcta ctcggggccg gaggtctcac

2521 caccacctgg aaagaataca accacgattt gcaactgaat gacgggaccg ttaaggccat

2581 ttgcgtggca ggttccttta aagtcacagc acttaatgtg gtcagtagga ggtatttggc

2641 atcattgcat aaggaggctt tacccacttc cgtggcattc gagctcctgt tcgacgggac

2701 caacccatca actgaggaaa tgggagatga cttcgggttc gggcagtgcc cgtttgatac

2761 gagtcctgtt gtcaagggaa agtacaatac aaccttgttg aacggtagtg ctttctatct

2821 tgtctgtcca atagggtgga cgggtgttat agagtgcaca gcagtgagcc caacaactct

2881 gagaacagaa gtggtaaaga ccttcaggag ggacaagccc tttccgcaca gaatgttttg

2941 tgtgaccaca acagtggaaa atgaagattt attctactgt aagttggggg gcaactggac

3001 atgtgtgaaa ggtgaaccag tggtctacac gggggggcta gtaaaacaat gcagatggtg

3061 tggctttgac ttcaatgagc ctgacggact cccacactac cccataggta agtgcatttt

3121 ggcaaatgag acaggttaca gaatagtgga ttcaacagac tgtaacagag atggtgttgt

3181 aatcagcaca gaggggagtc atgagtgctt gatcggtaac acaactgtca aggtgcatgc

3241 atcagatgaa agactgggcc ccatgccatg cagacctaaa gagatcgtct ctagtgcagg

3301 acctgtaagg aaaacttcct gtacattcaa ctacgcaaaa actttgaaga acaagtacta

3361 tgagcccagg gacagctact tccagcaata tatgcttaag ggcgagtatc agtactggtt

3421 tgacctggac gtgactgacc gccactcaga ttacttcgca gaatttgttg tcttggtggt

3481 ggtagcactg ttaggaggaa gatatatcct gtggctaata gtgacctaca tagttttaac

3541 agaacaaccc gccgctggtt taccattggg ccagggtgag gtagtgttga tagggaactt

3601 aattacccac acagacattg aggttgtagt atatttctta ctactctatt tggtcatgag

3661 ggatgagcct ataaagaaat ggatactgct gctattccat gctataacta acaatccagt

3721 caagactata acagtggcat tgctcatggt tagcggggtt gccaggggtg gaaagataga

3781 tggcggttgg cagcggctgc cggagaccag ctttgacatc caactcgcgc tgacagttat

3841 agtagtcgct gtgatgttgc tagcaaagag agatccgact actgtcccct tggttgtaac

3901 ggtggcaacc ctgagaacgg ctaagatgac taatggactt agtacggata tagccatagc

3961 tacagtgtca acagcgttgc taacctggac ctacattagt gactattata gatacaagac

4021 ttggctacag taccttatta gcacagtgac aggtatcttt ttaataaggg tactgaaggg

4081 aataggtgag ttggatttac acactccaac cttgccatct tatagacccc tcttcttcat

4141 tctcgtgtac ctcatttcca ctgcagtggt aacaagatgg aatctggaca tagccggatt

4201 gctgttgcag tgtgccccaa cccttttgat ggtttttacg atgtgggcag atattctcac

4261 cttgatcctc atactgccca cttacgagtt aacaaagcta tattacctca aggaagtgaa

4321 gattggggca gaaaggggct ggttatggaa gaccaacttc aagagggtaa acgacatata

4381 cgaggttgac caagctggtg aaggggtata ccttttcccg tcaaaacaaa aaacaagttc

4441 aataacaggt accatgttgc cattgatcaa agccatactc atcagctgca tcagtaataa

4501 gtggcagttc atatacctat tgtacttgat atttgaagtg tcttactacc tccacaagaa

4561 gatcatagat gaaatagcag gagggactaa cttcatctca agacttgtag ccgctttgat

4621 cgaagccaat tgggcctttg acaacgaaga agttagaggt ttaaagaagt tcttcctgtt

4681 gtctagtagg gttaaagaac tgatcatcaa acacaaagtg aggaatgaag taatggtcca

4741 ctggtttggt gacgaagagg tttatgggat gccgaagttg gttggcttag tcaaggcagc

4801 aacattgagt aaaaataaac attgcatttt gtgcaccgtc tgtgaagaca gagagtggag
4861 aggagaaacc tgcccaaaat gcgggcgttt tgggccacca atgacctgtg gcatgaccct
4921 agccgacttt gaagaaaaac actataagag gatctttttt agagaggatc aatcagaagg
4981 gccggttaga gaggagtacg cagggtatct gcaatacaga gccagagggc aattattcct
5041 gaggaatctc ccagtgctag caacaaaagt caagatgctc ctggtcggaa atcttgggac
5101 ggaggtggga gatttggaac accttggctg ggttcttaga gggcctgccg tttgcaagaa
5161 ggttaccgaa catgagaaat gcaccacatc cataatggat aaattgacag cttttttcgg
5221 tgttatgcca aggggcacca cacctagagc ccctgtgaga ttccccacct ctctcttaaa
5281 gataagaagg gggttagaaa ctggctgggc gtacacacac caaggtggca ttagttcagt
5341 ggaccatgtc acttgcggga aagacttact ggtatgtgac actatgggcc ggacaagggt
5401 cgtttgccaa tcaaataata agatgacaga cgagtccgag tatggagtta aaactgactc
5461 cggatgcccg gaaggagcta ggtgttatgt gttcaaccca gaggcagtta acatatcagg
5521 gactaaagga gccatggtcc acttacaaaa gactggagga gaattcacct gtgtgacagc
5581 atcaggaact ccagccttct ttgatctcaa gaacctcaaa ggctggtcag ggctaccgat
5641 atttgaggca tcaagtggaa gggtagtcgg cagggtcaag gtcgggaaga atgaggactc
5701 taaaccaacc aagcttatga gtggaataca aacagtctcc aaaagtacca cagacttgac
5761 agaaatggta aagaaaataa cgaccatgaa caggggagaa ttcagacaaa taacccttgc
5821 tacaggtgcc ggaaaaacca cggaacttcc taggtcggtc atagaagaga tagggaggca
5881 taagagagtc ttggtcttga tccctctgag ggcggcagca gagtcagtat accagtatat
5941 gagacaaaaa catccaagca tcgcatttaa cctgaggata ggggagatga aggaagggga
6001 catggccaca gggataacct atgcctcata cggttacttc tgtcagatgc cacaacctaa
6061 gttgcgagcc gcaatggttg agtactcctt catatttctt gatgagtacc actgcgccac
6121 cccagaacaa ttggctatca tgggaaagat ccacagattt tcagagaacc tgcgggtagt
6181 agccatgacc gcaacaccag caggcacggt aacaaccaca gggcagaaac accctataga
6241 agaattcata gccccagaag tgatgaaagg ggaagactta ggctcagagt acttggacat
6301 tgctggacta aagatacctg tagaggagat gaagagcaac atgctggttt ttgtgcccac
6361 taggaacatg gcggtggaga cagcaaagaa attgaaagct aagggctaca actcaggcta
6421 ctattatagt ggtgaggatc catctaacct gagagtggta acgtcacagt ccccatacgt
6481 ggtggtggca accaacgcga tagaatcagg tgttactctc ccggacttgg atgtggtcgt
6541 cgatacaggg cttaagtgtg aaaagagaat acggctgtca cctaagatgc ccttcatagt
6601 gacgggcctg aagaggatgg ctgtcacgat tggggaacaa gcccagagaa gggggagagt
6661 tgggagagta aagcctggga gatactacag gagtcaagaa actcccgttg gttctaaaga
6721 ttaccattac gatctactgc aagcacagag gtacggtatt gaagatggga taaacatcac
6781 caaatccttt agagagatga actatgattg gagcctttat gaggaggaca gtctgatgat
6841 tacacaattg gaaatcctca ataatttgtt gatatcagaa gaactaccga tggcagtaaa
6901 aaatataatg gccaggactg accacccaga accaattcag ctggcgtaca acagctacga
6961 aacacaagtg ccagtgctat tcccaaaaat aaaaaatgga gaggtgactg acagttacga
7021 taactatacc ttcctcaatg caagaaaatt gggggatgat gtaccccctt acgtgtatgc
7081 cacagaggat gaggacttag cggtagagct gctgggctta gactggccag accctggaaa
7141 ccaaggaacc gtagaggttg gcagagcact aaaacaagta gttggtctat caacagctga
7201 gaatgccctg ttagtagcct tattcggcta tgtaggatat caggcacttt caaagaggca
7261 tataccagta gtcacagata tatattcaat tgaagatcac aggttggaag acaccacaca
7321 cctacagtac gccccgaatg ctatcaagac ggaggggaag gagacagagt tgaaagagct

```
7381 agctcagggg gatgtgcaga gatgtgtggg agctatgacc aattatgcaa gagagggcat
7441 ccagttcatg aagtctcagg cactgaaggt gaaagaaacc cccacttaca aagagactat
7501 ggacactgtg acggactatg taaagaaatt catggaggcg ctgacagaca gtaaagagga
7561 catcataaaa tatgggttgt gggggacgca cacagcctta tataagagca tctgtgccag
7621 gctcgggagt gagactgcgt tcgctaccct ggtcgtgaag tggctggcat ttggggggga
7681 atcaatagca gaccatgtca aacaagcggc cacagacttg gtcgtttact atatcatcaa
7741 cagacctcag ttcccaggag acacggagac acaacaggaa ggaaggaaat atgtggccag
7801 cctactggtc tcagctctag ttacttacac atacaaaagc tggaattaca ataatctgtc
7861 caagatagtt gaaccggctt tagccactct gccctatgcc gccacagctc tcaaactatt
7921 cgctcccact cgattggaga gcgttgtcat attgagtacc gcaatctaca aaacctacct
7981 gtcaatcagg cgcggaaaaa gcgatggttt gctaggcaca ggggttagtg cggctatgga
8041 gatcatgtca caaaatccag tatccgtggg catagcagtc atgctagggg taggggccgt
8101 ggcagcccac aatgcaatcg aagccagtga gcagaagaga acactactca tgaaagtttt
8161 tgtaaagaac ttcttggacc aggcagccac tgatgaatta gtcaaggaga gtcctgagaa
8221 aataataatg gctttgtttg aagcagtgca gacagtcggc aaccctctta gactagtata
8281 ccacctttat ggagtttttt ataaggggtg ggaggcaaaa gagttggccc aaaggacagc
8341 cggtaggaac cttttcactt tgataatgtt cgaggctgtg gaactactag gagtagatag
8401 tgaaggaaag atccgccagc tatcaagtaa ttacattcta gagctcctgt ataagttccg
8461 tgacagtatc aagtctagcg tgagggatat ggcaatcagc tgggcccctg cccctttcag
8521 ttgtgattgg acaccgacgg atgacagaat agggctcccc caagacaatt tcctccaagt
8581 ggagacgaaa tgcccctgtg gttacaagat gaaggcagtt aagaattgtg ctggagagct
8641 aagactctta gaggaggaag gttcatttct ctgcagaaat aaattcggga gaggttcacg
8701 gaactacagg gtgacaaaat actatgatga caatctatca gaaataaagc cagtgataag
8761 aatggaaggg catgtggaac tatactacaa gggggccacc atcaaactgg atttcaacaa
8821 cagtaaaaca atattggcaa ccgataaatg ggaggttgat cactccactc tggtcagggt
8881 gctcaagagg cacacagggg ctggatatca tggggcatac ctgggcgaga aaccgaacca
8941 caaacatctg atagagaggg actgtgcaac catcaccaaa gataaggttt gttttctaaa
9001 aatgaagaga gggtgtgcat tcacttatga cttatccctt cacaacctta cccgactgat
9061 tgaattggta cacaagaata acttggaaga caaagagatc cctgctgtta cggttacaac
9121 ctggctggct tacacgtttg taaatgaaga tatagggacc ataaaaccag ccttcgggga
9181 gaaagtaaca ccggagatgc aggaggagat aaccttgcag cctgctgtag tggtggatac
9241 aactgacgtg accgtgactg tggtagggga agcccctact atgactacag gggagactcc
9301 gacagcgttc accagctcag gttcagaccc gaaaggccaa caagttttaa aactgggggt
9361 aggtgaagga caataccccg ggactaatcc acagagggca agcctgcacg aagccataca
9421 aggtgctgat gagaggccct cggtgctgat attggggtct gataaagcca cctctaatag
9481 agtaaaaact gcaaagaatg taaaggtata cagaggcagg gacccactag aagtgagaga
9541 tatgatgagg aggggaaaga tcctggtcat agccctgtct agggttgata atgctctatt
9601 gaaatttgtt gattacaaag gcacctttct aactagagag accctagagg cattaagttt
9661 gggtaggcct aaaaagaaaa acataaccaa ggcagaagca cagtggttgc tgtgccttga
9721 agaccaaatg gaagagctac ccgattggtt cgcagccggg gaacccattt ttctagaggc
9781 caacattaaa catgacaggt atcatctggt gggggatata gctactatca aggaaaaagc
9841 caaacagttg gggctacag actccacaaa gatatctaag gaggttggtg caaaagtgta
9901 ttctatgaaa ctgagtaatt gggtgatgca agaagaaaat aaacaggca acctgacccc
```

```
 9961 cttgttcgaa gagctcctgc aacagtgtcc acccgggggc cagaacaaaa ctgcacatat
10021 ggtctctgct taccaactag cccaagggaa ctggatgcca accagctgcc atgtttttat
10081 ggggaccata tctgccagga ggaccaagac tcatccatat gaagcatatg tcaagttaag
10141 ggagttggta gaggaacaca agatgaaaac attgtgtccc ggatcaagcc tgggtaagca
10201 caacgaatgg ataattggta agatcaaata ccagggaaac ctgaggacca aacacatgtt
10261 gaaccccggc aaggtggcag agcaactgtg cagagaggga cacagacgca atgtgtataa
10321 caagacaata ggctcagtaa tgacagctac tggtatcagg ttggagaaat tgcccgtggt
10381 tagggcccag acagacacaa ccaacttcca ccaagcaatc agggataaga tagacaagga
10441 agagaaccta cagaccccgg gtttacataa gaaactaatg gaggttttca atgcattgaa
10501 acgacccgag ttagagtcct cctacgatgc cgtggaatgg gaggaactgg agagaggaat
10561 aaacaggaag ggtgctgctg gtttcttcga acgcaaaaat ataggggaaa tattggattc
10621 agagaaaaac aaagtcgaag agattattga caatctgaaa aaaggtagaa acatcaaata
10681 ctatgaaact gcgatcccaa agaatgagaa gagggacgtc aatgatgact ggacctctgg
10741 tgacttcgtg gacgagaaga agcccagagt catacaatac cctgaagcaa aaacaaggct
10801 ggccatcacc aaggtgatgt ataagtgggt gaagcagaag ccagtagtta tacccgggta
10861 tgaagggaag acacctctgt tccaaatttt tgacaaagta aagaaggaat gggatcaatt
10921 ccaaaatcca gtggcagtga gcttcgacac taaggcgtgg gacacccagg taaccacaaa
10981 agatttggag ctgataaagg acatacaaaa gtactatttc aagaagaaat ggcataaatt
11041 tattgacacc ctgaccatgc atatgtcaga agtacccgta atcagcgccg atggggaagt
11101 atacataagg aaagggcaaa gaggcagtgg acaacctgac acaagcgcag gcaatagcat
11161 gctaaatgtg ttaacaatgg tttacgcctt ctgcgaggcc acgggagtac cctacaagag
11221 ctttgacagg gtggcaaaaa ttcatgtgtg cggggatgat ggtttcctga tcacagaaag
11281 ggctctcggt gagaaattcg cgagcaaggg agtccagatc ctatatgaag ctgggaagcc
11341 ccagaagatc actgaagggg ataaaatgaa agtggcctac caatttgatg atattgagtt
11401 ttgctcccac acaccaatac aagtaaggtg gtcagataac acctctagtt acatgccggg
11461 gagaaataca accacaatcc tggctaaaat ggccacaagg ttagattcca gtggtgagag
11521 gggcaccata gcatatgaga aagcagtagc attcagcttc ctgctgatgt actcctggaa
11581 cccactaatt agaaggatct gcttactggt gctatcaact gaactgcaag tgaaaccagg
11641 gaagtcaact acttactact atgaagggga cccgacatct gcctacaagg aagtcatcgg
11701 ccacaatctt tttgatctca agagaacaag cttcgagaag ctggccaaat taaatctcag
11761 catgtctgta ctcggggctt ggacaagaca caccagcaaa agactattac aagactgtgt
11821 caatatgggt gttaaagagg gcaactggct agttaatgca gacagactag tgagtagcaa
11881 gactggaaac aggtacatac ctggagaggg ccacaccctg caagggagac attatgaaga
11941 actggcgttg gcaagaaaac agatcaataa ctttcaaggg acagacaggt acaatctagg
12001 cccaatagtc aacatggtgt taaggaggct gagagtcatg atgatgaccc tgatagggag
12061 aggggtatga acgcgggcaa cccgggatct ggacccgcca gtaggaccct attgtagata
12121 acactaaatt tttttatttt tttttttttt tttttttatt tatttagata ttattattta
12181 tttatttatt tatttattga atgagtaaga actggtacaa actacctcga gttaccacac
12241 tacactcgtt tttaacagca ctttagctgg aaggaaaatt cctgacgtcc acagttggac
12301 taaggtaatt tctaacggcc c //
```

Classical swine fever virus strain Chinese strain, complete genome

GenBank: AY805221 12310 bp RNA

Country: China

375..12071/codon_start=1

Product: "polyprotein"

Translation : "MELNHFELLYKTNKQKPMGVEEPVYDATGRPLFGDPSEVHPQST
LKLPHDRGRGNIKTTLKNLPRKGDCRSGNHLGPVSGIYVKPGPVFYQDYMGPVYHRAP
LEFFDEVQFCEVTKRIGRVTGSDGKLYHTYVCIDGCILLKLAKRGEPRTLKWIRNFTD
CPLWVTSCSDDGASGSKEKKPDRINKGKLKIAPKEHEKDSRTRPPDATIVVEGVKYQV
KKKGKVKGKNTQDGLYHNKNKPPESRKKLEKALLAWAVIAIMLYQPVEAENITQWNLS
DNGTNGIQHAMYLRGVNRSLHGIWPGKICKGVPTHLATDVELKEIQGMMDASEGTNYT
CCKLQRHEWNKHGWCNWHNIDPWIQLMNRTQADLAEGPPVKECAVTCRYDKDADINVV
TQARNRPTTLTGCKKGKNFSFAGTVIESPCNFNVSVEDTLYGDHECGSLLQDAALYLV
DGMTNTIENARQGAARVTSWLGRQLSTAGKRLEGRSKTWFGAYALSPYCNVTSKIGYI
WYTNNCTPACLPKNTKIIGPGKFDTNAEDGKILHEMGGHLSEFLLLSLVVLSDFAPET
ASALYLILHYVIPQPHDEPEGCDTNQLNLTVELRTEDVIPSSVWNVGEYVCIRPDWWP
YETEVALLFEEAGQVVKLVLRALRDLTRVWNSASTIAFLICLIKVLRGQIVQGVVWLL
LVTGAQGRLACKEDYRYAISSTDEIGLLGAGGLTTTWKEYNHDLQLNDGTVKASCVAG
SFKVTALNVVSRRYLASLHKKALPTSVTFELLFDGTNPSTEEMGDDFRSGLCPFDTSP
VVKGKYNTTLLNGSAFYLVCPIGWTGVIECTAVSPTTLRTEVVKTFRRDKPFPHRMDC
VTTTVENEDLFYCKLGGNWTCVKGEPVVYTGGVVKQCRWCGFDFDGPDGLPHYPIGKC
ILANETGYRIVDSTDCNRDGVVISTEGSHECLIGNTTVKVHASDERLGPMPCRPKEIV
SSAGPVMKTSCTFNYTKTLKNRYYEPRDSYFQQYMLKGEYQYWFDLDATDRHSDYFAE
FVVLVVVALLGGRYVLWLIVTYVVLTEQLAAGLPLGQGEVVLIGNLITHTDIEVVVYF
LLLYLVMRDEPIKKWILLLFHAMTNNPVKTITVALLMVSGVAKGGKIDGGWQRLPGTS
FDIQLALTVIVVAVMLLAKRDPTTVPLVITVATLRTAKMTNGLSTDIAIATVSAALLT
WTYISDYYRYKTWLQYLISTVTGIFLIRVLKGIGELDLHTPTLPSHRPLFFILVYLIS
TAVVTRWNLDIAGLLLQCVPTLLMVFTMWADILTLILILPTYELTKLYYLKEVKIGAE
KGWLWKTNFKRVNDIYEVDQAGEGVYLFPSKQKTSSMTGTMLPLIKAILISCVSNKWQ
FIYLLYLIFEVSYYLHKKIIDEIAGGTNFISRLVAALIEVNWAFDNEEVRGLKKFFLL
SSRVKELIIKHKVRNEVMVRWFGDEEVYGMPKLVGLVKAATLSKNKHCILCTVCEDRE
WRGETCPKCGRFGPPMTCGMTLADFEEKHYKRIFFREDQSEGPVREEYAGYLQYRARG
QLFLRNLPVLATKVKMLLVGNLGTEVGDLEHLGWVLRGPAVCKKVTEHEKCTTSMMDK
LTAFFGVMPRGTTPRAPVRFPTSLLKIRRGLETGWAYTHQGGISSVDHVTCGKDLLVC
DTMGRTRVVCQSNNKMTDESEYGVKTDSGCPEGARCYVFNPEAVNISGTKGAMVHLQK
TGGEFTCVTASGTPAFFDLKNLKGWSGLPIFEASSGRVVGRVKVGKNEDSKPTKLMSG

IQTVSKSTTDLTEMVKKITTMSRGEFRQITLATGAGKTTELPRSVIEEIGRHKRVLVL
IPLRAAAESVYQYMRQKHPSIAFNLRIGEMKEGDMATGITYASYGYFCQMPQPKLRAA
MVEYSFIFLDEYHCATPEQLAIMGKIHRFSENLRVVAMTATPVGTVTTTGQKHPIEEF
IAPDVMKGEDLGSEYLDIAGLKIPVEEMKSNMLVFVPTRNMAVETAKKLKAKGYNSGY
YYSGEDPSNLRVVTSQSPYVVVATNAIESGVTLPDLDVVVDTGLKCEKRIRLSPKMPF
IVTGLKRMAVTIGEQAQRRGRVGRVKPGRYYRSQETPVGSKDYHYDLLQAQRYGIEDG
INITKSFREMNYDWSLYEEDSLMITQLEILNNLLISDELPMAVKNIMARTDHPEPIQL
AYNSYETQVPVLFPKIKNGEVTDSYDNYTFLNARKLGDDVPPYVYATEDEDLAVELLG
LDWPDPGNQGTVETGRALKQVVGLSTAENALLVALFGYVGYQALSKRHIPVVTDIYSI
EDHRLEDTTHLQYAPNAIKTEGKETELKELAQGDVQRCVEAMTNYAREGIQFMKSQAL
KVKETPTYKETMDTVTDYVKKFMEALADSKEDIIKYGLWGTHTALYKSISARLGGETA
FATLVVKWLAFGGESIADHVKQAATDLVVYYIINRPQFPGDTETQQDGRKFVASLLAS
ALATYTYKSWNYNNLSKIVEPALATLPYAATALKLFAPTRLESVVILSTAIYKTYLSI
RRGKSDGLLGTGVSAAMEIMSQNPVSVGIAVMLGVGAVAAHNAIEASEQKRTLLMKVF
VKNFLDQAATDELVKESPEKIIMALFEAVQTVGNPLRLVYHLYGVFYKGWEAKELAQR
TAGRNLFTLIMFEAVELLGVDSEGKVRQLSSNYILELLYKFRDSIKSSVREMAISWAP
APFSCDWTPTDDRIGLPQDNFHQVETKCPCGYKMKAVKNCAGELRLLEEEGSFLCRNK
FGRGSRNYRVTKYYDDNLLEIKPVIRMEGHVELYYKGATIKLDFNNSKTILATDKWEV
DHSTLVRVLKRHTGAGYHGAYLGEKPNHKHLIERDCATITKDKVCFLKMKRGCAFTYD
LSLHNLTRLIELVHKNNLEDKEIPAATVTTWLAYTFVNEDIGTIKPAFGEKVTLEMQE
EITLQPAVVVDTTDVAVTVVGEAPTMTTGETPTVFASSGSGLKSQQVLKLGVGEGQYP
GTNPQRASLHEAIQGADERPSVLILGSDKATSNRVKTAKNVKVYRGRDPLEVRDMMRR
GKILVVALSRVDNALLKFVDYKGTFLTREALEALSLGRPKKKNITKAEAQWLLCPEDQ
MEELPDWFAAGEPIFLEANIKHDRYHLVGDIATIKEKAKQLGATDSTKISKEVGAKVY
SMKLSNWVMQEENKQGNLTPLFEELLQQCPPGGQNKTAHMVSAYQLAQGNWMPTSCHV
FMGTISARRTKTHPYEAYVKLRELVEEHKMKTLCPGSSLGKHNDWIIGKIKYQGNLRT
KHMLNPGKVAEQLCREGHRHNVYNKTISSVMTATGIRLEKLPVVRAQTDPTNFHQAIR
DKIDKEENLQTPGLHKKLMEVFNALKRPELESSYDAVEWEELERGINRKGAAGFFERK
NIGEILDSEKNKVEEIIDNLKKGRNIKYYETAIPKNEKRDVNDDWTAGDFVDEKKPRV
IQYPEAKTRLAITKVMYKWVKQKPVVIPGYEGKTPLFQIFDKVKKEWDQFQNPVAVSF
DTKAWDTQVTTKDLELIRDIQKYYFKKKWHKFIDTLTTHMSEVPVISADGEVYIRKGQ
RGSGQPDTSAGNSMLNVLTMVYAFCEATGVPYKSFDRVAKIHVCGDDGFLITERALGE
KFASKGVQILYEAGKPQKITEGDKMKVAYQFDDIEFCSHTPIQVRWSDNTSSYMPGRN
TTTILAKMATRLDSSGERGTIAYEKAVAFSFLLMYSWNPLIRRICLLVLSTELQVKPG
KSTTYYYEGDPISAYKEVIGHNLFDLKRTSFEKLAKLNLSMSVLGAWTRHTSKRLLQD

```
CVNIGVKEGNWLVNADRLVSSKTGNRYIPGEGHTLQGRHYEELVLARKQINNFQGTDR
YNLGPIVNMVLRRLRVMMMTLIGRGA"

ORIGIN
        1 gtatacgagg ttagttcatt ctcgtataca cgattggaca aatcaaaatt ataatttggt
       61 tcagggcctc cctccagcga cggccgaact gggctagcca tgcccatagt aggactagca
      121 aaacggaggg actagccata gtggcgagct ccctgggtgg tctaagtcct gagtacagga
      181 cagtcgtcag tagttcgacg tgagcagaag cccacctcga gatgctacgt ggacgagggc
      241 atgcccaaga cacaccttaa ccctagcggg ggtcgctagg gtgaaatcac accacgtgat
      301 gggagtacga cctgataggg cgctgcagag gcccactatt aggctagtat aaaaatctct
      361 gctgtacatg gcacatggag ttgaatcact ttgaactttt atacaaaaca aacaaacaaa
      421 aaccaatggg agtggaggaa ccggtgtacg atgccacggg gagaccattg ttcggagacc
      481 cgagtgaggt acacccacaa tcaacactga agctaccaca tgataggggt agaggcaaca
      541 ttaaaacaac actgaagaac ctacctagga aaggcgactg caggagcggc aaccatctag
      601 gcccggtcag tgggatatat gtaaaacccg gccctgtctt ttaccaggac tacatgggcc
      661 cggtctacca tagagcccct ctggagtttt ttgacgaagt gcagttctgc gaggtgacca
      721 aaaggatagg tagggtgaca ggtagcgacg gaaagcttta ccatacatat gtgtgcatcg
      781 atggctgcat actgctgaag ctggccaaga ggggtgagcc aagaaccctg aagtggatta
      841 gaaatttcac cgactgtcca ttgtgggtta ccagttgctc cgatgatggc gcaagtggga
      901 gtaaagagaa gaagccagat aggatcaaca aaggcaaatt aaaaatagcc ccaaaagagc
      961 atgagaagga cagcagaact aggccacctg acgctacgat cgtggtggaa ggagtaaaat
     1021 accaggtcaa aaagaaaggt aaagttaaag gaaagaatac ccaagacggc ctgtaccaca
     1081 acaagaataa accaccagaa tccaggaaga aattagaaaa agccctattg gcatgggcgg
     1141 tgatagcaat tatgttgtac caaccagttg aagccgaaaa tataactcaa tggaacctga
     1201 gtgacaacgg cactaatggt atccagcatg ctatgtacct tagaggggtt aacagaagct
     1261 tgcatgggat ctggccgggg aaaatatgca aaggagtccc aacccacctg gccacagacg
     1321 tggagctgaa agaaatacag ggaatgatgg atgccagcga ggggacaaac tatacgtgct
     1381 gtaagttaca gagacatgaa tggaacaaac atggatggtg taactggcac aatatagacc
     1441 cctggataca gctgatgaat agaacccaag cagacttggc agaaggccct ccggtcaagg
     1501 agtgcgctgt gacttgcagg tacgataaag atgctgacat caacgtggtc acccaggcta
     1561 gaaacaggcc aacaaccctg accggctgca agaaagggaa aaatttttct tttgcgggta
     1621 cagttataga gagcccatgt aatttcaatg tttccgtgga ggataccttg tatggggatc
     1681 atgagtgcgg cagtttactc caggacgcag ctctgtacct agtagatgga atgaccaaca
     1741 ctatagagaa tgccagacag ggagcagcga gggtgacatc ttggctcggg aggcaactca
     1801 gcactgctgg gaagaggttg gagggtagaa gcaaaacctg gtttggcgct tatgccctat
     1861 cgccttactg taatgtaaca agcaagatag ggtacatatg gtacactaac aactgcaccc
     1921 cagcttgcct ccccaaaaac acaaagataa taggccctgg taaatttgac accaatgcag
     1981 aagacggaaa gattctccat gagatggggg gccacctatc agaatttctg ctgctttctc
     2041 tggttgttct gtctgacttc gcccctgaaa cagccagcgc gttatacctc attttgcact
     2101 acgtgattcc tcaaccccat gatgaacctg aaggctgcga tacgaaccag ctgaatctaa
     2161 cagtagaact caggactgaa gacgtaatac cgtcatcagt ctggaatgtt ggtgaatatg
     2221 tgtgtattag accagactgg tggccatatg aaaccgaggt ggctctgtta tttgaagagg
```

```
2281 caggacaggt cgtaaagtta gtcttacggg cgctgaggga tttgactagg gtctggaata
2341 gcgcatcaac cattgcattc ctcatctgct tgataaaagt attaagggga cagatcgtgc
2401 aaggtgtggt atggctgtta ctagtaactg gggcacaagg ccggctagcc tgcaaggaag
2461 attacaggta cgcaatatcg tcaaccgatg agatagggct acttggggcc ggaggtctca
2521 ccaccacctg gaaggaatac aaccacgatt tgcaactgaa tgacgggacc gtcaaggcca
2581 gttgcgtggc aggttccttt aaagtcacag cacttaatgt ggtcagtagg aggtatttgg
2641 cgtcattgca taagaaggct ttacccactt ccgtgacatt cgagctcctg ttcgacggga
2701 ccaacccatc aactgaggaa atgggagatg acttcaggtc cgggctgtgc ccgtttgata
2761 cgagtcctgt tgttaaggga aagtacaata cgaccttgtt gaacggtagt gctttctatc
2821 ttgtctgccc aatagggtgg acgggtgtca tagagtgcac agcagtgagc ccaacaactc
2881 tgaggacaga agtggtaaag accttcagga gagacaagcc ctttccgcac agaatggatt
2941 gtgtgaccac cacagtggaa aatgaagatt tattctattg taagttgggg ggcaactgga
3001 catgtgtgaa aggcgagcca gtggtctaca caggggggt agtaaaacaa tgtagatggt
3061 gtggcttcga cttcgatggg cctgacggac tcccgcatta ccccataggt aagtgcattt
3121 tggcaaatga gacaggttac agaatagtag attcaacgga ctgtaacaga gatggcgttg
3181 taatcagcac agaggggagt catgagtgct tgatcggtaa cacgactgtc aaggtgcatg
3241 catcagatga aagactgggc cctatgccat gcagacctaa agagattgtc tctagtgctg
3301 gtcctgtaat gaaaacctcc tgtacattca actacacaaa aactttgaag aacaggtact
3361 atgagcccag ggacagctac ttccagcaat atatgcttaa gggtgagtat cagtactggt
3421 ttgacctgga tgcgactgac cgccactcag attacttcgc agaatttgtt gtcttggtgg
3481 tggtagcact gttaggagga agatatgtcc tgtggctgat agtgacctac gtagttctaa
3541 cagaacaact cgccgctggt ttaccattgg gccagggtga ggtagtgttg atagggaact
3601 taatcaccca cacagacatt gaggtcgtag tatattttt actactctat ttggtcatga
3661 gggatgaacc tataaagaaa tggatactgc tgctgttcca tgctatgact aacaatccag
3721 tcaagaccat aacagtggca ttgctcatgg ttagtggagt tgccaagggt ggaaagatag
3781 atggcggttg gcagcgactg ccggggacca gctttgacat ccaactcgcg ctgacagtta
3841 tagtagtcgc tgtgatgttg ctggcaaaga gagatccgac tacggtcccc ttggttataa
3901 cagtggcgac cctgagaaca gctaagatga ctaacggact tagtacggat atagccatag
3961 ccacagtgtc agcagcgttg ctaacctgga cctacattag tgactattac agatacaaga
4021 cctggctaca gtaccttatc agcacagtga caggtatctt tttaataagg gtactgaagg
4081 gaataggtga gttggattta cacactccga ccttgccatc tcatagaccc ctctttttca
4141 ttctcgtgta ccttatttcc actgcagtgg taacgagatg gaatctggac atagctggat
4201 tgctgttgca gtgtgttcca acccttttga tggtttttac gatgtgggca gacattctca
4261 ccctgatcct catactgccc acttacgagt taacgaagct atattatctt aaggaagtga
4321 agattggggc agaaaagggc tggttatgga agaccaactt caagagggta aacgacatat
4381 acgaagttga ccaagctggt gaaggggtat acctattccc gtcaaaacaa aagacaagtt
4441 caatgacagg caccatgttg ccattgatca aagccatact tatcagctgc gtcagtaata
4501 agtggcagtt catatatcta ctgtacttga tatttgaagt atcttactac ctccacaaga
4561 agatcataga tgaaatagca ggagggacca acttcatctc aagacttgta gccgctttga
4621 tcgaagtcaa ttgggccttt gacaacgaag aagttagggg tttgaagaag ttcttcctgt
4681 tgtctagtag ggttaaagaa ctgatcatca aacacaaagt gaggaatgaa gtaatggtcc
4741 gctggtttgg tgacgaagag gtctatggga tgccgaagtt ggttggccta gtcaaggcag
4801 caacattgag taaaaataaa cattgtattt tgtgcaccgt ctgtgaagac agagagtgga
```

```
4861 gaggagaaac ctgcccaaaa tgcgggcgtt ttgggccacc aatgacctgt ggcatgaccc
4921 tagccgactt tgaagagaaa cactataaga ggatcttttt tagagaggat caatcagaag
4981 ggccggttag ggaggagtac gcagggtatc tgcaatatag agccagaggg caattgttcc
5041 tgaggaatct cccagtgcta gcaacaaaag ttaagatgct cctggtcggc aatcttggga
5101 cggaggtggg agatttggaa caccttggct gggtgctcag agggcctgcc gtttgcaaga
5161 aggtcactga acatgagaaa tgtaccacat ccatgatgga caaattgact gctttttttcg
5221 gtgttatgcc gaggggcacc acacctagag cccctgtgag attccctacc tctctcttaa
5281 agataagaag ggtttggaa actggctggg cgtacacaca ccaaggtggc atcagttcag
5341 tggaccatgt cacttgtgga aaagacttac tggtatgtga cactatgggc cggacaaggg
5401 ttgtttgcca gtcaaataat aagatgacag atgagtccga gtatggagtt aaaactgatt
5461 ccggatgccc ggaaggagct aggtgttatg tgttcaaccc agaggcagtt aacatatcag
5521 ggactaaagg agccatggtc cacttacaaa aaactggagg agaattcacc tgtgtgacag
5581 catcaggaac tccggctttc tttgatctta aaaaccttaa aggctggtca gggctaccga
5641 tatttgaggc atcaagtgga agggtagtcg gcagggtcaa ggtcggtaag aatgaggact
5701 ctaaaccaac caagcttatg agtggaatac aaacagtttc caaaagtacc acagacttga
5761 cagaaatggt aaagaaaata acgaccatga gcaggggaga attcagacaa ataacccttg
5821 ctacaggtgc cggaaaaacc acggaactcc ctaggtcagt catagaagag atagggaggc
5881 ataagagagt cttggtcttg atccctctga gggcggcagc agagtcagta taccaatata
5941 tgagacaaaa gcatccaagc atcgcattta acctgaggat aggggagatg aaggaagggg
6001 acatggccac agggataact tatgcttcat acggttactt ctgtcagatg ccacaaccta
6061 agttgcgagc cgcaatggtt gagtactcct tcatattcct tgatgagtac cactgtgcca
6121 cccctgaaca attggctatc atgggaaaga ttcacagatt ttcagagaac ctgcgggtgg
6181 tggccatgac cgcaacacca gtaggcacgg taacgaccac agggcagaaa caccctatag
6241 aagaattcat agccccagat gtgatgaaag gggaagactt aggttcagag tacttggaca
6301 ttgctggact aaagatacca gtagaggaga tgaagagcaa tatgctggtt tttgtgccca
6361 ccaggaacat ggcagtggag acagcaaaga aattgaaagc taagggttat aactcaggct
6421 actattatag tggtgaggat ccatctaacc tgagggtggt aacatcgcag tccccgtacg
6481 tggtggtggc aaccaacgcg atagaatcag gtgttactct cccggacttg gatgtggttg
6541 tcgatacagg gcttaagtgt gaaaagagaa tacggctgtc acctaagatg cctttcatag
6601 tgacgggcct gaagagaatg gctgtcacga ttggggaaca agcccagaga agggggagag
6661 ttgggagagt aaagcctgga agatactaca ggagtcaaga aacccccgtt ggttctaaag
6721 attaccatta tgatttactg caagcacaga ggtacggtat tgaagatggg ataaacatca
6781 ccaaatcctt tagagagatg aactatgatt ggagccttta tgaggaggac agtctgatga
6841 ttacacaatt ggaaattctt aataatttgt tgatatcaga tgaactacca atggcagtaa
6901 aaaatataat ggccaggact gaccacccag aaccaattca gctggcgtac aacagctacg
6961 aaacacaagt gccagtgcta ttcccaaaaa taaagaatgg agaggtgact gacagttacg
7021 ataactatac cttcctcaac gcaagaaaat taggggatga tgtacccct tacgtgtatg
7081 ccacagagga tgaggactta gcggtagagc tgctgggctt agactggccg gaccctggaa
7141 accaagggac cgtagagact ggcagagcac taaaacaggt agttggtcta tcaacagctg
7201 agaatgccct gttagtagcc ttattcggct acgtaggata tcaggcgctt tcaaagaggc
7261 atataccagt agtcacagac atatactcaa ttgaagatca caggttggaa gacaccacac
7321 acctacagta cgccccaaat gctatcaaga cggaggggaa ggagacagaa ttgaaggagc
7381 tagctcaggg ggatgtgcag agatgtgtgg aagccatgac caattatgca agagagggta
```

7441 tccaatttat gaagtctcag gcactgaagg tgaaggaaac ccctacttac aaggagacaa

7501 tggacactgt gacggactat gtaaagaaat tcatggaggc gctggcagac agtaaagaag

7561 acatcataaa atatgggctg tgggggacgc acacagcctt atataagagc atcagtgcca

7621 ggcttggggg tgagactgcg ttcgctaccc tggtagtgaa gtggctggca tttgggggtg

7681 aatcaatagc agaccatgtc aaacaagcgg ccacagactt ggtcgtttac tatatcatca

7741 acagacctca gttcccagga gacacggaga cacaacaaga cggaaggaaa tttgtggcca

7801 gcctactggc ctcagctcta gctacttaca catacaaaag ctggaattac aataacctgt

7861 ccaagatagt tgaaccggct ttggccactc tgccctatgc cgccacagct ctcaaattat

7921 ttgcccccac ccgattagag agcgttgtca tattaagtac cgcaatctac aagacctacc

7981 tatcaatcag gcgcggaaaa agcgatggtt tgctaggcac gggggttagt gcggctatgg

8041 agatcatgtc acaaaatcca gtatccgtgg gcatagcagt catgctaggg gtaggggccg

8101 tggcagccca caatgcaatc gaagccagtg agcagaaaag aacactactc atgaaagtct

8161 ttgtaaagaa cttcttggac caggcagcca cagatgaatt agtcaaggag agtcctgaga

8221 aaataataat ggctttgttt gaggcagtgc agacagtcgg caaccctctt agacttgtat

8281 accaccttta tggagttttt tataaggggt gggaggcaaa agagttggcc caaaggacag

8341 ccggtaggaa ccttttcact ttaataatgt tcgaggctgt ggaactgctg ggagtagaca

8401 gtgaaggaaa ggtccgccag ctatcaagta attacatact agagcttttg tataagttcc

8461 gtgacagtat caagtctagc gtgagggaga tggcaatcag ctgggcccct gcccctttca

8521 gttgtgattg gacaccgacg gatgacagaa tagggctccc ccaagacaac ttccaccaag

8581 tggagacgaa atgcccctgt ggttacaaga tgaaggcagt taagaattgt gctggagaac

8641 tgagactctt ggaggaggaa ggttcatttc tctgcagaaa taaattcggg agaggttcac

8701 ggaactacag ggtgacaaaa tattatgatg acaacctatt agaaataaag ccagtgataa

8761 gaatggaagg gcatgtggaa ctctactaca agggggccac catcaaactg gatttcaaca

8821 acagcaaaac aatattggca accgataaat gggaggttga tcactccact ctggtcaggg

8881 tgctcaagag gcacacaggg gctggatatc atggggcata cctgggcgaa aaaccgaacc

8941 acaaacacct gatagagagg gactgtgcaa ccatcaccaa agataaggtc tgtttctca

9001 aaatgaagag agggtgcgca tttacttatg acttatccct tcacaacctt acccgactga

9061 ttgaattggt acacaagaat aacttggaag acaaagagat tcccgctgct acggttacaa

9121 cctggctggc ttacacattt gtaaatgagg atatagggac cataaaacca gccttcgggg

9181 agaaagtaac gctggagatg caggaggaga taaccttgca gcctgccgtt gtggtggata

9241 caacagacgt agccgtgact gtggtagggg aagcccccac tatgactaca ggggagacac

9301 cgacagtgtt cgccagctca ggttcaggcc tgaaaagcca acaagttttg aaactagggg

9361 taggtgaagg ccaatatcca gggactaatc cacagagggc aagcctgcac gaagccatac

9421 aaggtgcaga tgagaggccc tcggtgctga tattggggtc tgataaagcc acctctaata

9481 gagtgaagac tgcaaagaat gtaaaggtat acagaggcag ggacccacta gaagtgagag

9541 atatgatgag gaggggaaag atcctggtcg tagccctgtc tagggttgat aatgctctat

9601 tgaaatttgt tgactacaaa ggcacctttc taactaggga ggctctagag gcattaagtt

9661 tgggcaggcc taaaaagaaa aacataacca aggcagaagc gcagtggttg ctgtgccccg

9721 aggaccaaat ggaagagcta cccgactggt tcgcagccgg ggaacccata tttttagagg

9781 ccaacattaa acatgacagg taccatctgg tgggggatat agctaccatc aaggaaaaag

9841 ccaaacagtt gggggctaca gactccacaa agatatctaa ggaggttggt gccaaagtgt

9901 attctatgaa actgagtaat tgggtgatgc aagaagaaaa taaacagggc aacctgaccc

9961 ccttgtttga agagctcctg caacagtgtc cacccggggg ccagaacaaa accgcacaca

```
10021 tggtctctgc ttaccaactg gctcaaggga actggatgcc gaccagctgc catgttttca
10081 tggggaccat atctgccagg agaaccaaga cccacccata tgaagcatac gtaaagttaa
10141 gggagttggt agaggaacac aagatgaaaa cactgtgtcc cggatcaagc ctgggtaagc
10201 acaacgattg gataattgga aaaattaaat accagggaaa cctgaggacc aaacacatgt
10261 tgaaccccgg caaggtggca gagcaactgt gcagagaggg acacagacac aatgtgtata
10321 acaagacaat aagctcagta atgacagcta ctggtatcag gttggagaag ttgcccgtgg
10381 ttagggccca gacagaccca accaacttcc accaagcaat aagggataag atagacaagg
10441 aagagaacct acaaaccccg ggtttacata agaaattaat ggaagttttc aacgcattga
10501 aacgacccga gttagagtcc tcctacgacg ccgtggaatg ggaggaactg gagagaggaa
10561 taaacaggaa gggtgctgct ggtttttttcg aacgcaaaaa tataggggaa atattggatt
10621 cagagaaaaa taaagtcgaa gagattattg acaatctgaa aaaaggtaga aacattaaat
10681 attatgaaac cgcgatccca aagaatgaga agagggacgt caacgatgac tggaccgccg
10741 gtgatttcgt ggacgagaag aaacctagag tcatacaata ccctgaagca aaaacaagac
10801 tggccatcac caaggtgatg tataagtggg tgaagcagaa gccagtagtt atacccgggt
10861 atgaagggaa gacacctcta ttccaaattt ttgacaaagt gaagaaggaa tgggatcaat
10921 ttcaaaatcc agtggcagtg agcttcgaca ctaaggcgtg ggacacccag gtaaccacaa
10981 aagatttgga gctgataagg gacatacaaa agtattattt caagaagaaa tggcacaaat
11041 ttattgacac cctgaccacg catatgtcag aagtacccgt gattagtgct gatggggaag
11101 tatacataag gaaagggcaa agaggcagtg gacaacctga cacaagtgcg ggcaacagca
11161 tgctaaatgt cttaacaatg gtttacgcct tctgcgaggc cacaggagta ccctacaaga
11221 gctttgacag ggtggcaaaa attcatgtgt gcggggatga tggcttcctg atcacagaaa
11281 gagctctcgg tgagaaattt gcaagtaagg gagtccagat cctttatgaa gctgggaagc
11341 cccagaagat cactgaaggg gacaaaatga aagtggccta ccaatttgat gatattgagt
11401 tttgctccca tacaccaata caagtaagat ggtcagataa cacttctagt tacatgccgg
11461 ggagaaatac aaccacaatc ctggcaaaga tggccacgag gttagattcc agcggtgaaa
11521 ggggtaccat agcatatgag aaagcagtag catttagctt cctgctgatg tattcctgga
11581 acccactaat tagaaggatc tgcttactgg tgctatcaac tgaactgcaa gtgaaaccag
11641 ggaagtcaac tacttactat tatgaaggag acccgatatc tgcctacaag gaagtcatcg
11701 gccacaatct ttttgatctt aagagaacaa gctttgagaa gctggccaag ttaaatctta
11761 gcatgtctgt actcggggcc tggactagac acaccagtaa aagactacta caagactgtg
11821 tcaatatagg tgttaaagag ggcaactggc tagtcaatgc agacagacta gtaagtagca
11881 agaccgggaa taggtacata cccggagagg gtcacaccct gcaaggaaga cattatgaag
11941 aactggtgtt ggcaagaaaa cagatcaaca actttcaagg gacagacagg tacaacctag
12001 gcccaatagt caacatggtg ttaaggaggc tgagagtcat gatgatgacg ctgataggga
12061 gaggggcatg agcgcgggta acccgggatc tgaacccgcc agtaggaccc tattgtagat
12121 aacactaatt ttcttttttc tttttatttt atttagatat tattatttat ttatttattt
12181 atttattgaa tgagtaagaa ctggtataaa ctacctcaag ttaccacact acactcattt
12241 ttaacagcac tttagctgga aggaaaattc ctgacgtcca cagttggact aaggtaattt
12301 ctaacggccc //
```

BIOGRAPHY

Name	Amonrat Dechpan
Home Address	95 Puttaraksa Road, Huaichorakhe, Muang, Nakhonpathom, 73000

Education Background

2007	Bachelor degree of science (B.Sc), Department of Biology, Faculty of Science, Silpakorn University
2011	Master degree of science (M.Sc), Department of Biology, Faculty of Science, Silpakorn University