Prediction of TGEV Spike Protein Secondary Structure and B Cell Epitopes*

Tian-fei Yu[1], Shu-li Shao[1], Xing-jun Xu[1], Jian-wei Lv[1], and Ming Li[2]

[1] College of Life Science and Agriculture Forestry University of Qiqihar Qiqihar, Heilongjiang Province, China
[2] College of Computer and Control Engineering, University of Qiqihar, Qiqihar, Heilongjiang Province, China
{yutianfei2001,shshl32}@163.com, fiona001lee@yahoo.cn

Abstract. The spike gene of TGEV TH98 strain was translated into amino sequence by Editseq. The secondary structure and B cell epitope of spike protein of TGEV TH98 strain were predicted by Protean. Combining the results according to these methods, the spike protein of TGEV TH98 strain has complicated secondary structure. There are several epitopes of the B-cells in spike protein, including 43-56aa, 97-104aa, 117-128aa, 132-173aa, 238-257aa, 391-398aa, 535-706aa, 779-799aa, 918-987aa, 1165-1200aa, 1257-1266aa and 1430-1446aa.

Keywords: Lasergene, TGEV, Spike protein, Secondary structure, Epitope.

1 Introduction

Lasergene was exploited by DNASTAR, Inc. It was committed to providing innovative and easy to use desktop software tools for today's life researchers. The modules of Lasergene were: SeqBuilder, visualization and sequence editing Video; SeqMan Pro, sequence assembly and SNP discovery Video; MegAlign, sequence alignment; PrimerSelect, oligo primer design; Protean, protein structure analysis & prediction; GeneQuest, gene finding; EditSeq, utility for importing unusual file types. The Data Manager enables data integration between the Lasergene modules so that edits, additions and deletions made to a sequence in one module will synchronize and automatically update when opened in most other modules. Bioinformatics researchers paid close attention to the software from the day of its exploitation. It has been widely used in these days.

Transmissible gastroenteritis (TGE) is a highly contagious viral disease of swine characterized by vomiting, diarrhea, and dehydration. Its causative agent is transmissible gastroenteritis virus (TGEV), considered the principal etiologic agent responsible for dramatic outbreaks of diarrhea and high mortality of newborn pigs, in which mortality approaches 100% [1].

* This work is supported by National High-Tech Research and Development Plan of China #2008AA10Z326 to Y Dai and National Natural Science Foundation of China #20776112 to S Jia.

E. Zhu and S. Sambath (Eds.): Information Tech. and Agricultural Eng., AISC 134, pp. 57–60.
springerlink.com

TGEV spike protein is a major viral antigen and plays a crucial role in the induction of neutralizing antibodies. Its binding to the cellular receptor porcine aminopeptidase N (pAPN) is required or the initial stage of infection. Therefore, the spike protein of TGEV is a good target for vaccine design and antigen detection [2-6].

In this research, the secondary structure and B cell epitopes of spike protein were predicted by Protean. The work settled the foundation for diagnose of TGEV and functional analysis of related protein.

2 Materials and Methods

2.1 Prediction of Secondary Structure

Nucleotide sequence of TGEV TH98 spike gene (GenBank number: AF494337) was translated into amino acid sequence by EditSeq, one module of Lasergene. The Gamier-Robson method, Chou-Fasman method and Karplus-Schulz method were used to predict the secondary structure of spike protein.

2.2 Prediction of B Cell Epitopes

The hydrophilicity and surface probability of spike protein were analyzed by Kyte-Doolittle method and Emini method. Antigenic index was showed by Jameson-Wolf method. Finally, synthetic conclusion was got to evaluate the potential epitopes of spike protein.

3 Results

3.1 Prediction of Secondary Structure

Alpha helical structure appeared at 312-316aa, 890-902aa, 910-914aa, 1094-1099aa, 1141-1145aa, 1173-1177aa, 1347-1352aa, 1440-1444aa according to Gamier-Robson method and Chou-Fasman method. Beta folds were well-distributed according to these methods. Among the beta fold units there are several turn regions.

3.2 Hydrophilicity of Spike Protein

The hydrophobicity of spike protein was analyzed by Kyte-Doolittle method. The results showed there were many hydrophilicity plots, 19-32aa, 45-51aa, 67-76aa, 87-105aa, 119-125aa, 136-172aa, 194-205aa, 236-247aa, 325-335aa, 347-354aa, 391-402aa, 481-490aa, 538-546aa, 554-568aa, 587-596aa, 613-621aa, 640-651aa, 668-674aa, 683-693aa, 699-705aa, 765-796aa, 819-829aa, 862-874aa, 918-935aa, 946-959aa, 973-987aa, 1082-1091aa, 1134-1146aa, 1164-1198aa, 1236-1243aa, 1256-1277aa, 1306-1326aa, 1342-1358aa, 1378-1389aa, 1429-1444aa. These regions could be potential epitopes.

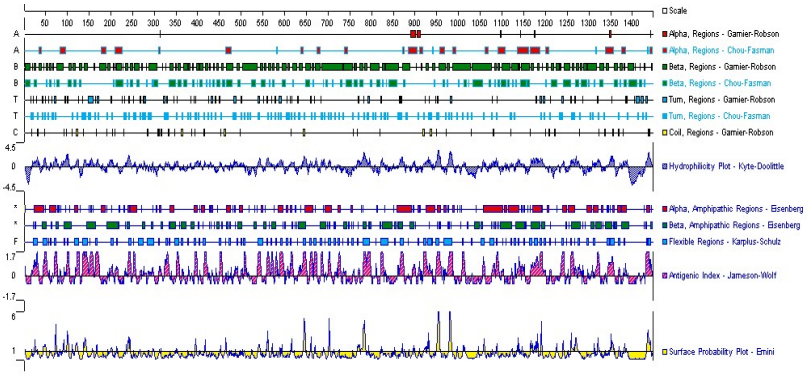

Fig. 1. Prediction of TGEV spike protein secondary structure and B cell epitopes by Lasergene

3.3 Surface Probability of Spike Protein

The surface probability plots of spike protein appeared at 43-50aa, 97-103aa, 117-124aa, 136-145aa, 234-243aa, 557-563aa, 641-647aa, 765-786aa, 949-958aa, 975-985aa, 1163-1192aa, 1257-1273aa, 1347-1356aa, 1432-1441aa. The probability of other regions was low.

3.4 Coil Regions of Spike Protein

There were a lot of coil regions in TGEV TH98 spike protein. These coil regions distributed evenly. The extents of coil regions in N terminal were wider.

3.5 Prediction of B Cell Epitopes

According to Jameson-Wolf method and former results, the potential epitopes at 43-56aa, 97-104aa, 117-128aa, 132-173aa, 238-257aa, 391-398aa, 535-706aa, 779-799aa, 918-987aa, 1165-1200aa, 1257-1266aa and 1430-1446aa. Some regions showed antigenicity by Jameson-Wolf method only, but other indexes were lower. Such regions were questionable.

4 Discussion and Conclusion

Viruses are major factors of human infectious diseases. Understanding of the structure-function correlation in viruses is important for the identification of potential anti-viral inhibitors and vaccine targets. In virology research, virus-related databases and bioinformatic analysis tools are essential for discerning relationships within complex datasets about viruses and host-virus interactions. Bioinformatic analyses on viruses include the identification of open reading frames, gene prediction, homology searching, sequence alignment, and motif and epitope recognition. The predictions of features such as transmembrane domains, glycosylation sites, and protein secondary and tertiary structure are important for analyzing the structure-function relationship of proteins encoded in viral genomes. Biochemical pathway analysis can help elucidate information at the biological systems level [7].

The secondary structure of protein had great influence on epitope. The chemical bond energy of alpha helical and beta fold was high. They could maintain fast higher structure of protein and could not be the epitopes. Meanwhile, the turn and coil regions were noncohesive structure and easily twist. The structures usually hovered over and displayed on the surface of protein. Therefore, antigenic indexes of these domains antigenic index were high [8]. In this research, the N terminal of spike protein had more turns and high indexes of hydrophilicity and surface probability. The characteristics of such regions could be the indexes of potential epitopes. In a word, the potential epitopes of spike protein were at 43-56aa, 97-104aa, 117-128aa, 132-173aa, 238-257aa, 391-398aa, 535-706aa, 779-799aa, 918-987aa, 1165-1200aa, 1257-1266aa and 1430-1446aa. The 43-56aa domains were in coincidence with already definite epitope [9].

Because all the theories of predictable methods were based on protein primary structure, the intermolecular force was neglected. The prediction of conformational epitopes was not suitable by these methods. Up to now, there were no better effective ways to predict conformational epitopes. Even the predicted epitopes presented on the surface of protein actually, but, these epitopes could be barricaded in posse by the tertiary structure.

References

1. Schwegmann-Wessels, C., Herrler, G.: Transmissible gastroenteritis virus infection: a vanishing spectre. Dtsch Tierarztl Wochenschr 113(4), 157–159 (2006)
2. Ballesteros, M.-L., Sanshes, C., Enjuanes, L.: Two amino changes at the N-terminal of TGEV spike protein result in the loss of enteric tropism. Virology 227(2), 378–388 (1997)
3. Izeta, A., et al.: The spike protein of transmissible gastroenteritis coronavirus controls the tropism of pseudorecombinant virions engineered using synthetic minigenomes. Adv. Exp. Med. Biol. 440, 207–214 (1998)
4. Sanchez, C.-M., et al.: Targeted recombination demonstrates that the spike gene of transmissible gastroenteritis coronavirus is a determinant of its enteric tropism and virulence. J. Virol. 73(9), 7607–7618 (1999)
5. Krempl, C., et al.: Characterization of the sialic acid binding activity of transmissible Gastroenteritis coronavirus by analysis of haemagglutination-deficient mutants 81(2), 489–496 (2000)
6. Schwegmann-Wessels, C., Zimmer, G., Schroder, B., Breves, G., Herrler, G.: Binding of transmissible gastroenteritis coronavirus to brush border membrane sialoglycoproteins. J. Virol. 77(21), 11846–11848 (2003)
7. Persson, B.: Bioinformatics in protein analysis. EXS 88, 215–231 (2000)
8. Poland, G.-A., Ovsyannikova, I.-G., Johnson, K.-L., Naylor, S.: The role of massspectrometry in vaccine development. Vaccine 19(17-19), 2692–2700 (2001)
9. Gebauer, F., et al.: Residues involved in the antigenic sites of transmissible gastroenteritis virus S glycoprotein. Virology 183(1), 225–238 (1991)