Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Medicine

journal homepage:www.elsevier.com/locate/apjtm

Document heading doi:

Bioinformatics analysis on ORF1 protein of Torque teno virus (SANBAN isolate)

Chen Jinling[1#], Zhu Dandan[1#], Shen Pei[2], Sun Wei[1], Xiao Gengfu[1], Duan Yinong[1*], Zhu Ying[1]

[1]*Department of Pathogen Biology, School of Medicine, Nantong University, Nantong, Jiangsu, 226001, People's Republic of China*
[2]*Zhenjiang Key Laboratory of Medical Immunology, School of medical science and laboratory medice, Jiangsu University, Zhenjiang, Jiangsu, China*

ARTICLE INFO

Article history:
Received 9 July 2011
Received in revised form 11 September 2011
Accepted 15 October 2011
Available online 20 November 2011

Keywords:
TTV
ORF1
Bioinformatics

ABSTRACT

Objective: To analyze the sequence of ORF1 protein of Torque teno virus to prepare for the future hybrid experiments. **Methods:** The sequence of ORF1 protein of Torque teno virus was analyzed by bioinformatics using some web tools. **Results:** The most likely cleavage site was between position 14aa and 15aa and signal peptide may be position 1aa–14aa. Two possible transmembrane helices from inside to outside and three possible transmembrane helices from outside to inside were found. The position 509 (NKTN) was the potential *N*-glycosylation site. The speculative molecular weight of TTV ORF1 protein, which may be a kind of unstable protein was 88 705.7 Da. 1aa–91aa and 278aa–361aa were localized in non-regular secondary structure region. **Conclusions:** TTV ORF1 protein may be a nuclear protein which contains two non-regular secondary structure region. 265aa to 486aa and 510aa to 679aa may be the two approciate fragments to construct the plasmids, which would be prepared for the future hybrid experiments to study the functional positions of the protein and the interactions between TTV and its hosts. Bioinformatics analysis would possibly make it easier to study the protein's function.

1. Introduction

Torque teno virus (or transfusion-transmitted virus, TTV) was first identified by representational-difference display analysis of the serum from a Japanese patient, who had acute posttransfusion hepatitis of unknown etiology[1]. It is a small, nonenveloped, single-stranded, and circular DNA virus with a genome of approximately 3 800 nucleotides, which generates at least six proteins[2-4]. The functions or antigenicities of these potential proteins are still unknown. TTV was originally classified as belonging to the Circoviridae[2-4] family and could cause the increasing of alanine aminotransferase and TTV viremia. It is potentially related to many diseases. However, the infection mechanisms and pathogenicity of TTV are still poorly understood. Due to its global presence in healthy populations and the lack of morphological or molecular abnormalities of TTV-infected cells, the study of the biological functions of TTV is particularly challenging[5, 6].

In an attempt to shed light on TTV-host interactions, we proceeded to investigate the TTV proteins by the method of yeast two-hybrid system. In this study, we first analyzed the bioinformatics characterizations of ORF1 protein of TTV (SANBAN isolate, which belongs to a novel group 3 TTV genotype). The results of the bioinformatics analysis were then used to construct the plasmids including the whole TTV ORF1 plasmids to prepare for screening the functional positions of the protein.

2. Materials and methods

2.1. Cloning of TTV ORF1 gene

The whole TTV gene was synthesized and presented as a gift by Professor Gengfu Xiao. It was used as template

#Chen Jinling and Zhu Dandan contributed equally to this work.
Chen Jinling, Lecturer, Master of Medicine, E-mail: yinongduan@gmail.com.
*Corresponding author: Duan Yinong, 19 Qixiu Road, Department of Pathogen Biology, School of Medicine, Nantong University, Nantong City, Jiangsu Province, 226001, People's Republic of China.
Tel: +86+513+85051738
Fax: +86+513+85051876
E-mail: yinongduan@gmail.com

DNA. Then TTV ORF1 gene was amplified to PGEM-T vector (Promega) using the previous methods[7] and sequenced by Invitrogen (Shanghai). The sequence was blasted in GenBank.

2.2. Bioinformatics analysis of TTV ORF1 protein

To predict and analyze the signal peptide cleavage sites, we used the SignalP 3.0 server provided by EXPASY (http://www.cbs.dtu.dk/services/SignalP/). To predict and analyze the transmember domain of this protein, TMPRED software (http://www.ch.embnet.org/software/TMPRED_form.html) was used. An *N*-glycosylation site predictor for human proteins is available at http://www.cbs.dtu.dk/services/NetNGlyc/. ProtParam tool (http://us.expasy.org/tools/protparam.html/) is a tool which allows the computation of various physical and chemical properties for a user entered sequence. ProtScale tool (http://web.expasy.org/protscale/) computes the profile in the form of a two-dimensional plot. The NetPhos 2.0 server of center for biological sequence analysis(http://www.cbs.dtu.dk/services/NetPhos/) produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in proteins, while kinase specific phosphorylation predictions are available at: http://www.cbs.dtu.dk/services/NetPhosK/. PSORT (http://psort.nibb.ac.jp/) is a computer program for the prediction of protein localization sites in cells and PSORTⅡ is a new version of PSORT to predict the subcellular localization sites of proteins from their amino acid sequences. Besides these, we also used PredictProtein platform (http://www.predictprotein.org/) to predict the TTV ORF1 protein.

3. Results

3.1. Cloning of TTV ORF1 gene

PGEM-T-TTV-ORF1 plasmid was constructed and sequenced. It showed high homology (99.73%) to the complete ORF1 gene sequence of TTV DNA isolated as TTV SANBAN (GenBank accession number: AB025946). The sequence of ORF1 is 2 238 bp in length and the inconsistant base pair with the ORF1 gene sequence of TTV DNA of SANBAN (GenBank accession number: AB025946) were: 537 S-C, 864 Y-T, 876 Y-T, 982 R-A, 1000 M-C, 1092 W-A. The amino sequence of TTV ORF1 protein was obtained by Primer 5.0 software according to nucleotide sequence and it was as follows:

MAWGWWRRWRRWPARRWRRRRRRRPLRRRRAGRPARRYR
RRRTVRTRRRRWGRRRYRRGWRRRTYVRKGRHRKKKKRLIL
RQWQPATRRRCTITGYLPIVFCGHTKGNKNYALHSDDYTPQGQ
PFGGALSTTSFSLKVLFDQHQRGLNKWSFPNDQLDLARYRGCKF
IFYRTKQTDWIGQYDISEPYKLDKYSCPNYHPGNLIKAKHKFLIP
SYDTNPRGRQKIIVKIPPPDLFVDKWYTQEDLCSVNLVSLAVSAA
SFLHPFGSPQTDNPCYTFQVLKEFYYQAIGFSATDESRNYVFNVL
YEENSYWESNITPFYVINVKKGSNT**R**DYMSP**Q**ISDSHFRNKVNT
NYNWYTYNAKSHKNDLH**E**LRRAYFKQLTTEGPQQTSSEKGYAS
QWTTPTTDAYEYHLGMFSTIFLAPDRPVPRFPCAYQDVTYNPLM
DKGVGNHVWFQYNTKADTQLIVTGGSCKAHIEDIPLWAAFYGY
SDFIESELGPFVDAETVGLICVICPYTKPPMYNKTNPMMGYVFYD
RNFGDGKWIDGRGKIEPYWQVRWRPEMLFQETVMADIVQTGP
FSYKDELKNSTLVAKYKFYFTWGGNMMFQQTIKNPCKTDGRPT
DSDRHPRGIQVADPEQMGPRWVFHSFDWRRGYLSEGAIKRLHE
KPLDYESYFTQPKRPRIFPPTEAAEGEFREPEKGSYSEEERSQAS
AEEQATEETVLLLKRRLREQRKLEQQLQFLTREMFKTQAGLHIN
PMLLSQR.

The number of amino acids is 745 with R at 328aa, Q at 334aa and E at 364aa , while X were in these positions in amino sequence of TTV ORF1 protein of SANBAN (GenBank accession number: AB025946).

3.2. SignalP 3.0 server

The prediction of the signal peptide was performed by SignalP 3.0 server. S-score output from SignalP for the signal peptide prediction is reported for every single amino acid position in the submitted sequence, with high scores indicating that the corresponding amino acid is part of a signal peptide, and low scores indicating that the amino acid is part of a mature protein. S-mean score is the average of the S-score, ranging from the *N*-terminal amino acid to the amino acid assigned with the highest Y axis score, thus the S-mean score is calculated for the length of the predicted signal peptide. Using the neural networks (NN), the SignalP-NN results showed that the most likely cleavage site was between position 14 and 15 (WPA-RR) and signal peptide may be position 1-14 which S-mean value was 0.706.

3.3. TMPRED software

The analysis results of the given amino sequence provided by TMPRED software demonstrated that two possible transmembrane helices from inside to outside were found (from position 244 to 264, score: 1 453; from position 487 to 508, score: 490) (Figure 1) and three possible transmembrane helices from outside to inside were found (from position 121 to 140, score: 421; from position 239 to 260, score: 643; from position 487 to 508, score: 381)(Figure 1). Also two suggested models for transmembrane topology were performed. The first model showed (strongly preferred model: *N*-terminus inside) 1 strong transmembrane helix from position 244 to 264 (score: 1 453, orientation: inside >outside) was found, while the second model (alternative model) revealed the strong transmembrane helix from position 239 to 260 (score: 643, orientation: outside >inside) was found.

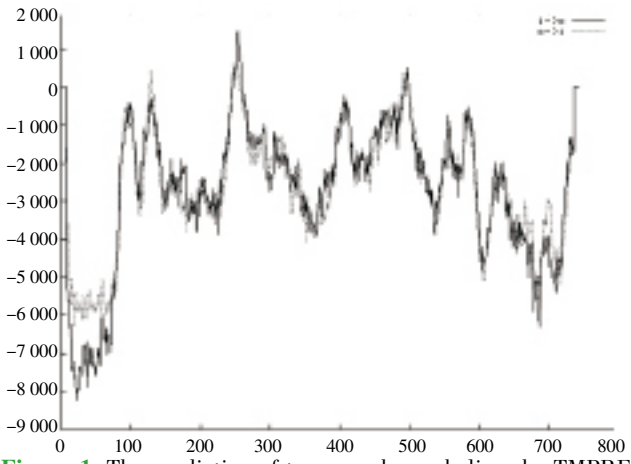

Figure 1. The prediction of transmembrane helices by TMPRED software.
X-axis represents protein length from *N*- to C-terminal. *Y*-axis represents the score computed by TMPRED.

3.4. NetNGlyc

The manual of NetNGlyc demonstrates that a position with a potential (vertical lines) crossing the threshold (horizontal line at 0.5) is predicted glycosylated. So the *N*-glyc results (Figure 2) illustrated that the position 509 (NKTN) was the potential *N*-glycosylation site predicted as *N*-glycosylated (++) with 0.677 6 score. The score of the other two positions (572, NSTL; 312, NITP) were all less than 0.5 (0.462 3 and 0.243 4, respectively).

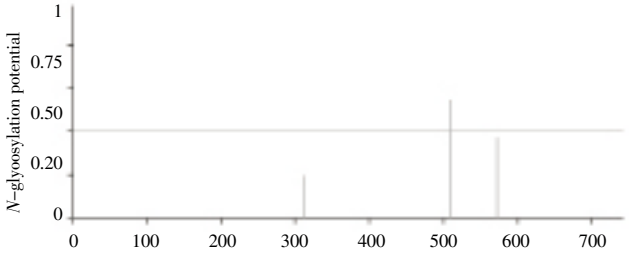

Figure 2. The prediction of *N*-glycosylation site by NetNGlyc.
X-axis represents protein length from *N*- to *C*-terminal. Horizontal line at 0.5 represents the threshold. The graph illustrates that the predicted *N*-glyc sites are the positions, whose values of *N*-glycosylation potential are crossing and above the threshold.

3.5. ProtParam tool

The parameters computed by ProtParam tool include the

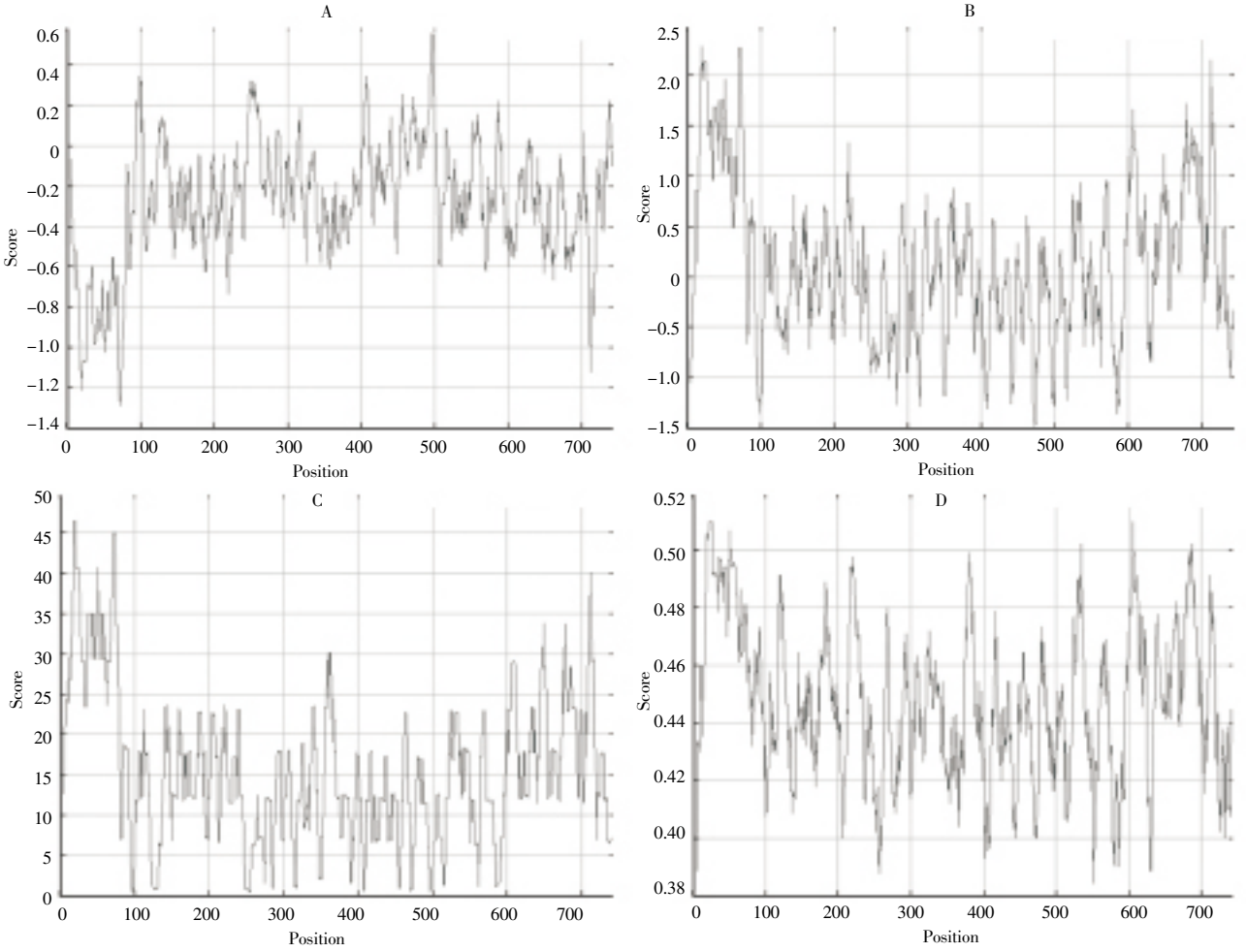

Figure 3. The hydrophilicity, polarity and flexibility parameters predicted by ProtScale platform.
X-axis represents protein length from *N*- to *C*-terminal. Y-axis represents the score computed by certain algorithm. The higher score, the higher probability.
A:hphob./Janin; B: hphob./Hopp&Woods; C: polarity/Zimmerman; D:Average flexibility.

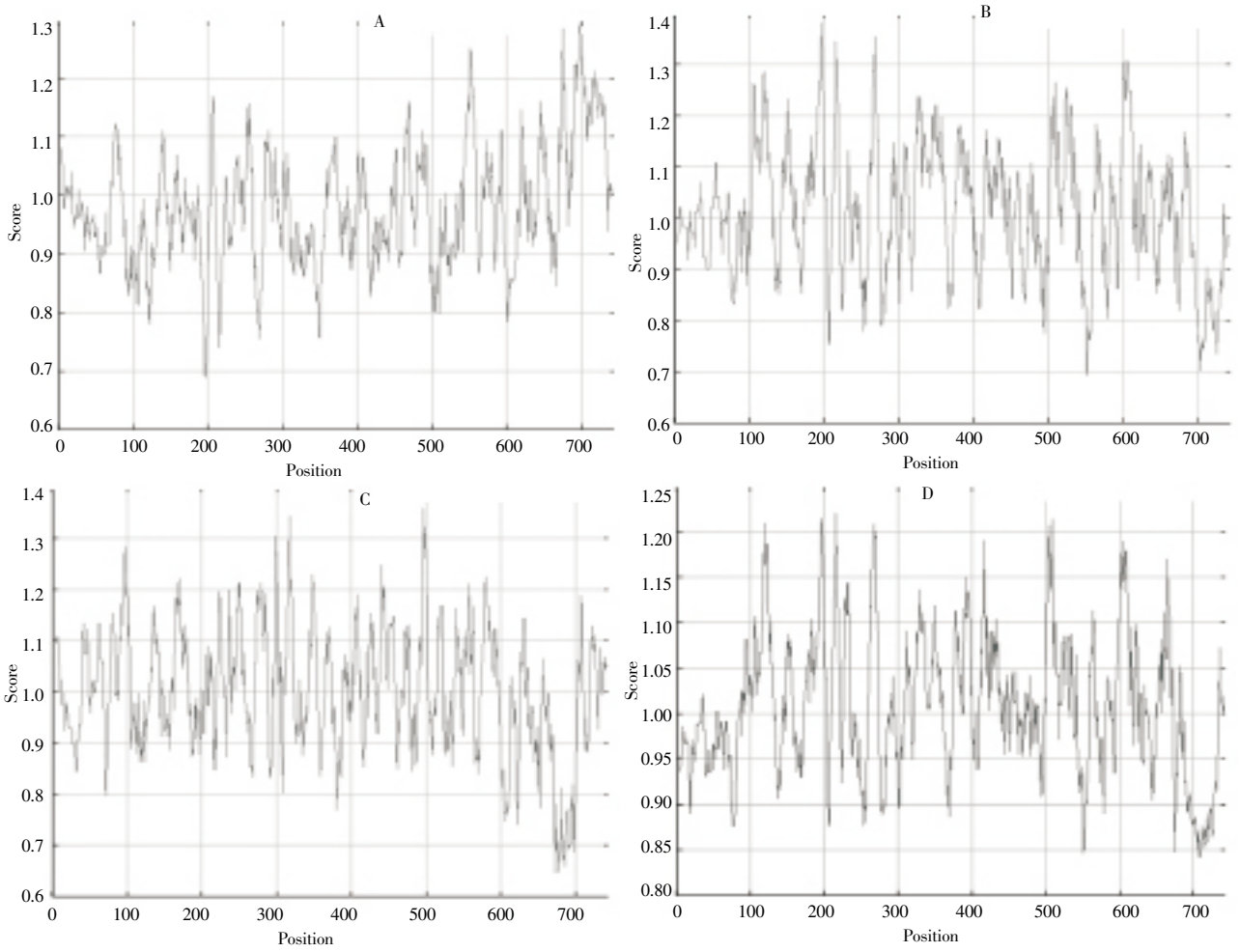

Figure 4. The secondary structures of the protein including alpha-helix, beta-turn, beta-sheet and coil predicted by ProtScale platform.
X-axis represents protein length from *N*- to *C*-terminal. Y-axis represents the score computed by certain algorithm. The higher score, the higher probability.
A: alpha-helix/Chou&Fasman; B: beta-turn/Chou&Fasman; C: beta-sheet/Chou&Fasman; D: coil/Deleage.

molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). The speculative molecular weight of TTV ORF1 protein was 88 705.7 Da. The theoretical pI was 9.89. Total number of negatively charged residues (Asp + Glu) was 74, while the total number of positively charged residues (Arg + Lys) was 126. Two values of extinction coefficients were produced by ProtParam, both for proteins measured in water at 280 nm. The first computed value [Ext. coefficient: 195 665, Abs 0.1% (=1 g/L): 2.206] was based on the assumption that all cysteine residues appeared as half cystines (*ie.* all pairs of Cys residues form cystines), and the second one [Ext. coefficient: 195 040, Abs 0.1% (=1 g/L): 2.199] was based on the assumption that no cysteine appeared as half cystine. The *N*-terminal of the sequence considered was M (Met). The estimated half-life was: 30 hours (mammalian reticulocytes, *in vitro*), >20 hours (yeast, *in vivo*) and >10 hours (*Escherichia coli*, *in vivo*). According to the formula, a protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. The instability index of ORF1 was computed to be 63.15 and this classified the protein as unstable. The aliphatic index was 57.07, while grand average of hydropathicity (GRAVY) was -0.0.895.

3.6. ProtScale tool

ProtScale allows to compute and represent (in the form of a two-dimensional plot) the profile produced by any amino acid scale on a selected protein. As a result, the minimum value of hphob./Janin was -1.289, while the maximum one was 0.556 (Figure 3A); the minimum value of hphob./Hopp & Woods was -1.478, while the maximum one was 2.289 (Figure 3B); the minimum value of polarity/Zimmerman was 0.429, while the maximum one was 46.456 (Figure 3C); the minimum value of average flexibility was 0.384, while the maximum one was 0.510 (Figure 3D); the minimum value of alpha-helix/Chou & Fasman was 0.690, while the maximum one was 1.294 (Figure 4A); the minimum value of beta-turn/Chou & Fasman was 0.692, while the maximum one was 1.381 (Figure 4B); the minimum value of beta-sheet/Chou & Fasman was 0.646, while the maximum one was 1.358 (Figure 4C); the minimum value of coil/Deleage & Roux was 0.842, while the maximum one was 1.219 (Figure 4D).

3.7. NetPhos 2.0 server and NetPhosK 1.0 server

The phosphorylation sites were predicted as follows: 1) Serine predictions: 16 sequences (including position 135, 289, 332, 336, 338, 357, 382, 383, 459, 565, 608, 640, 684, 686, 691, 694, maximum score: 0.997, minimum score: 0.670) were found; 2) Threonine predictions: 9 sequences (including position 43, 46, 64, 87, 92, 381, 447, 594, 606) were found with the maximum score was 0.988 and the minimum score was 0.525; 3) Tyrosine predictions: 16 sequences, including position 65, 118, 192, 197, 283, 297, 303, 308, 330, 387, 399, 424, 517, 539, 654, 685 (maximum score: 0.982, minimum score: 0.527) were found. Using the method of NetPhosK without ESS filtering, the kinase PKC was predicted at position 46 with the highest score 0.90.

3.8. PSORT: Results of subprograms

The possible cleavage site was between 14aa and 15aa and the transmembrane region was also 244aa-260aa. The results of membrane topology prediction indicated that *C*-terminal side of the protein would be inside and it was type 1b (cytoplasmic tail 244aa to 745aa). There was no transport motif from cell surface to Golgi; No *N*-myristoylation pattern was found. In the *C*-terminus, no ER retention motif and peroxisomal targeting signal were found. No possible vacuolar targeting motif was found. Also there was no RNA-binding motif, no actin-binding motif, no *N*-myristoylation pattern and no prenylation motif. However, too long tyrosines in the tail and two dileucine motifs (LL at 705 and LL at 706) in the tail were found. Then the κ-nearest neighbor (κ-NN) algorithm for assessing the probability of localizing at each candidate sites was performed. The result showed that these κ data points contained nuclear proteins with 60.9%, cytoplasmic proteins with 8.7% and mitochondrial proteins with 8.7%, the query was predicted to be localized to the nucleus with the probability of 60.9% and the prediction for query was nucleus.

3.9. PredictProtein

α-helix accounted for 16.4%, α strands for 15.7% and loop for 67.9% in the secondary structure of ORF1 protein of TTV. 1aa-91aa and 278aa-361aa were localized in non-regular secondary structure region. ORF1 protein was also predicted to be nuclear using PredictNLS (protein has a nuclear localization signal) at the position 10aa-85aa (RRWP ARRWRRRRRRRPLRRRRAGRPARRYRRRRTVRTRRRRWGRR RYRRGWRRRTYVRKGRHRKKKKRLILRQWQP).

4. Discussion

TTV is spread worldwide in nature and has been demonstrated in more than 90% of serum samples from healthy individuals[8-10]. TTV is a non-enveloped human virus with a circular ssDNA genome, which consists of a 2.6 kb coding and 1.2 kb non-coding region[11,12]. Previous studies have shown that TTV can produce three mRNA species by alternative splicing and then six proteins by alternative translation initiation[13,14]. Among the six proteins, ORF1 protein, which is the longest open reading frame, is considered to encode the nucleocapsid as a structural protein of TTV[15], while ORF2 is expected to represent a nonstructural protein that is necessary for viral replication[16]. There is a hydrophilic domain rich in arginine in the *N*-terminal amino acid sequence of ORF1. Such an arginine-rich and hydrophilic domain in the core protein is suggested to have DNA-binding activity and to function in packaging of the viral DNA[17], just as those in the *N*-terminus of the core protein of hepatitis C virus[18].

To make a study of TTV proteins and determine the interactions between TTV and host, we first analyzed the bioinformatics characterizations of ORF1 protein of TTV (SANBAN isolate). Since functional protein is a mature protein with no transmembrane domains, in the present study, we first used SignalP 3.0 server to predict the signal peptide. SignalP 3.0 server, which based on a combination of several artificial neural networks and hidden Markov models, predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms[19]. Then we used TMPRED to predict the transmembrane regions and orientation. The TMPRED program makes a prediction of membrane-spanning regions and their orientation based on the statistical analysis of TMbase, a database of naturally occurring transmembrane proteins[20]. To our knowledge, as an important post-translational modification, glycosylation has been seen as a very important mechanism determining the folding of proteins, localization and trafficking, protein solubility, antigenicity, biological activity and half-life, as well as cell-cell interactions. So we used NetNGlyc platform to predict the *N*-glycosylation sites in TTV ORF1 proteins. *N*-glycosylation is known to occur on asparagines, which occur in the Asn-Xaa-Ser/Thr stretch (where Xaa is any amino acid except Proline). While these consensus tripeptides (also called the *N*-glycosylation sequon) may be a requirement, it is not always sufficient for the asparagine to be glycosylated. NetNGlyc attempts to distinguish glycosylated sequons from non-glycosylated ones[21]. ProtParam, based on either compositional data, or on the *N*-terminal amino acid, is a tool which allows various physico-chemical properties deduced from a given protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and GRAVY. The extinction coefficient, which indicates how much light a protein absorbs at a certain wavelength, is one of the most important parameters computed by ProtParam. It is useful to have an estimation of this coefficient for a protein when purifying it[22]. The half-life is a prediction of the time, which takes for half of the amount of protein in a cell to disappear after its synthesis in the cell. ProtParam relies on the ´*N*-end rule´, which relates the half-life of a protein to the identity of its *N*-terminal residue. The instability index provides an estimate of the stability of the protein in a test tube. Guruprasad *et al*[23] analyzed 12 unstable and 32 stable proteins and considered that the occurrence of some

dipeptides was significantly different in the unstable proteins compared with those in the stable ones. The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins. ProtScale allows to predict and represent the profile produced by any amino acid scale (hydrophobicity scale, secondary structure conformational parameter scale and many other scales based on different chemical and physical properties of the amino acids) on a selected protein. The NetPhos 2.0 server is useful for serine, threonine and tyrosine phosphorylation sites prediction[24]. Residues having a prediction score above the threshold of 0.500 are indicated by 'S' (serine), 'T' (threonine) or 'Y' (tyrosine), respectively. Sequence context, one of the outputs from NetPhos is shown as a 9-residue sequence centered on the residue being analyzed (All the sequences were not shown in the results). The NetPhosK 1.0 server produces neural network predictions of kinase specific eukaryotic protein phosphoylation sites. Currently NetPhosK covers the following kinases: PKA, PKC, PKG, CK Ⅱ, Cdc2, CaM-Ⅱ, ATM, DNA PK, Cdk5, p38 MAPK, GSK3, CK Ⅰ, PKB, RSK, INSR, EGFR and Src[25]. PSORT, a computer program for the prediction of protein localization sites in cells, analyzes the input sequence by applying the stored rules for various sequence features of known protein sorting signals. PSORT Ⅱ is a long-awaited new version of PSORT, a program to predict the subcellular localization sites of proteins[26]. And the prediction of PSORT Ⅱ is performed using κ-nearest neighbor (κ-NN) algorithm for assessing the probability of localizing at each candidate sites. If these κ data points contain nuclear proteins with 50%, the query is predicted to be localized to the nucleus with the probability of 50%. PredictProtein is a service for sequence analysis, structure and function prediction[27-35] and it also showed that ORF1 protein was predicted to be nuclear.

Besides these, InterPro Scan (Gene ontology terms) was also performed to predict the functional information of TTV (http://www.ebi.ac.uk/Tools/InterProScan/). This platform is an extremely valuable and complex resource that integrates a wide variety of protein signature databases. Since protein function prediction can be done using protein signatures, InterPro Scan was produced to combine different protein signature recognition methods into one resource with look up of corresponding InterPro and gene ontology annotation[36]. However, when the amino sequence was inputted into the specified locations of InterPro Scan platform, the computer program was executed to recognize the sequence to be a TTV ORF1 protein which cantains a signal peptide at position 1-14, but with unknown function, no parent and no children (Data not shown). Then we also used ScanProsite platform and ProtFun 2.2 server (http://prosite.expasy.org/scanprosite/; http://www.cbs.dtu.dk/services/ProtFun/) to predict the TTV ORF1 protein. ScanProsite web tool, which displays for each found motif (profile/pattern/rule) hit and biological features within a protein sequence, showed no hits in the input sequence. The ProtFun 2.2 server, which produces ab initio predictions of protein function from sequence and indicates the cellular role, enzyme class and selected gene ontology categories of the submitted sequence[37], indicated the same results about phosphorylation sites, *N*-glycosylated site as the results from other web tools. And from the results of Protfun, we also obtained that no *O*-glycosylated sites were predicted in the sequence.

As we know, some disparities may exist in different web toots to predict the functional protein. And all the bioinformatics analysis results, which are just speculated by certain algorithms, couldn't absolutely represent the experimental results. For example, we obtained the prediction data from PSORT Ⅱ web server , which indicated that the ORF1 protein may be nuclear protein with 60.9% possibility, instead of cytoplasmic protein. Our results were inconsistent with the previous study[14], the result of which revealed that the ORF1 and ORF2 proteins were all predominately localized in the cytoplasm, perhaps consistent with a role as a structural component[14]. However, aggregate analysis could diminish the limitations of the single parameter prediction and improve the accuratissime of prediction. And in conclusion, all the results of the bioinformatics analysis were integrated to analyze in our study. According to the results of the bioinformatics analysis, some segments (*ie.* from position 265 to 486 and from position 510 to 679) and the whole TTV ORF1 gene would be chosen to be cloned into PGBKT7 DNA-BD cloning vector and these would be prepared to screen the functional positions of the ORF1 protein. It will provide a basis for the study on the infection mechanism of TTV. In short, bioinformatics analysis would be helpful to choose the functional and targeted segments of the proteins and make it easier to study the protein function.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgements

This work was supported by a grant from a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] Nishizawa T, Okamoto H, Konishi K, Yoshizawa H, Miyakawa Y, Mayumi M. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochem Biophys Res Commun* 1997; **241**: 92-97.

[2] Miyata H, Tsunoda H, Kazi A, Yamada A, Khan MA, Murakami J, et al. Identification of a novel GC-rich 113-nucleotide region to complete the circular, single-stranded DNA genome of TT virus, the first human circovirus. *J Virol* 1999; **73**: 3582-3586.

[3] Takahashi K, Iwasa Y, Hijikata M, Mishiro S. Identification of a new human DNA virus (TTV-like mini virus, TLMV) intermediately related to TT virus and chicken anemia virus. *Arch*

Virol 2000; **145**: 979-993.

[4] Kakkola L, Tommiska J, Boele LC, Miettinen S, Blom T, Kekarainen T, et al. Construction and biological activity of a full-length molecular clone of human Torque teno virus (TTV) genotype 6. *FEBS J* 2007; **274**(18): 4719-4730.

[5] Gergely PJr, Perl A, Poor G. Possible pathogenic nature of the recently discovered TT virus: does it play a role in autoimmune rheumatic diseases? *Autoimmun Rev* 2006; **6**: 5-9.

[6] Irshad M, Joshi YK, Sharma Y, Dhar I. Transfusion transmitted virus: a review on its molecular characteristics and role in medicine. *World J Gastroenterol* 2006; **12**: 5122-5134.

[7] Duan Y, Zhu D, Chen J, Qin Y, Wang J. Gene cloning, expression, and immunological characterization of kexin in rat Pneumocystis carinii. *Labmedicine* 2010; **41**: 99-103.

[8] Abe K, Inami T, Asano K, Miyoshi C, Masaki N, Hayashi S, et al. TT virus infection is widespread in the general populations from different geographic regions. *J Clin Microbiol* 1999; **37**: 2703-2705.

[9] Kakkola L, Hedman K, Vanrobaeys H, Hedman L, Sŏderlund-Venermo M. Cloning and sequencing of TT virus genotype 6 and expression of antigenic open reading frame 2 proteins. *J Gen Virol* 2002; **83**: 979-990.

[10]Kakkola L, Bonden H, Hedman L, Kivi N, Moisala S, Julin J, et al. Expression of all six human Torque teno virus (TTV) proteins in bacteria and in insect cells, and analysis of their IgG responses. *Virology* 2008; **382**: 182-189.

[11]Kamada K, Kamahora T, Kabat P, Hino S. Transcriptional regulation of TT virus: promoter and enhancer regions in the 1.2-kb noncoding region. *Virology* 2004; **321**: 341-348.

[12]Suzuki T, Suzuki R, Li J, Hijikata M, Matsuda M, Li TC, et al. Identification of basal promoter and enhancer elements in an untranslated region of the TT virus genome. *J Virol* 2004; **78**: 10820-10824.

[13]Kamahora T, Hino S, Miyata H. Three spliced mRNAs of TT virus transcribed from a plasmid containing the entire genome in COS1 cells. *J Virol* 2000; **74**: 9980-9986.

[14]Qiu J, Kakkola L, Cheng F, Ye C, Sŏderlund-Venermo M, Hedman K, et al. Human circovirus TT virus genotype 6 expresses six proteins following transfection of a full-length clone. *J Virol* 2005; **79**: 6505-6510.

[15]Zheng H, Ye L, Fang X, Li B, Wang Y, Xiang X, et al. Torque teno virus (SANBAN isolate) ORF2 protein suppresses NF-κB pathways via interaction with IkB kinases. *J Virol* 2007; **81**: 11917-11924.

[16]Okamoto H, Nishizawa T, Ukita M. A novel unenveloped DNA virus (TT virus) associated with acute and chronic non-A to G hepatitis. *Intervirology* 1999; **42**: 196-204.

[17]Erker JC, Leary TP, Desai SM, Chalmers ML, Mushahwar IK, et al. Analyses of TT virus full-length genomic sequences. *J Gen Virol* 1999; **80**: 1743-1750.

[18]Okamoto H, Okada S, Sugiyama Y, Yotsumoto S, Tanaka T, Yoshizawa H, et al. The 5′-terminal sequence of the hepatitis C virus genome. *Jpn J Exp Med* 1990; **60**: 167-177.

[19]Bendtsen JD, Nielsen H, Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**: 783-795.

[20]Hofmann K, Stoffel W. TMbase-A database of membrane spanning proteins segments. *Biol Chem Hoppe-Seyler* 1993; **374**: 166.

[21]Salim A, Bano A, Zaidi ZH. Prediction of possible sites for posttranslational modifications in human gamma crystallins: effect of glycation on the structure of human gamma-B-crystallin as analyzed by molecular modeling. *Proteins* 2003; **53**: 162-173.

[22]Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 1999; **112**: 531-552.

[23]Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng* 1990; **4**: 155-161.

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

[25]Blom N, Sicheritz-Ponten T, 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.

[26]Harms JS, Ren X, Oliveira SC, Splitter GA. Distinctions between bovine herpesvirus 1 and herpes simplex virus type 1 VP22 tegument protein subcellular associations. *J Virol* 2000; **74**: 3301-3312.

[27]Rost B, Yachdav G, Liu J. The PredictProtein server. *Nucleic Acids Res* 2004; **32** (Web Server issue): W321-W326.

[28]Okoror LE, Eniolorunda TA, Okoror OI. Molecular evolutionary studies of Lassa virus nucleoprotein 2. *Asian Pac J Trop Dis* 2011; **1**(1): 28-34.

[29]Owolabi RS, Daniel O, Araoye MO, Osagbemi GK, Odeigah L, Ogundiran A. Self-reported reasons for seeking HIV testing by people living with HIV/AIDS(PLWHA) in a tertiary hospital in Nigeria. *Asian Pac J Trop Dis* 2011; **1**(1): 59-62.

[30]Ravikumar H, Ramachandraswamy N, Puttaraju HP. Molecular strain typing of *Wolbachia* infection from Indian mosquitoes using *wsp* gene. *Asian Pac J Trop Dis* 2011; **1**(2): 106-109.

[31]Daniel O, Osman E, Adebiyi P, Mourad G, Declarcq E, Bakare R. Non tuberculosis mycobacteria isolates among new and previously treated pulmonary tuberculosis patients in Nigeria. *Asian Pac J Trop Dis* 2011; **1**(2): 113-115.

[32]Magana-Arachchi DN, Medagedara D, Thevanesam V. Molecular characterization of *Mycobacterium tuberculosis* isolates from Kandy, Sri Lanka. *Asian Pac J Trop Dis* 2011; **1**(3): 181-186.

[33]Gbotosho GO, Okuboyejo TM, Happi CT, Sowunmi A. Recrudescent *Plasmodium falciparum* infections in children in an endemic area following artemisinin-based combination treatments: Implications for disease control. *Asian Pac J Trop Dis* 2011; **1**(3): 195-202.

[34]Rahimi MT, Sharifdini M, Ahmadi A, Laktarashi B, Mahdavi SA, Kia EB. Hydatidosis in human and slaughtered herbivores in Mazandaran province, northern Iran. *Asian Pac J Trop Dis* 2011; **1**(3): 212-215.

[35]Antony B, Prasad BPMR. An outbreak of neonatal septicaemia by Enterobacter cloacae. *Asian Pac J Trop Dis* 2011; **1**(3): 227-229.

[36]Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. *Nucleic Acids Res* 2005; **33**(Web Server issue): W116-W120.

[37]Jensen LJ, Gupta R, Blom N, Devos D, Tamames J, Kesmir C, et al. Prediction of human protein function from post-translational modifications and localization features. *J Mol Biol* 2002; **319**: 1257-1265.