Preliminary Bioinformatic Analysis of Gallus gallus (Chicken) CD8a Gene

Zhongling Gong,Wenjie Xu

Gas group co., LTD of Shenzhen City · 518049 · china

Abstract: The experimental object of this study is Gallus gallus (chicken) CD8a gene. In this study, we performed sequence analysis of CD8a gene and analyzed the primary structure, secondary structure and tertiary structure of CD8a encoded protein by using bioinformatics. Through the analysis, we concluded that the sequence length is 648bp with various restriction sites identified, which is similar to CD8a of other species in Galliformes order. The sequence encodes a total of 216 amino acids. The amino acids have low hydrophobicity, with 12 phosphorylation sites and a transmembrane region, and no signal peptide. The protein secondary structure mainly consists ofαspiral, accounting for 22.69%; extension chain, accounting for 15.74%; random coil, accounting for 61.57 %, in non-continuous distribution. The relationship between CD8a and other species homologous sequences is basically consistent with the establishment of the system tree by the establishment of three trees.

Key words: Gallus gallus (chicken) Bioinformatics CD8a

1. Overview

Bioinformatics has emerged as a new frontier of science. With the advancement of computer technology, bioinformatics has been gradually applied in various researches including agriculture and forestry, biology and medicine [1]. Its increasingly important role in science and technology has been shown in variety of software tools, databases, servers for biological molecules data analysis and functional prediction.

In this study, we performed a preliminary analysis of Gallus gallus (chicken) CD8a gene by using bioinformatics methods. Gallus gallus (chicken) has high adaptability and is able to breed and to be kept in tropical and subtropical regions, as well as alpine region regardless of season. Their resistance to disease is strong with high survival rate reaching 98% during brooding period. They are protected species of secondary level in China.

CD8 molecule is type I transmembrane protein on T lymphocyte surface [2], which is also an important marker of T lymphocyte [3,4]. Chicken CD8 molecules are mainly expressed on the surfaces of cytotoxic T lymphocytes [5], inhibitory T cells and dendritic cells [6]. CD8 molecules are dimers which have two forms of expression: CD8aa homodimer and CD8aβ heterodimer. CD8 interacts with endogenous antigens presented by MHC I which plays an important role in chicken immunity system as well as other mammals [12,13]. CD8a is involved in thymic differentiation and signaling of T cell activation, [7] while CD8β is primarily involved in facilitating CD8a biologic functions.

Through analyzing CD8a nucleic acid sequence of different species, it has been found that CD8a is a more conserved transmembrane protein with similar function. It consists of 230 amino acids with a regulated signal peptide, an Ig hypervariable region, Ig hypervariable region exposure to the homologous immunoglobulin CDR1, 2, 3 of the hook ring region; followed by proline-rich hinge Ig hypervariable region and subsequently a transmembrane region constituted by twenty amino acids.

In this study, the nucleic acid sequence of CD8a was first analyzed. Subsequently identification of cleavage site allowed in vitro amplification of CD8a for molecular experiments. Through the analysis of amino acid sequences, the location of phosphorylation, presence or absence of signal peptide, and subcellular localization and its secondary structure, could be determined. The predicted protein structure and function were analyzed accordingly. The evolutionary relationship between different genes, and the relationship between CD8a and immune system can be better understood through bioinformatic analysis of CD8a gene and by comparing the same genes and alleles of different genes. Therefore, the study contributes to more promising prospect for poultry farming and breeding [8].

2. Materials and Methods

2.1 Research Materials

CD8a gene of Gallus gallus (chicken) with sequence length of 648bp was used in this study. The accession number of CD8a in GenBank is NM_205235.1.

2.2 Research Tools

In this study, the structure, function and evolution of CD8a (648bp) of Gallus gallus (chicken) were analyzed using bioinformatic methods. The following were the database, software and server used in the study:

Database:

NCBI (National Center for Biotechnology Information, National Biotechnology Information Center)

Software:

Bioedit (for nucleic acid sequence editing and analysis)

Discover studio visualizer (for protein structure analysis)

MEGA (for molecular evolutionary genetic analysis)

Server (mainly for protein structure prediction):

BLAST program (for comparison of homologous sequences)

Scratch protein predictor (protein structure prediction)

Swiss-model;

CBS server NetPhos2.0 Server

Exposé server for Prostate

CBS Server TMHMM Server v. 2.0

CBS server SignalP3.0 server

PBIL LYON-GERLAND database

SMART

CBS server CPHmodels

Wolf PSORT

CDART: Conserved Domain Architecture Retrieval Tool

2.3 Research Methods

First of all, NCBI was accessed and the nucleic acid sequence to be analyzed was saved in FASTA format and named as 48, while the

amino acid sequence was saved in FASTA format and named as CD8a.

2.3.1. Analysis of the primary structure of the nucleic acid sequence

(1) The basic analysis of the nucleic acid sequence using Bioedit software, open Bioedit, select the file, save the FASTA format nucleic acid sequence open. After selecting this sequence, click Sequence → Nucleic Acid → Nucleotide Composition to display the result of the molecular weight and base number of the nucleic acid sequence.

(2) Analysis of the nucleic acid sequence by restriction enzyme digestion

Using the Chromaspro software to analyze the restriction sites, open the FASTA format of the sequence in the Chromaspro software, then click Analysis → Restriction Enzyme to select all enzyme to display the results and save the results.

(3) search and comparison of homologous sequences

First open the NCBI home page, then open Blast, in the Basic Blast selected Nucleotide blast. After the page opens, browse the file in the dialog box, select the saved FASTA format nucleic acid sequence, select others in the Database (nr etc), the default other parameters. Finally point BLAST, display the analysis results then save.

2.3.2. A Preliminary Analysis of Amino Acid Sequences

(1) The composition of the amino acid sequence

Open the BioEdit software, open the amino acid sequence encoded by this amino acid, select the sequence, point → Protein → Amino Acid Composition, display the results, save.

(2) Analysis of phosphorylation sites of amino acid sequences

Use the NetPhos 2.0 server program on the CBS server of the Danish University of Science and Technology (DTU).

Website: http://www.cbs.dtu.dk/services/NetPhos/

Open the URL, browse the file, select the FASTA format amino acid sequence, then select the parameters: tyrosine (tyrosine), threonine (phenylalanine), serine (serine) these three amino acids in order to see these three amino acids It is phosphorylated, click Submit to submit, and the final result is output as an image.

(3) Amino acid sequence hydrophobicity and hydrophilicity analysis

A hydrophobicity analysis of amino acids was performed using the ProtScale program on the ExPAsy server

Website: http://www.expasy.org/cgi-bin/protscale.pl

Open the URL, then the amino acid sequence copy and paste to the dialog box, in the parameter selection, there are different Window size and amino acid scale, select different parameters will get different results, here we choose: Hphob. / Roseman, click Submit submission, results and save.

(4) Transmembrane region analysis of amino acid sequences

Use the TMHMM Server v. 2.0 program on the CBS server of the Danish University of Science and Technology (DTU).

Website: http://www.cbs.dtu.dk/services/TMHMM/

Open the URL, browse the file, the amino acid sequence FASTA format selected dialog box, the output form selected Extensive, with graphic, and finally click Submit, will appear later analysis results, save the results.

(5) Amino acid sequence of the signal peptide predicted

Use the SignalP3.0 server program.

Website: http://www.cbs.dtu.dk/services/SignalP/

First open the URL, browse the file, the amino acid sequence FASTA format selected dialog box, select the parameters:

Organism group: election Eukaryotes (eukaryotic cell);

Method (Neural Network Algorithm): In this experiment,

Graphics: Select GIF (inline);

Output format: Select Standard output mode;

Truncation: Because of the unpredictable signal peptide length, the general choice (0) is blocked, and this experiment also chooses (30) as a blocking constant for comparison. Click the Submit button to display the result and save it.

2.3.3. Analysis of high-energy structure of amino acids

(1) Analysis and prediction of secondary structure of amino acid sequence

The secondary structure of amino acid sequence was predicted by PBIL LYON-GERLAND database.

URL: http: //npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl page = / NPSA / npsa_hnn.html

First, open the URL and paste the amino acid sequence in the dialog box. Click Submit to display the results and save them.

(2) Structural domain analysis of amino acid sequences

The structure of the amino acid sequence was analyzed by simple module architecture search tool SMART.

Website: http://smart.embl-heidelberg.de/,

Open the URL, click on the home page through this page, and enter the new page. Paste the amino acid sequence in Sequence, and finally click the Sequence SMART to get the result and save the result.

(3) Subcellular localization of proteins

The subcellular localization of the amino acid sequence was performed using the WoLF PSORT program.

URL: http://wolfpsort.seq.cbrc.jp/

First, open the URL and select 'Animal' in 'select an organism type'. After the 'Please select input method' select 'From File'; browse the file, select the FASTA format amino acid sequence, and finally, the results will be saved.

(4) Amino acid sequence of tertiary structure analysis

The tertiary structure of the amino acid sequence was analyzed by CPHmodel 3.0 program on the CBS server of the Danish University of Science and Technology (DTU).

URL: http://www.cbs.dtu.dk/services/CPHmodels/

First open the URL, the amino acid sequence into the dialog box, point Submit, the new page in the Email address filled in a few minutes after the results will open, point query.pdb download pdb format files.

Open Accelrys Discovery Studio Visualizer3.1 software, open the file in OPEN just downloaded, you can appear in the 3D structure of the protein, save the results.

(5) The construction of the system tree and evolutionary status analysis

The nucleic acid sequence and amino acid sequence of MEGA4 software were used for homology analysis, and the phylogenetic tree was drawn. The evolutionary status analysis was carried out by NCBI database, BioEdit and MEGA software.

First, a sequence similar to that of the EST amino acid sequence was found by Protein Blast in NCBI and the amino acid sequences were downloaded together with the predicted sequences in a Notepad, opened in BioEdit, and stored in FASTA format.

Constructing phylogenetic tree with MEGA software. Open the MEGA → Alignment → Alignment explorer / CLUSTAL → select create a new alignment, click OK → in the pop-up dialog box point NO (that the sequence is placed in the protein) → there has been a dialog box, open the previously saved FASTA Format of the amino acid sequence → all selected, then point W (Align selected block by clustalW) → point OK → and then point Data → export alignment → MEGA format, save → in the previous small MEGA open on the steps to save the meg → phylogeny → construct Phylogeny → finally choose NJ or MP to get the phylogenetic tree, save the results.

Third, the forecast results and analysis

Landed NCBI, accession number NM_205235.1, the nucleic acid sequence and amino acid sequence saved as FASTA format, used for the following analysis.

(A) Nucleic acid sequence primary structure analysis

1. Nucleic acid sequence analysis

The molecular weight, base composition and base distribution of nucleic acid sequences were analyzed by Bioedit software. The results are as follows:

Figure 1 nucleic acid sequence 4 base percentage histogram

Through the analysis, the statistics are obtained

Table 1: Base composition analysis table

	A	T	C	G
Quantity	170	193	162	123
Percentage	26.23%	29.78%	25.00%	18.98%

The sequence has a total length of 648bp, single chain molecular weight of 198449.00Daltons, double chain molecular weight of 394483.00Daltons.

2. Analysis of nucleic acid sequence by restriction enzyme digestion

The results shown in Figure, contains a variety of enzyme restriction sites.

Figure 2 Nucleic acid sequence cleavage site prediction map

3. Preliminary analysis of base homology

In the NCBI, the sequence BLAST, the following results:

Figure 3 nucleic acid sequence base homology comparison chart (1)

Figure 4 Comparison of nucleic acid sequence base homology (2)

In this paper, the sequence predicted and the chicken CD8a gene similarity of up to 100%, we track one of the sequences, the following results:

Figure 5 nucleic acid sequence base homology comparison chart (3)

This is the chicken chicken CD8a gene and the nucleic acid sequence of the control, in the original chicken of the 62th base to 709 between the two of the same sequence.

(B) A preliminary analysis of amino acid sequences

1. Amino acid sequence

The results of amino acid composition analysis using Bioedit software are as follows:

Figure 6 Histogram of amino acids

Table 2 Analysis of amino acid composition

Amino Acid	Number	Mol%	Amino Acid	Number	Mol%
Ala A	15	6.94	Met M	4	1.85
Cys C	9	4.17	Asn N	13	6.02
Asp D	5	2.31	Pro p	15	6.94
Glu E	9	4.17	Gln Q	14	6.48
Phe F	12	5.56	Arg R	16	7.41
Gly G	12	5.56	Ser S	17	7.87
His H	4	1.85	Thr T	16	7.41
Ile I	12	5.56	Val V	11	5.09
Lys K	11	5.09	Trp W	2	0.93
Leu L	15	6.94	Tyr y	4	1.85

2. Prediction of Phosphate Sites

The results showed that there were nine serine (Ser), two threonine (Thr) and one tyrosine (Tyr) could be phosphorylated. The probabilities were as follows:

Possible phosphorylation sites in sequence of Figure 7 (1)

Possible phosphorylation sites in sequence of Figure 8 (2)

3. Analysis of hydrophobicity and hydrophilicity of amino acids

Get the following results

Figure 9 Amino acid affinity / hydrophobicity profile

The hydrophobicity of the amino acid sequence determines the structure of the protein. From the output, the amino acid sequence is concentrated in the vicinity of -1, so the hydrophobicity is not strong and belongs to the hydrophilic amino acid.

4. Distribution of amino acid transmembrane regions

The transmembrane region of the protein sequence was analyzed using the TMHMM Server v.2.0 program on the CBS server of the Danish University of Science and Technology (DTU).

The results are shown in Figure 9:

Figure 9, the amino acid sequence transmembrane analysis

From the analysis results, there is a transmembrane region in this amino acid sequence.

5. Signal peptide region prediction

Figure 10 Analysis of amino acid sequence signal peptide

By analysis, this sequence does not have a signal peptide.

(3) Analysis of high energy structure of amino acids

1. Secondary Structure Analysis of Amino Acids

Figure 11 Secondary structure analysis of amino acids

Table 3 Secondary structure prediction results

A helix (Hh)	49	22.69%
310 Spiral (Gg)	0	0.00%
Pi Spiral (Ii)	0	0.00%
βsheet (Bb))	0	0.00%
Extension chain (Ee)	34	15.74%
βRotation angle (Tt)	0	0.00%
Curved field (Ss)	0	0.00%
Random Curl (Cc)	133	61.57%
Any form	0	0.00%
Other Form	0	0.00%

Figure 12 Amino acid secondary structure prediction

Figure 13 Secondary structure analysis Visual results

Through the table and the figure we can see that the secondary structure of this amino acid sequence is mainly a spiral, accounting for 22.69%; extension chain, accounting for 15.74%; random curl, accounting for 61.57%, and the distribution is not continuous.

2. Structural Functional Domain Analysis of Protein Sequences

With the output, we conclude that the functional domain of this sequence is the IG domain encoded by the 23rd amino acid to the 133th amino acid. E value is 1.63e-03.

3. Protein subcellular localization

The subcellular localization of the amino acid sequence was performed using the WoLF PSORT program. The results are as follows:

Figure 15 subcellular localization analysis results (1)

Figure 16 subcellular localization analysis results (2)

Figure 17 subcellular localization analysis results (3)

By analyzing the results, we conclude that the subcellular localization analysis protein is located on the cell membrane.

4. Analysis of tertiary structure of amino acid sequence

Results and analysis

Figure 18 Three-level structure model

The above is a model of the tertiary structure of the protein, with α helix and a lot of random curl, the rest for the extension chain.

5. Construction of phylogenetic tree and its evolutionary analysis

Figure 19 Absolute conservative area (partial)

Figure 19 N-J method to build the tingling tree

Figure 20 M-P method to build the phylogenetic tree

The results of the above analysis show that the relationship between the sequences studied in this paper and the homologous sequences of other species is basically the same. Predictive proteins are closest to turkeys, so they are predicted to be in the same evolutionary position in terms of structure and function.

Discussion

(A) Basic analysis of nucleic acid sequences

Through the analysis of the above aspects of the sequence, we can conclude that the sequence is a full length of 648bp mDNA, containing a variety of enzyme cleavage sites, through the information obtained by the digestion site, the nucleic acid sequence of this article digestion and in-vitro amplification, for molecular experiments. Through the homology analysis of the base, compared with other chicken head CD8a sequence is relatively similar, relatively conservative.

(B) the basic analysis of amino acid sequence

Through the analysis of amino acid sequence, we found that the sequence of the encoded protein has a strong conservative. There are nine serine (Ser), two threonine (Thr), one tyrosine (Tyr) may be phosphorylated, and amino acid sequence of hydrophobic water determines the structure of the protein, from the output results, amino acid sequence value of less than 0, hydrophobic is not strong, are hydrophilic amino acids. This section of the amino acid sequence has a transmembrane region, there is no signal peptide region, suspected sequence structure and functional domain is not complete, need to be further analysis.

(C) Advanced structure analysis and functional prediction of protein

The protein secondary structure is mainly α-helix, accounting for 22.69%; extension chain, accounting for 15.74%; random curl, accounting for 61.57%, and the distribution is not continuous. The functional domain of this sequence is the IG domain encoded by the 23rd amino acid to the 133th amino acid. E value is 1.63e-03.After subcellular localization analysis, the protein is located in mitochondria. The phylogenetic tree constructed by the M-P method predicts that the protein is closest to the turkey, so it is predicted that they are in the same evolutionary position in terms of structure and function.

Conclusion

CD8 molecule is the type I transmembrane protein on the surface of T lymphocytes [11, 14], which is also an important landmark molecule on the surface of T lymphocytes [9]. CD8 molecules are dimers, and CD8aa homodimer and CD8aβ heterodimer two forms of expression [15]. Among them, CD8a is a similar and more conservative transmembrane protein, composed of about 230 amino acids, contains a standard signal peptide, an Ig hypervariable region, Ig hypervariable region exposed to homologous immunoglobulin CDR1, 2,3 of the hook ring area; followed by Ig hypervariable region is rich in proline-hinge region, followed by twenty amino acids constitute the transmembrane region. This paper predicts that the protein contains 216 amino acids. From the results, there is no signal peptide, cannot carry out membrane function, but there is a transmembrane region, as to how it is completed through the membrane, and we need further analysis. Through the above analysis, it is speculated that the protein may be combined with other aorta-encoded a chain and β chain to complete the immune function. It is now widely believed that CD8 molecules are involved in TCR signaling cofactors. As a secondary receptor, CD8 plays an important role in signal transduction and activates TCR through the tyrosine phosphorylation pathway between p56lck and CD3 chains [10]. Recent data suggest that CD8 has a series of important immunological functions of transmembrane molecules.

REFERENCES

[1]. Bernard, V. and M. Michaut, Explain bioinformatics to your grandmother! PLoS Comput Biol. 9 (10): e1003305.

[2]. Wu, T.L., et al., CD8 T Cell Recognition of Epitopes Within the Capsid of Adeno-associated Virus 8-based Gene Transfer Vectors Depends on Vectors 'Genome. Mol Ther.

[3]. Zou, Q., et al., CD8 Treg cells suppress CD8 T cell-responses by IL-10-dependent mechanism during H5N1 influenza virus infection. Eur J Immunol.

[4]. Snauwaert, S., et al., In vitro generation of mature, naive antigen-specific CD8 T cells with a single T-cell receptor by agonist selection.

[5]. Trifari, S., et al., MicroRNA-directed program of cytotoxic CD8 + T-cell differentiation. Proc Natl Acad Sci U S A. 110 (46): 18608-13.

[6]. Brown, A.S., et al., CD8 subunit expression by plasmacytoid dendritic cells is variable, and does not define stable subsets. Mucosal Immunol.

[7]. Cho, J.H., et al., Unique Features of Naive CD8 + T Cell Activation by IL-2. J Immunol.

[8]. Meng, X.T., et al., Increased hepatic yolk precursor synthesis, secretion and facilitated uptake by follicles are involved in the rejuvenation of reproductive performance of molted hens (Gallus gallus domesticus). Gen Comp Endocrinol. 194C: 198-207.

[9]. Zelinskyy, G., T. Werner, and U. Dittmer, Natural regulatory T cells inhibiting production of cytotoxic molecules in CD8 + T cells during low-level Friend retrovirus infection. Retrovirology. 10 (1): 109.

[10]. Gonzalez-Serna, A., et al., CD8 TCR beta chain repertoire expansions and deletions are related with immunato markers in HIV-1-infected patients during treatment interruption. J Clin Virol.

[11]. HUANG J.; Cloning of chicken CD4 and CD8α gene and construction of its eukaryotic expression plasmid. Chinese Society of Animal Husbandry and Veterinary Medicine 2004 Annual Symposium and the 5th National Symposium on Animal Science and Veterinary Young Scientists and Technologists Veterinary Society 2004 Annual Conference and the Fifth National Animal Husbandry and Veterinary Young Scientists and Technicians Academic Seminar (Volume 1): Yangzhou University School of Biological Science and Technology;

[12]. Xu Q.;. Preliminary analysis of cloning and expression regulation of duck CD8α gene. See: Chen Guohong, ed., Special economic animal breeding.

[13]. Jiao X. Advances in studies on chicken CD4 and CD8 molecules. Advances in animal medicine. 2005. (04).

[14]. Xie G. Cloning and prokaryotic expression of chicken CD4, CD8α gene. See: Li Heian; editor-in-chief.

[15]. Tang C. Fluorescence quantitative RT-PCR detection of chicken CD4, CD8 gene expression level. Journal of Animal Husbandry and Veterinary Medicine. 2008 (06).