Computational Biology and Chemistry 47 (2013) 207–214

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Research Article

Bioinformatics analysis of a non-specific nuclease from *Yersinia enterocolitica subsp. palearctica*

Zhen-Hua Li[a], Zhen-Xing Tang[b,c], Xiu-Juan Fang[a], Zhi-Liang Zhang[a], Lu-E. Shi[a,*]

[a] College of Life and Environmental Sciences, Hangzhou Normal University, 310016 Hangzhou, Zhejiang, China
[b] Date Palm Research Center, King Faisal University, P.O. Box 420, Al-hasa 31982, Saudi Arabia
[c] Department of Food Science, Anqing Vocational & Technical College, 246003 Anqing, Anhui, China

ARTICLE INFO

Article history:
Received 30 July 2013
Received in revised form
18 September 2013
Accepted 18 September 2013

Keywords:
Yersinia enterocolitica subsp. palearctica
Non-specific nuclease
Bioinformatics analysis
Thermostability

ABSTRACT

In this paper, the physical and chemical characteristics, biological structure and function of a non-specific nuclease from *Yersinia enterocolitica subsp. palearctica* (*Y. NSN*) found in our group were studied using multiple bioinformatics approaches. The results showed that *Y. NSN* had 283 amino acids, a weight of 30,692.5 ku and a certain hydrophilic property. *Y. NSN* had a signal peptide, no transmembrane domains and disulphide bonds. Cleavage site in *Y. NSN* was between pos. 23 and 24. The prediction result of the secondary structure showed *Y. NSN* was a coil structure-based protein. The ratio of α-helix, β-folded and random coil were 18.73%, 16.96% and 64.31%, respectively. Active sites were pos. 124, 125, 127, 157, 165 and 169. Mg^{2+} binding site was pos. 157. Substrate binding sites were pos. 124, 125 and 169. The analysis of multisequencing alignment and phylogenetic tree indicated that *Y. NSN* shared high similarity with the nuclease from *Y. enterocolitica subsp. enterocolitica 8081*. The enzyme activity results showed that *Y. NSN* was a nuclease with good thermostability.

1. Introduction

Nuclease is a class of enzymes that can degrade DNA and/or RNA molecules by cleaving the phosphodiester bonds that link adjacent nucleotides. According to consensus criteria, nucleases can be classified into RNase, DNase, and non-specific nuclease. Non-specific nucleases can attack both DNA and RNA in the presence of Mg^{2+} or other bivalent cations (Li et al., 2005). Various non-specific nucleases, such as *Serratia marcescens* nuclease that is one of the best-studied non-specific nuclease (Friedhoff et al., 1994, 1996, 1999; Miller and Krause, 1996; Miller et al., 1994; Shlyapnikov et al., 2000, 2002), nuclease from white spot syndrome virus (Li et al., 2005), and mitochondrial EndoG that is a sugar non-specific nuclease (Schäfer et al., 2004), have been reported.

Protein structure, function and biological properties of *Y. NSN* are very important for its industrial applications. In biotechnology field, bioinformatics can be used as a tool to process biological data and provide answers to some biological questions. It has been playing an important role in analyzing sequences of nucleic acids and proteins, and predicting the characteristics of protein structure and function (Ouyang et al., 2009; Bai et al., 2012; Yang and Shang, 2007). Bai et al. (2012) studied the physical and chemical characteristics, signal peptide, transmembrane domain and topological structure of a novel protein TgIMP1 through bioinformatics methods. Hu et al. (2013) found that the second thermonuclease from *Staphylococcus aureus* (Nuc2) was a more conserved protein when compared with the first thermonuclease from *S. aureus* (Nuc1) by sequence alignment and phylogenetic methods.

Yersinia enterocolitica is a common food-borne pathogen which can cause yersiniosis in humans (Hanifian and Khani, 2012). It has been paid much attention due to the cause of intestinal infectious disease outbreak. A non-specific nuclease from *Y. enterocolitica subsp. palearctica* (*Y. NSN*) has been found in our group. Therefore, in this study, bioinformatics method was used to analyze and predict protein structure and function, including hydrophilicity/hydrophobicity, signal peptide, transmembrane structure, secondary structure, structure domain and the three dimensional structure. Meanwhile, the evolution relationship of the nucleases was studied by the analysis of multisequencing alignment and phylogenetic tree. Thermostability of the available nucleases was also investigated. It is the first report on the study of the characteristics of structure and function from *Y. NSN* using bioinformatics methods.

2. Materials and methods

2.1. Source of sequences

The gene sequence and amino acids sequence of *Y. NSN* were obtained from the GenBank (http://www.ncbi.nlm.nih.gov). The

* Corresponding author. Tel.: +86 15268553605.
E-mail address: shilue@126.com (L.-E. Shi).

1476-9271/$ – see front matter.
http://dx.doi.org/10.1016/j.compbiolchem.2013.09.003

Table 1
The bioinformatics web sites and software used in this study.

Name	Explanation	Website
ProtParam	Prediction of physical and chemical characteristics	http://web.expasy.org/protparam/
ProtScale	Prediction of hydrophilicity/hydrophobicity	http://web.expasy.org/protscale/
Signal P	Prediction of signal peptide	http://www.cbs.dtu.dk/services/SignalP/
TMpred	Prediction of transmembrane structure	http://www.ch.embnet.org/software/TMPRED_form.html
PredictProtein	Prediction of secondary structure	https://www.predictprotein.org/
CDD	Prediction of structure domain	http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
SWISS-MODEL	Prediction of three dimensional structure	http://swissmodel.expasy.org//SWISS-MODEL.html/
MEGA 5.05	Phylogenetic tree analysis	/
DNAMAN	Sequencing alignment	/

Table 2
Nucleases selected for multiple sequence alignment and phylogenetic tree analysis.

Number	Species or subspecies	Accession no.
1	*Yersinia enterocolitica subsp. palearctica*	YP_006005337.1
2	*Arsenophonus nasoniae*	CBA74571.1
3	*Brenneria sp. EniD312*	ZP_09015352.1
4	*Enterobacter aerogenes EA1509E*	YP_007389540.1
5	*Klebsiella oxytoca M5al*	ZP_16164117.1
6	*Pectobacterium wasabiae CFBP 3304*	ZP_15539849.1
7	*Pseudomonas brassicacearum subsp. brassicacearum NFM421*	YP_004353219.1
8	*Pseudomonas fluorescens A506*	YP_006323691.1
9	*Pseudomonas sp. PAMC 25886*	ZP_10429599.1
10	*Salmonella bongori NCTC 12419*	YP_004730055.1
11	*Salmonella enterica subsp. enterica serovar Typhimurium str. D23580*	YP_005232392.1
12	*Serratia marcescens WW4*	YP_007405584.1
13	*Serratia odorifera DSM 4582*	ZP_06640950.1
14	*Serratia plymuthica A30*	ZP_16219248.1
15	*Serratia proteamaculans 568*	YP_001478041.1
16	*Serratia sp. AS12*	YP_004500194.1
17	*Yersinia enterocolitica subsp. enterocolitica 8081*	YP_001007112.1
18	*Yersinia frederiksenii ATCC 33641*	ZP_04632529.1
19	*Yersinia pseudotuberculosis IP 31758*	YP_001393388.1
20	*Yersinia ruckeri ATCC 29473*	ZP_04616648.1

accession numbers are FR729477.2 and YP_006005337.1, respectively.

2.2. Analytical methods

The bioinformatics web sites and software used in this study are showed in Table 1. The nucleases for the analysis of multiple sequence alignment and phylogenetic tree are shown in Table 2.

2.3. Enzyme activity analysis

Enzyme activity was determined by the degradation degree of DNA. Approximately 2.0 μl of diluted enzyme solution was incubated in 36 μl DNA of calf thymus (100 ng/μl) as a substrate. After incubation at 37 °C for 5 min, the reaction was stopped by adding 8.0 μl of 6× DNA loading buffer (10 mM Tris–HCl, 60 mM EDTA, 40% sucrose, 0.05% bromophenol blue, pH 7.6). The reaction products were then subjected to 1.0% agarose gel electrophoresis. The degradation degree of nucleic acid could be visualized and calculated by staining with 0.25 μg/ml ethidium bromide.

3. Results

3.1. Physical and chemical characteristics of Y. NSN

Y. NSN had 283 amino acids residues. The theoretical pI was 5.94. The formula and molecular weight of *Y. NSN* was $C_{1357}H_{2142}N_{372}O_{425}S_7$ and 30,692.5 ku, respectively. As regards amino acid composition of *Y. NSN*, the most abundant amino acids were Ala, Leu, Thr and Asn, with a proportion of 11.3%, 9.5%, 9.2%, 7.1%, respectively. The calculated instability index (II) was 28.56.

3.2. Prediction of hydrophilicity/hydrophobicity

Protscale was used to analyze the hydrophilicity/hydrophobicity of *Y. NSN*. Ampholyte amino acids were between −0.5 and +0.5. The strongest hydrophilic amino acid and hydrophobic amino acid was Lys (−2.533) and Leu (2.522), respectively (Fig. 1).

3.3. Prediction of signal peptide and transmembrane structure

In this study, *C*-max 0.215, *Y*-max 0.437, *S*-max 0.960, *S*-mean 0.887 and *D* 0.680 were obtained by the tool of SignalP. Cleavage site was between pos. 23 and 24 (Fig. 2a). According to the results presented in Fig. 3, *Y. NSN* only had one transmembrane domain (the score of the region above 600). And transmembrane helix was N-terminus inside, a 24 bp sequence from 3 to 26.

3.4. Prediction of secondary structure and structure domain

The results of PredictProtein showed that *Y. NSN* was a coil structure-based protein. The percent of α-helix, β-folded and random coil in the secondary structure of *Y. NSN* was 18.73%, 16.96%, 64.31%, respectively (Fig. 4). 59.36% of the residues were exposed to more than 16% of their surface according to the predicted solvent accessibility composition. Meanwhile, it was observed that disulphide bonds did not exist in the protein sequence. PredictProtein also was used to predict structure domain of *Y. NSN*. The results showed there were four potential motifs named N-glycosylation site (70), protein kinase C phosphorylation site (231, 271), casein kinase II phosphorylation site (18, 34, 215) and N-myristoylation site (51, 136, 220). In addition, the function domains of *Y. NSN* were also analyzed by CDD. The results indicated *Y. NSN* had two function domains. One had a high homology with NUC superfamily (48. . .274), the other was Endonuclease_NS (63. . .263). Furthermore, the information of *Y. NSN*, such as active sites (pos. 124, 125, 127, 157, 165 and 169), Mg^{2+} binding site required for enzyme activity (pos. 157) and substrate binding sites with hydrolyzing DNA or RNA (pos. 124, 125 and 169), were also obtained through CDD (Fig. 5).

3.5. Prediction of three-dimensional structure

In order to study three-dimensional structure of *Y. NSN*, SIWSS-MODEL was used in this study. Sequence identity was 60.17%. QMEAN *Z*-score was −1.67 (Fig. 6a). Besides, Phyre was also used to study three-dimensional structure of *Y. NSN*. The confidence was 100%, and the coverage was 84% (Fig. 6b).