

Author's Accepted Manuscript

Identification of thermophilic proteins by incorporating evolutionary and acid dissociation information into Chou's general pseudo amino acid composition

Guo-Liang Fan, Yan-Ling Liu, Hui Wang



www.elsevier.com/locate/jtbi

PII: S0022-5193(16)30193-X
DOI: <http://dx.doi.org/10.1016/j.jtbi.2016.07.010>
Reference: YJTBI8735

To appear in: *Journal of Theoretical Biology*

Received date: 16 February 2016
Revised date: 24 June 2016
Accepted date: 7 July 2016

Cite this article as: Guo-Liang Fan, Yan-Ling Liu and Hui Wang, Identification of thermophilic proteins by incorporating evolutionary and acid dissociation information into Chou's general pseudo amino acid composition, *Journal of Theoretical Biology*, <http://dx.doi.org/10.1016/j.jtbi.2016.07.010>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Revised manuscript: No. JTB-D-16-00162

Identification of thermophilic proteins by incorporating evolutionary and acid dissociation information into Chou's general pseudo amino acid composition

Guo-Liang Fan*, Yan-Ling Liu, Hui Wang

Department of Physics, School of Physical Science and Technology, Inner Mongolia University, Hohhot 010021, China

Abstract

Thermophilic proteins can thrive stably at the high temperatures. Identification of thermophilic protein could be helpful to learn the function of protein. Automated prediction of thermophilic protein is an important tool for genome annotation. In this work, a powerful predictor is proposed by combining amino acid composition, evolutionary information, and acid dissociation constant. The overall prediction accuracy of 93.53% was obtained for using the algorithm of support vector machine. In order to check the performance of our method, two low-similarity independent testing datasets are used to test the proposed method. Comparisons with other methods show that the prediction results were better than other existing methods in literature. This indicates that our approach was effective to predict thermophilic proteins.

Keywords Jackknife validation; Auto covariance; Thermophile; Mesophilic proteins

1 Introduction

A thermophile is a type of extremophile that thrives at the temperatures between 41 and 122 °C (Takai et al., 2008). Many thermophiles are archaea and suggested to have been among the earliest bacteria (Horiike et al., 2009). They are the most unique organism on the planet in terms of their contribution to modern life.

As a prerequisite for their survival, thermophiles contain enzymes that can function at high temperatures. Some of these enzymes are used in molecular biology, like heat-stable DNA polymerases for PCR and in washing agents.

How cells adapt to heat is a hot topic. Changing in ambient temperature affect nearly all cellular and biochemical processes and drive adaptive responses to maintain cellular homeostasis. Proteins from thermophilic organisms usually show high intrinsic thermal stability but have structures that are very similar to their mesophilic homologues. Several studies (Hasegawa et al., 2000) have been pointed out that the increase in Gibbs free energy change of hydration and shape enhanced the stability of thermophilic proteins, which has been validated by experiment (Hasegawa et al., 2000). Recently, Ma et al. (Ma et al., 2015) identified the gene ACDH-11 (acyl-CoA dehydrogenase) to find out the responds of *C. elegans* to conditions of changing oxygen. They discovered that ACDH-11 functions in homeoviscous adaptation and does so by regulating levels of the stearic Co-A desaturase (SCD) FAT-7. This demonstrated that specific intracellular fatty acids link ACDH-11 in a metabolic pathway to NHRs for transcriptional control of homeoviscous heat adaptation in *C. elegans*. The ACDH-11 pathway drives heat adaptation by linking temperature shifts to regulation of lipid desaturase levels and membrane fluidity via an unprecedented mode of fatty acid signaling. The other researchers reported that the specific amino acid composition corresponding to the stability of thermophilic proteins. Thus, Identification of thermophilic from mesophilic proteins by computational algorithms provided a new thought for the theoretical description of protein folding and stability.

Recently, many researchers have endeavored to develop the algorithm to distinguish the thermophilic from mesophilic proteins (Liu et al., 2011; Wang et al., 2011). Zhang and Fang (Zhang and Fang, 2007) constructed the dataset and discriminated the mesophilic and thermophilic proteins by using the dipeptide composition and amino acid composition and achieved 86.6% accuracy for 5-fold cross validation. Then, Gromiha and Suresh

Download English Version:

<https://daneshyari.com/en/article/6368946>

Download Persian Version:

<https://daneshyari.com/article/6368946>

[Daneshyari.com](https://daneshyari.com)