Journal of Molecular Graphics and Modelling 73 (2017) 166–178

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Identifying the molecular functions of electron transport proteins using radial basis function networks and biochemical properties

Nguyen-Quoc-Khanh Le*, Trinh-Trung-Duong Nguyen, Yu-Yen Ou*

Department of Computer Science and Engineering, Yuan Ze University, Chung-Li, Taiwan

ARTICLE INFO

Article history:
Received 4 November 2016
Received in revised form
26 December 2016
Accepted 4 January 2017
Available online 2 February 2017

Keywords:
Electron transport proteins
Transporter
Annotation
Feature selection

ABSTRACT

The electron transport proteins have an important role in storing and transferring electrons in cellular respiration, which is the most proficient process through which cells gather energy from consumed food. According to the molecular functions, the electron transport chain components could be formed with five complexes with several different electron carriers and functions. Therefore, identifying the molecular functions in the electron transport chain is vital for helping biologists understand the electron transport chain process and energy production in cells. This work includes two phases for discriminating electron transport proteins from transport proteins and classifying categories of five complexes in electron transport proteins. In the first phase, the performances from PSSM with AAIndex feature set were successful in identifying electron transport proteins in transport proteins with achieved sensitivity of 73.2%, specificity of 94.1%, and accuracy of 91.3%, with MCC of 0.64 for independent data set. With the second phase, our method can approach a precise model for identifying of five complexes with different molecular functions in electron transport proteins. The PSSM with AAIndex properties in five complexes achieved MCC of 0.51, 0.47, 0.42, 0.74, and 1.00 for independent data set, respectively. We suggest that our study could be a power model for determining new proteins that belongs into which molecular function of electron transport proteins.

1. Introduction

Cellular respiration is the procedure for generating adenosine triphosphate (ATP) and allows cells to gain energy from foods. When we carry out all the activities in our life, cellular respiration is used to make energy inside the shape of ATP (allow our living organism to work). During cellular respiration, cells damage food molecules, such as sugar, and release energy. The goal of cellular respiration is to reap electrons from natural compounds to create ATP, which is used to provide energy for most cellular reactions. As cells go through cellular respiration, they require a pathway to keep and transport electrons (i.e., the electron transport chain). The electron transport chain produces a transmembrane proton electrochemical gradient because of oxidation-reduction reactions. If protons flow back via the ATP synthase through the membrane, ATP synthase converts this mechanical energy into chemical energy through generating ATP, which presents energy in several cellular procedures.

The electron transport chain is a number of protein complexes embedded inside the inner membrane of the mitochondria. Fig. 1 indicates the electron transport chain system. Electrons captured from donor molecules are transferred via these complexes. These complexes are organized into Complex I, Complex II, Complex III, Complex IV, and ATP synthase (which may be called Complex V). Each complex includes numerous specific electron carriers with different molecular functions. At the mitochondrial inner membrane, electrons from nicotinamide adenine dinucleotide (NADH) and succinate bypass through the electron transport chain to oxygen. The most famous molecular function in complex I and complex II are NADH dehydrogenase and succinate dehydrogenase, respectively. Electrons bypass from complex I to a carrier (coenzyme Q) that embeds itself inside the membrane. From coenzyme Q, electrons are handed to complex III (cytochrome b, c1 complex). The pathway from complex III ends in cytochrome c then to complex IV (cytochrome oxidase complex). At the end, the proton electrochemical gradient allows ATP synthase to apply the flow of H+ to generate ATP.

Electron transport proteins and membrane proteins have attracted the interest of numerous researchers due to their relevance in cellular respiration and our existence. For example, Gromiha [1] provided a simple statistical method for discriminat-

* Corresponding author.
E-mail addresses: khanhlee87@gmail.com (N.-Q.-K. Le), yienou@gmail.com (Y.-Y. Ou).

http://dx.doi.org/10.1016/j.jmgm.2017.01.003
1093-3263/

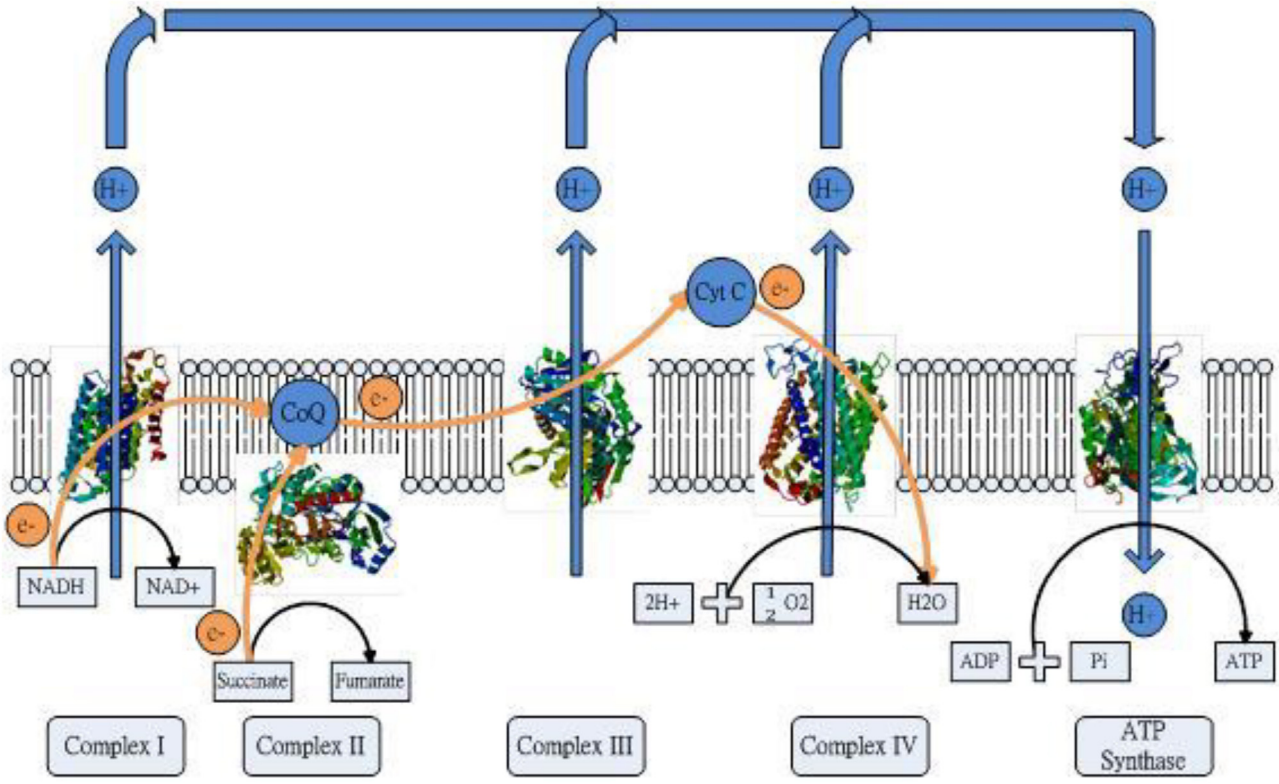

Fig. 1. The process of electron transport Chain.

ing outer membrane proteins with excessive accuracy. Moreover, Ou [2] tried to discriminate of beta-barrel membrane proteins transport by using radial basis function networks (RBFNs) and position specific scoring matrices (PSSM) profiles. The study from Chen [3] divided electron transport proteins into four varieties of transport proteins to behavior prediction and analysis. After the prediction and evaluation, Chen categorized the transport proteins and determined the functions of each protein type inside the transport protein using PSSM profiles and biochemical properties. Then, Ou [4] integrated significant amino acid pairs to identify O-linked glycosylation sites on transmembrane proteins and non-transmembrane proteins.

This study proposes a method primarily based on PSSM profiles and biochemical properties for identifying the category in electron transport proteins from their molecular function. In the first section, we used the set of 2277 transport proteins and 354 electron transport proteins to identify electron transport proteins in transport proteins. This section performed sensitivity of 74.6%, specificity of 95.8%, and accuracy of 92.9%, with Matthews Correlation Coefficient (MCC) of 0.7 for cross-validation dataset. And for the independent dataset, our method achieved sensitivity of 73.2%, specificity of 94.1%, and accuracy of 91.3%, with MCC of 0.64. With second section, we used the variety of electron transport proteins recognized from the first section to do experiment, with 101 electron transport proteins as training dataset and 31 electron transport proteins for the independent test dataset. We implemented the independent dataset to evaluate the performance of the proposed approach, which established an MCC of 0.51, 0.47, 0.42, 0.74, and 1.00, respectively for 5 complexes. In these stages, the essential approach is that using F-score to select 544 biochemical properties adding to PSSM profiles to improve prediction effects. The proposed method has an extensive result and gives beneficial information for biologists. The proposed approach can serve as a powerful model for predicting the categories in electron transport proteins and may help biologists recognize electron transport chain functions, especially the categories in electron transport protein.

2. Materials and methods

This work consist of two stages for discriminating electron transport proteins from transport proteins and classifying categories of five complexes in electron transport proteins. Fig. 2 displays the whole architecture of this work, consists of three subprocesses in each stage: data collection, feature set generation, and model evaluation. From this architecture, we have evolved a novel approach based on PSSM profiles and biochemical properties for discriminating electron transport proteins from transporters and classifying categories of five complexes in electron transport proteins.

2.1. Data collection

First of all, we accumulated transport proteins from the UniProt database [5]. In this section, we eliminated the sequences without the annotation “evidence at protein level” or “complete”. Next, BLAST [6] was used to exclude sequences with a sequence identity of greater than 20% from the dataset. Finally, 2277 transport proteins and 354 electron transport proteins were used in this work. Alternatively, only 132 proteins, which include annotation of complex, are used for second stage to classify categories of five complexes in electron transport proteins. The annotation of complex retrieved from GeneOntology, which contains the descriptions of many gene products for biologists.

We divided the accumulated protein sequences into two data sets: the training dataset and the independent test dataset. In these stages, the training dataset is used for identifying electron transport proteins and evaluating biochemical properties. The independent test dataset is used to assess the overall performance of the pro-