journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Computational determination of the effects of virulent *Escherichia coli* and salmonella bacteriophages on human gut

Marwa Mostafa Mostafa *, *Mohammad Nassef*, *Amr Badr*

Department of Computer Science, Faculty of Computers and Information, Cairo University, Egypt

ARTICLE INFO

Article history:
Received 23 February 2016
Received in revised form
15 June 2016
Accepted 7 July 2016

Keywords:
Salmonella
Escherichia coli (*E. coli*)
Bacteriophage
Domain–Domain Interactions (DDI)
Human gut
Protein–Protein Interaction (PPI)

ABSTRACT

Background and objective: Salmonella and *Escherichia coli* are different types of bacteria that cause food poisoning in humans. In the elderly, infants and people with chronic conditions, it is very dangerous if Salmonella or *E. coli* gets into the bloodstream and then they must be treated by phage therapy. Treating Salmonella and *E. coli* by phage therapy affects the gut flora. This research paper presents a system for detecting the effects of virulent *E. coli* and Salmonella bacteriophages on human gut.

Methods: A method based on Domain–Domain Interactions (DDIs) model is implemented in the proposed system to determine the interactions between the proteins of human gut bacteria and the proteins of bacteriophages that infect virulent *E. coli* and Salmonella. The system helps gastroenterologists to realize the effect of injecting bacteriophages that infect virulent *E. coli* and Salmonella on the human gut.

Results: By testing the system over Enterobacteria phage 933W, Enterobacteria phage VT2-Sa and Enterobacteria phage P22, it resulted in four interactions between the proteins of the bacteriophages that infect *E. coli* O157:H7, *E. coli* O104:H4 and *Salmonella typhimurium* and the proteins of human gut bacterium strains.

Conclusion: Several effects were detected such as: antibacterial activity against a number of bacterial species in human gut, regulation of cellular differentiation and organogenesis during gut, lung, and heart development, ammonia assimilation in bacteria, yeasts, and plants, energizing defense system and its function in the detoxification of lipopolysaccharide, and in the prevention of bacterial translocation in human gut.

1. Introduction

Protein–Protein Interaction (PPI) is becoming one of the major objectives of system biology. It refers to lasting and specific physical contacts established between two or more proteins as a result of biochemical events and/or electrostatic forces. Protein–Protein Interactions play key roles for mediating functions such as sensing the environment, mediating signal transduction, adjusting the activity of metabolic and signaling enzymes, converting energy into physical motion, and maintaining cellular organization.

Salmonella and *Escherichia coli* (*E. coli*) are different types of bacteria and they are considered as the main reasons for food poisoning. Salmonella is spread by ingesting foods that are contaminated by salmonella such as raw eggs, raw meat, eggs, fruits, vegetables, and contaminated water. Contamination takes place when these foods come into contact with animal or

* *Corresponding author.* Department of Computer Science, Faculty of Computers and Information, Cairo University, Egypt.
E-mail addresses: m.mostafa_fci@yahoo.com (M.M. Mostafa); m.nassef@fci-cu.edu.eg (M. Nassef); amr.badr@fci-cu.edu.eg (A. Badr).

http://dx.doi.org/10.1016/j.cmpb.2016.07.018
0169-2607/

human feces and are not cooked properly [1]. *E. coli* infects humans by eating undercooked meat, drinking impure water, drinking unpasteurized (raw) milk and working with cattle [2].

Most people infected by Salmonella get better without treatment. It can be more serious in the elderly, infants and people with chronic conditions. If Salmonella gets into the bloodstream, it can be serious or even life-threatening [3–7]. *E. coli* lives in intestine and most types of *E. coli* are harmless. However, some types can make people sick and cause diarrhea. The worst type of *E. coli* causes bloody diarrhea, and can sometimes cause kidney failure and even death. These problems are most likely to occur in children and in adults with weak immune systems [8–11].

For salmonella species, over 2500 different strains (called "serotypes") have been identified to date. *Salmonella typhimurium*, one of the most important serotypes of salmonellosis transmitted from animals to humans in most parts of the world. *S. typhimurium* causes serious problems to the elderly people and patients with weakened immunity [12].

E. coli O157:H7 is one of hundreds of strains of the bacterium *E. coli*. Although most strains are harmless, this strain produces a powerful toxin that can cause severe illness. *E. coli O157:H7* has been found in the intestines of healthy cattle, deer, goats, and sheep. *E. coli O157:H7* was first recognized as a cause of illness in 1982 during an outbreak of severe bloody diarrhea; the outbreak was traced to contaminated hamburgers [13,14].

Also *E. coli O104:H4* bacteria is one of the bacterium *E. coli* strains that caused a serious outbreak of food borne illness focused in northern Germany in May through June 2011. The illness was characterized by bloody diarrhea, with a high frequency of serious complications, including hemolytic-uremic syndrome (HUS), a condition that requires urgent treatment [15].

Phage therapy which is a process for treating pathogenic bacterial infections becomes a need for treating *S. typhimurium*, *E. coli O157:H7* and *E. coli O104:H4* [16–19]. Despite the effectiveness of phage therapy in treating *S. typhimurium*, *E. coli O157:H7* and *E. coli O104:H4*, it becomes a threatening factor on the diversity and abundance of *microflora* in the gastrointestinal tract.

This research paper presents a system for detecting the effects of bacteriophages that infect *E. coli O157:H7*, *E. coli O104:H4* and Salmonella on human gut.

The remainder of the paper is organized as follows: in section 2, an overview of the previous work related to our subject is presented. In section 3, the architecture of the proposed system is described. In section 4, testing the system and the results are produced, before drawing conclusions and future work in section 5.

2. Related work

Several research studies have been targeted to the subject of domain based model that determines the PPIs.

Jonsson et al. [20] used methods that were established to identify the network topology of a cancer protein network. They report a study of cancer proteins in an extensive human protein–protein interaction (PPI) network constructed by computational methods. These observations indicate an underlying evolutionary distinction between the two groups of proteins, reflecting the central roles of proteins, whose mutations lead to cancer.

Clancy et al. [21] designed a statistical method to infer the physical interactions between two complexes for the human and yeast species. They integrated manually curated protein complexes and physical protein interaction networks and designed a statistical method to identify pairs of protein complexes where the number of protein interactions between a complex pair is due to an actual physical interaction between the complexes.

Lee et al. [22] proposed a one to one domain–domain interaction (DDI) model to obtain specific sets of DDI for oncoproteins and tumor suppressor proteins respectively. Cross-validation test was conducted to benchmark the prediction sensitivity, specificity and F1 measure. They suggested that PPI, which is mediated by DDI, may be affected due to domain removal through the alternative splicing mechanism. They studied the domain removal effects on liver cancer isoforms PPI.

Pohane et al. [23] demonstrated that the interaction between N-terminal catalytic and C-terminal cell wall binding domains of mycobacteriophage D29 endolysin makes the enzyme inactive in *E. coli*. Also, they demonstrated that such interaction occurs intra-molecularly and is facilitated by a charged linker that connects the two domains. They showed that linker composition is crucial for inactivation of PG hydrolase enzyme in *E. coli*. Such knowledge will immensely help in bio-engineering of endolysins with narrow or broad spectrum antimicrobial activity.

Huang et al. [24] proposed a method that extends the Aragues's method by employing the PPI data, DDI data, weighted Domain Frequency Score (DFS), and Cancer Linker Degree (CLD) data to predict cancer proteins. They demonstrated the accuracy of the proposed method on two independent datasets. The best algorithm can achieve a hit ratio of 89.4% and 72.8% for lung cancer dataset and lung cancer microarray study, respectively.

Deng et al. [25] studied the large-scale conserved patterns of interactions between protein domains. Using evolutionarily conserved domains defined in a protein–domain database, called PFAM, they apply a Maximum Likelihood Estimation method to infer interacting domains that are consistent with the observed protein–protein interactions. They estimated the probabilities of interactions between every pair of domains and measured the accuracies of their predictions at the protein level. The authors got several PPIs such as RPS0A interacting with APG17 and TAF40 interacting with SPT3, which are consistent with the functions of the proteins.

Huo, T. et al. [26] reported a systematic flow to predict the Host Pathogen Interactions (HPIs) *between Mycobacterium tuberculosis* and *Homo sapiens* based on sequence motifs. They developed a web accessible database named PATH that holds Protein interactions of *M. tuberculosis* and Human to store these predicted interactions and proteins.

A number of computational methods have been developed for predicting PPIs based on interacting protein sequences and using machine learning techniques [27,28]. Other studies have been done to develop methods based only on information about amino acid sequences [29–31].

Elżbieta Pajtasz-Piasecka et al. [32] presented a study focused on the application of mouse DCs which were activated with