

PUBLISHED BY

# INTECH

open science | open minds

World's largest Science,  
Technology & Medicine  
Open Access book publisher



**2,950+**  
OPEN ACCESS BOOKS



**100,000+**  
INTERNATIONAL  
AUTHORS AND EDITORS



**94+ MILLION**  
DOWNLOADS



**BOOKS**  
DELIVERED TO  
151 COUNTRIES

AUTHORS AMONG

**TOP 1%**  
MOST CITED SCIENTIST



**12.2%**  
AUTHORS AND EDITORS  
FROM TOP 500 UNIVERSITIES



Selection of our books indexed in the  
Book Citation Index in Web of Science™  
Core Collection (BKCI)

Chapter from the book *Current Topics in Salmonella and Salmonellosis*

Downloaded from: <http://www.intechopen.com/books/current-topics-in-salmonella-and-salmonellosis>

Interested in publishing with InTechOpen?  
Contact us at [book.department@intechopen.com](mailto:book.department@intechopen.com)

---

# Computational Identification of Indispensable Virulence Proteins of *Salmonella* Typhi CT18

---

Shrikant Pawar, Izhar Ashraf,  
Kondamudi Manobhai Mehata and  
Chandrajit Lahiri

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/66489>

---

## Abstract

Typhoid infections have become an alarming concern with the increase of multidrug resistant strains of *Salmonella* serovars. The new pathogenic Gram-negative strains are resistant to most antibiotics such as chloramphenicol, ampicillin, trimethoprim, ciprofloxacin and even co-trimoxazole and their derivatives thereby causing numerous outbreaks in the Indian subcontinent, Southeast Asian and African countries. Conventional and modern methods of typing had been adopted to differentiate outbreak strains. However, identifying the most indispensable proteins from the complete set of proteins of the whole genome of *Salmonella* sp., comprising the *Salmonella* pathogenicity islands (SPI) responsible for virulence, has remained an ever challenging task. We have adopted a network-based method to figure out, albeit theoretically, the most significant proteins which might be involved in the resistance to antibiotics of the *Salmonella* sp. An understanding of the above will provide insight into conditions that are encountered by this pathogen during the course of infection, which will further contribute in identifying new targets for antimicrobial agents.

**Keywords:** *Salmonella*, *Salmonella* pathogenicity island, SicA, eigen vector centrality, k-core analysis

---

## 1. Introduction

Food-borne infections are quite common and widely distributed worldwide, though there can be several sources of such diseases. Human Salmonellosis or typhoid, causing systemic infection of the human gastrointestinal tract and diarrhoea, is one such common disease caused by *Salmonella enterica* serovar Typhi. With a prevalence of probably 10 millions of

---

cases and hundreds of thousands of deaths every year [1], the disease has turned out to be a major cause for concern with the emergence of multidrug-resistant (MDR) *Salmonella* strains [2]. Such new strains are resistant to chloramphenicol, ampicillin, trimethoprim, ciprofloxacin and even co-trimoxazole and their derivatives, thereby causing numerous outbreaks in the Indian subcontinent, Southeast Asian and African countries [3, 4]. Thus, newer drugs like cephalosporins and quinolone derivatives needed to be explored to combat the situation [5].

To deal with the threats of multidrug resistance, several health intervention strategies have been undertaken. However, the prospects for finding new antibiotics for several classes of Gram-negative pathogens are especially poor due to the blockades provided by their outer membrane to the entry of some existing antibiotics and expulsion of many of the remainder by their efflux pumps [6]. It has become imperative that the conventional strategies for dealing with such pathogens are less effective or even at times, ineffective completely, to emerge victorious against the strategies for the war waged out by them. In such cases, the complexities posed can be solved by adopting some non-conventional approaches of finding the drug targets for these pathogens. Proteins, being the functional unit of the cell of any living organism, have always been good targets for combating diseases. Diseases, on the other hand, serve as interesting examples of complex protein interactions among several other heterogeneous entities of and between organisms. However, understanding the complexity of such interacting protein partners, especially with respect to the combat against the pathogens, has always been elusive. Thus, analyses of the mosaic mesh or network of interacting proteins, commonly known as protein interaction networks (PINs) can provide sufficient insight to reveal the indispensable virulent proteins for valuable drug targets [7].

Analyses of a PIN, to highlight important and/or indispensable proteins, can be as simple as centrality measurements with respect to the biological scenario. These can start by determining the number of interacting partners of a particular protein to identify its *degree* centrality (DC) which correlates with its biological importance. Thus, high-degree proteins (or hubs) are known to correspond to proteins that are essential [8]. As a protein can be affected locally while interacting with its other partners in the global network, other centrality measures are also given importance based on their relevance. Thus, we have discussed the importance of the measures like *closeness centrality* (CC), *betweenness centrality* (BC) and *eigenvector centrality* (EC) [8] parameters for PIN comprising the *Salmonella* pathogenicity islands (SPI) harbouring the specialized virulent proteins characterized by the type III secretion system (T3SS) among others. Till date, 17 such discrete sets have been reported for *S. Typhi* [9] along with the five SPI (1 till 5) characterized experimentally [10] among which SicA has been identified as the indispensable one in the phylogenetically closest neighbour, *S. enterica* serovar Typhimurium strain LT2 [11].

Again, extracting knowledge of the most indispensable virulence proteins from among the stipulated sets of SPI proteins could be quite insufficient. Thus, we have carried out further analyses of the whole genome of *S. Typhi* CT18 encompassing the decomposition of the whole genome protein interactome to a core of highly interacting proteins through the k-core

analysis approach [12]. We have performed cartographic analyses further to identify the functional modules in the network [13] and predicted the indispensability of certain sets of proteins, which have been shown to be sharing similar functional modules empirically important for drug targets.

## 2. Approach

### 2.1. Dataset collection

Proteins for 17 *Salmonella* pathogenicity islands (SPIs) were collected from an in silico study of SPI for *S. enterica* serovar Typhi strain CT18 [9]. The locus tag of all the proteins of SPI for *S. Typhi* CT18 was fed as queries to the STRING 10.0 biological meta-database [14] to get all the possible interactions of a particular protein (date and time of access: Jul 28 2016 13:07:15). Detailed protein links file under the accession number 220341 in STRING was used to collect all the interactions of the whole genome proteins of *S. Typhi*.

The number of proteins from the different genomic islands starting from SPI-1 till -13 and -15 till -18 were 54, 43, 8, 7, 10, 55, 144, 12, 4, 23, 16, 4, 14, 9, 7, 2 and 97, respectively, with all the combined SPI amounting to a total of 502. The total number of protein interactions obtained from STRING v10 were 334, 339, 3, 21, 9, 192, 1193, 12, 6, 69, 19, 1, 19, 5, 3, 1, 343, for the 17 SPI loci mentioned above and 2570 interactions for all of these combined together. The whole genome of *S. Typhi* had 1041274 interaction information arising out of 4529 unique proteins.

### 2.2. Interactome construction

All individual protein interaction data, with medium confidence values obtained by default from String 10.0, were imported into Cytoscape version 3.3.0 [15] to integrate and build the interactomes of network comprising SPI-1 till -13 and -15 till -18, individually and all these 17 SPI collectively (AS). The interaction information, weighted by their strength as per STRING, of all the proteins of *S. Typhi* genome was imported into Gephi 0.9.1 [16] to construct and visualize the interactome of the whole genome. An interactome of proteins can be perceived as the protein interaction network (PIN) and can be represented as an undirected graph  $G = (V, E)$  consisting of a finite set of  $V$  vertices (or nodes) and  $E$  edges. An edge  $e = (u, v)$  connects two vertices (nodes)  $u$  and  $v$ . Each protein in the above PIN is represented as a vertex/node. The number of connections/interactions/associations/links a node has with other nodes comprises its degree  $d(v)$  [17].

### 2.3. Network analyses

#### 2.3.1. SPI-PIN

All the interactomes of SPI-PIN have been viewed by Cytoscape version 3.3.0 in the form of graphs of aforementioned interconnected proteins. The networks were subsequently analysed via the Cytoscape integrated java plugin CytoNCA [18] to compute values for the network centrality parameters namely EC, DC, CC and BC. Combined scores from different

parameters considered in STRING were taken as edge weights for computing CytoNCA scores. Top 20 proteins for each of the centrality measures were taken for drawing Venn diagrams to find common proteins from each measure.

### 2.3.2. *WhoG-PIN*

As few (21) nodes out of the whole genome were isolated from the major part of network, these were considered to have less impact on the overall topology and thus ignored. Further analyses were based on the large connected component (LCC) of network comprising 4508 protein partners having 1041182 interactions. The analytical study has been done by using MATLAB version 7.11, a programming language developed by MathWorks [19].

For the primary understanding of the network, the distributions of network degree ( $k$ ) were plotted by Complementary Cumulative Distribution Function (CCDF). To extract significant information from the topology of the large and complex Whole Genome Protein Interaction Network (*WhoG-PIN*), knowledge of the role of each protein was derived from the cartographic representation of within-module degree  $z$ -score of the protein versus its participation coefficient as per the methodology described by Guimera et al. [20]. Participation of each protein reflected its positioning within own module and with respect to other modules, where modules were calculated based on Rosvall method [21]. To have an idea of the core group of the very specific proteins which might have variety of role to play in the whole genome context, a  $k$ -core analysis was performed following the network decomposition (pruning) techniques to produce a sequence of subgraph of gradually increasing cohesion [12].

## 3. Features of the 17 SPIs

The virulence proteins of *Salmonella* are spread across the 17 *Salmonella* pathogenicity islands (SPIs) in *S. Typhi* as implied by Ong et al. [9]. Among these, five have been well characterized and reported to have *SicA* as the most indispensable one as identified computationally by Lahiri et al. [11]. A detailed insight into these SPI proteins would reveal SPI-1 and -2 to encode the proteins of the type III secretion systems (T3SSs), while SPI-4 encodes those of type I secretion system (T1SS) mediated by a giant non-fimbrial adhesin, which is co-regulated by the invasion genes encoded by the SPI-1 [22]. The *sit* gene cluster proteins of SPI-1 T3SS, encoding an iron uptake system, are involved in the invasion into the eukaryotic host non-phagocytic cells mediated by the delivery of effectors that directly engage host cell signalling pathways [10]. For the systemic phase of infection, proteins of the SPI-2 cluster are essential for the survival and replication in eukaryotic host cells [23], which are aided by the high-affinity magnesium uptake system encoded by *mg<sub>t</sub>CB*, harboured by SPI-3 [24]. The effector proteins of enteropathogenesis are harboured by SPI-5 and are induced by distinct regulatory cues and targeted to different TTSS, namely, *SopB*,

secreted by SPI1 T3SS and PipB, translocated by SPI-2 T3SS to the *Salmonella*-containing vacuole and *Salmonella*-induced filaments.

The 59 kb SPI-6 consists of a type VI secretion system (T6SS), the *safABCD* fimbrial gene cluster, the invasin *pagN*, two pseudogenes as transposase remnants (*STY0343* and *STY0344*), the fimbrial operon *tfABCD* and the genes *tinR* and *tioA* [25–29]. The largest SPI identified till date is that of SPI-7 with 134 kb size [25, 30, 31] and 150 genes inserted between duplicated *pheU* tRNA sequences [30, 32] containing the Vi capsule biosynthesis genes [33], a type IVB pilus operon [34] and the *SopE* prophage (*ST44*) [35]. SPI-9 is a 16 kb locus containing three genes encoding for a T1SS and one for a large protein [36]. SPI-10 is an island found next to the *leuX* tRNA gene at centisome 93. It is a 33 kb fragment [25] carrying a full P4-related prophage, termed *ST46* [37–39]. *ST46* harbours the *prpZ* cluster as cargo genes encoding eukaryotic-type Ser/Thr protein kinases and phosphatases involved in *S. Typhi* survival in macrophages [40]. SPI-11 is a 10 kb fragment in *S. Typhi* and includes phoP-activated genes *pagD* and *pagC* involved in intramacrophage survival [41, 42]. The 6.3 kb SPI-12 contains the effector *SspH2* [43] along with the three ORFs are pseudogenes (*STY2466a*, *STY2468* and *STY2469*). SPI-13 was initially identified in serovar Gallinarum [44]. In *S. Typhi*, it is a 25-kb gene cluster found next to the *pheV* tRNA gene on centisome 67. The 8-kb portion of this island corresponds to SPI-8 whose virulence function is unknown, and it harbours two bacteriocin immunity proteins (*STY3281* and *STY3283*) and four pseudogenes [25]. SPI-14 is absent in *S. Typhi* [36, 44]. SPI-15 in *S. Typhi* is a 6.5 kb island of five ORFs encoding hypothetical proteins [44]. SPI-16 is a 4.5 kb fragment inserted next to an *argU* tRNA site, and encodes five or seven Open reading frames (ORFs), four of which are pseudogenes, the three remaining ORFs show a high level of identity with P22 phage genes involved in seroconversion [45]. SPI-17 is a 5-kb island encoding six ORFs inserted next to an *argW* tRNA site [45]. SPI-18 was recently identified in *S. Typhi* as a 2.3 kb fragment harbouring only two ORFs: *STY1498* (*clyA*) and *STY1499* [46] of which the former encodes a 34 kDa pore-forming secreted cytolysin [46, 47].

#### 4. The individual and the combined SPI-PINs

To focus upon the most indispensable proteins of the highly complex virulent phenotype as that of *Salmonella*, an integrated picture comprising the involvement of all the SPI and the connected associated proteins must be taken into account. Thus, with an ultimate goal to identify the indispensable virulent proteins for potential candidates of therapeutic targets, we have constructed the PINs or interactomes of the 17 individual SPI mentioned above, along with and a combined network of all of these SPI-PINs (AS). These were then analysed to identify the most important proteins among a group of highest number of interacting partners. This was done by utilizing the four important concepts of centrality applied to biological networks, namely eigenvector centrality (EC), degree centrality (DC), closeness centrality (CC) and betweenness centrality (BC) [48–50].

Amongst the four centrality measures being mentioned above, DC is the most basic as it brings out the involvement of the protein in a large number of interactions in a network. However, in a biological scenario of *Salmonella* infection, having the primary stages as attachment and invasion, the interactions of those proteins may not be in a sequential order so as to carry out a particular function as reflected through DC parametric analyses. In such cases, analyses of CC could be a good measure, which would reveal the close proximities of the proteins expected to communicate sequentially with other network proteins essential for a particular function. Again, a one-to-many type simultaneous interaction of a protein, rendering different functions, is imperative from the complexities of biological phenotype like virulence. Thus, the protein with a high proportion of interactions lying 'in between' and thereby connecting many other proteins in the network would be revealed through BC measures. This could have reflected to be quite an important protein, though it lacks the idea of connecting other important proteins in the network. EC measures the last concept and reflects the indispensable protein connecting other important proteins. A comparative picture of the parametric values of the top 20 rank holders in their descending order have been consolidated and put in a tabular form (**Table 1**). These rankers in either of the cases have the proteins reflected to be important.

There have been three clear trends observed across the topmost rankers of the SPI-PINs for the measures of DC, BC, CC and EC, respectively. In most of the cases, there is a unanimous decision for the top ranking protein showing its utmost importance nearing to indispensability. SPI-PINs of these categories are -1, -3, -4, -5, -7, -8, -9, -10 to -13 and -15 to -17. The other categories have either three or two of the centrality measures conforming to the unanimity of the top ranking proteins. SPI-2, -18 and the all SPI (AS-PIN) have BC differing in the top ranking position whereas SPI-6 and -10 have segregation of DC and EC against CC and BC for the top ranking positions. The common top ranking proteins across these 17 SPI and the AS has been reflected in **Figure 1** with Venn diagrams.

It has been observed that with SPI-1, protein HilA is ranked highest. HilA is the central regulator in SPI-1, which activates the sip operon that is responsible in encoding secreted proteins, as well as the *inv/spa* and *prg* operons encoding components of the secretion apparatus [51, 52]. SPI-2 till -4 has all the secretion apparatus inner membrane proteins SsaG, FidL and STY4452 as the top rankers, respectively. Among the other top rankers, the inositol phosphate phosphatase, SopB, of SPI-5, an atypical fimbria chaperone protein SafB and ImpA-related N-family protein, STY0286, of SPI-6, the pilin protein, PilL, of SPI-7, bacteriocin immunity protein, STY3281, of SPI-8, a large repetitive protein with six Bacterial\_Ig-like domains, t2643, of SPI-9, bacteriophage gene regulatory protein, STY4826, of SPI-10, cytolethal distending toxin protein, CdtB, of SPI-11, uronate isomerase, UxaC, of SPI-13 and the sensory histidine kinase protein, having role in motility and virulence, BarA, of SPI-18 are noteworthy.

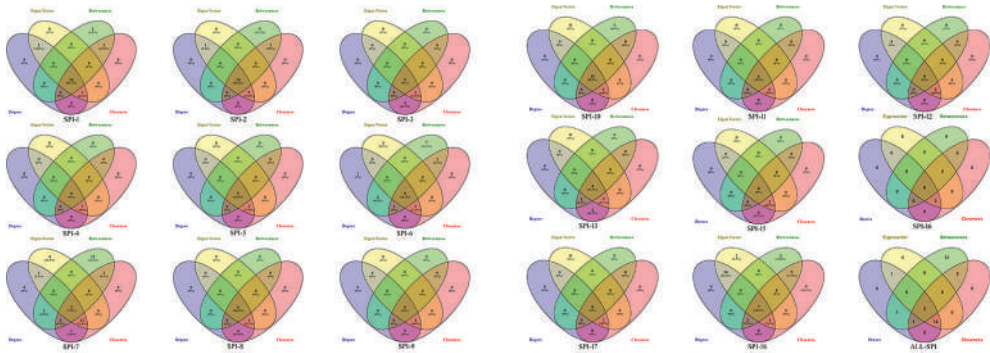
With respect to the above analyses of the individual interactomes of the SPI, an idea about the importance of these proteins in their individual SPI and finally across all SPI could be obtained. However, for a drug to be effective, the indispensability issue of these proteins needs to be taken care of. Thus, a broader picture with respect to the whole genome proteins of *S. Typhi* is then delineated to address the concern.





SPI	Degree	Betweenness	Closeness	Eigenvector
9	12643,STY2876,STY2877,STY2878,	12643,STY2876,STY2877,STY2878,	12643,STY2876,STY2877,STY2878,	12643,STY2876,STY2877,STY2878,
10	STY4826,STY4832,STY4830,STY4822,STY4821,STY4834,STY4849,STY4852,STY4829,STY4823,STY4825,STY4827,STY4823,sefC,sefB,	STY4832,sefC,STY4826,STY4830,STY4843,STY4822,STY4849,STY4852,sefB,STY4821,14521,STY4834,STY4829,STY4828,STY4851,STY4833,STY4825,STY4827,STY4823,	STY4832,STY4826,STY4830,STY4822,STY4849,STY4852,STY4829,STY4828,STY4851,STY4825,STY4827,STY4823,	STY4826,STY4830,STY4822,STY4822,STY4852,STY4821,STY4834,14521,STY4825,STY4827,STY4823,STY4833,STY4829,12655,sefB,
11	cdtB,pagC,envE,STY1879,STY1880,pagD,STY1889,STY1890,STY1891,espH,msgA,STY1887,	cdtB,pagC,envE,STY1879,STY1880,pagD,STY1889,STY1890,STY1891,espH,msgA,STY1887,	cdtB,pagC,envE,STY1879,STY1880,pagD,STY1889,STY1890,STY1891,espH,msgA,STY1887,	cdtB,pagC,envE,STY1879,STY1880,pagD,STY1889,STY1890,STY1891,espH,msgA,STY1887,
12	sppH2,STY2468,	sppH2,STY2468,	sppH2,STY2468,	sppH2,STY2468,
13	uxaC,ordL,STY3296,STY3294,STY3295,STY3298,STY3293,uxuA,uxuB,exuI,STY3302,STY3303,	uxaC,ordL,STY3296,STY3294,STY3295,STY3298,STY3293,uxuA,uxuB,exuI,STY3302,STY3303,	uxaC,ordL,STY3296,STY3294,STY3295,STY3298,STY3293,uxuA,uxuB,exuI,STY3302,STY3303,	uxaC,ordL,STY3296,STY3294,STY3295,STY3298,STY3293,uxuA,uxuB,exuI,STY3302,STY3303,
15	STY0605,gtfB,gtfA,STY3188,STY3189,STY3192,STY3193,	STY0605,gtfB,gtfA,STY3188,STY3189,STY3192,STY3193,	STY0605,gtfB,gtfA,STY3188,STY3189,STY3192,STY3193,	STY0605,gtfB,gtfA,STY3188,STY3189,STY3192,STY3193,
16	STY0605,gtfB,gtfA	STY0605,gtfB,gtfA	STY0605,gtfB,gtfA	STY0605,gtfB,gtfA
17	gtfA2,STY2629,	gtfA2,STY2629,	gtfA2,STY2629,	gtfA2,STY2629,
18	barA,cpX,csrA,flag,flhA,flhB,fljA,fljF,fljH,fljI,fljK,fljO,fljR,fljZ,phoQ,rcsB,rcsC,rpoD,STY1297,yojN,	acrR,baeR,barA,clpP,csrA,dnaK,flag,fljA,hns,mgfA,mntH,ompF,phoQ,rcsB,rcsC,rpoN,rpoS,soxS,STY1297,STY1678,	barA,clpP,cpX,rcsA,dnaK,flag,fljA,grnL,hns,ompB,phoB,phoQ,ompF,phoQ,rcsB,rcsC,rpoN,rpoS,sirA,STY1297,STY1678,yojN,	barA,flag,flhA,flhB,fljD,fljA,fljF,fljH,fljI,fljK,fljO,fljR,fljQ,fljR,fljZ,rcsB,rcsC,rpoS,STY1297,yojN,
All SPI	pill,STY4521,STY4523,STY4526,STY4528,STY4530,STY4534,STY4562,STY4564,STY4569,STY4571,STY4572,STY4573,STY4575,STY4576,STY4577,STY4579,STY4665,STY4666,14268,	barA,pilL,pilV,rpoS,sicA,STY4521,STY4523,STY4526,STY4561,STY4586,STY4592,STY4618,STY4622,STY4644,STY4645,STY4658,STY4664,STY4666,14317,141d,	pill,STY4521,STY4523,STY4528,STY4558,STY4559,STY4562,STY4563,STY4564,STY4568,STY4569,STY4571,STY4572,STY4573,STY4575,STY4576,STY4577,STY4579,STY4665,14268,	pill,STY4521,STY4523,STY4528,STY4558,STY4559,STY4562,STY4563,STY4564,STY4568,STY4569,STY4571,STY4572,STY4573,STY4575,STY4576,STY4577,STY4579,STY4665,14268,

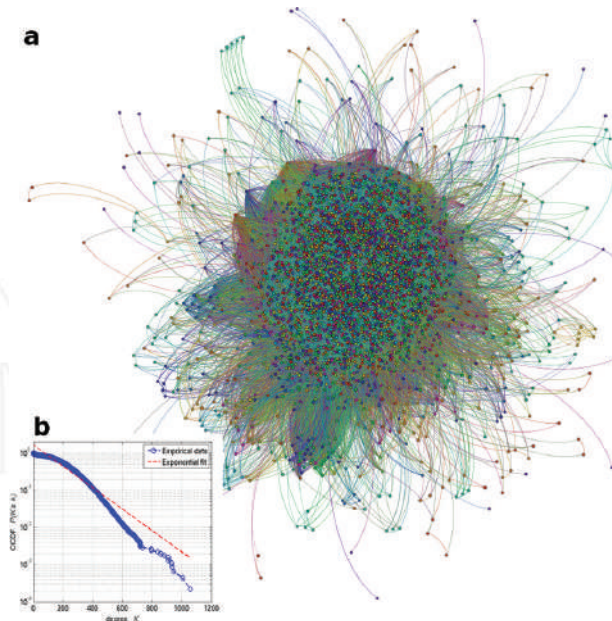
Table 1. Details of the 17 groups of SPI proteins involved in the network.



**Figure 1.** Venn diagram representation for the top rankers of DC, CC, BC and EC parametric analyses of 17 SPI-PINs and AS-PIN.

## 5. Feature of the WhoG-PIN

It is imperative that the WhoG-PIN, built from the empirical and theoretical results of physical and functional interactions among proteins laid down in STRING, can be random like that



**Figure 2.** (a) Protein-protein interaction network of the whole genome of *Salmonella* Typhi CT18 with inset (b) showing degree distribution of the proteins from the large connected component.

proposed by Erdos and Renyi [53] or a small-world type proposed by Watts and Strogatz [54]. The idea was to see if the connectivity distribution,  $P(k)$ , of a node in a network getting connected to  $k$  other nodes, decays exponentially for large values of  $k$ . It was observed that the WhoG-PIN roughly follows the power law and is free of a characteristic scale [55] with a tailed degree distribution (**Figure 2**).

## 6. Decomposition of WhoG-PIN

In order to get an idea of the indispensable ones from the barrage of proteins involved in the individual SPI-PINs and AS, we have performed a  $k$ -core analysis for them. A  $k$ -core is a subgraph whose nodes have degree at least equal to  $k$ . Nodes which are part of  $k$ -core, but not in the  $k+1$  core, is called,  $k$ -shell. This is able to classify the nodes (proteins, in our study) based on the variety of their interacting partners. Proteins, which belong to outer shell, have lower  $k$  value and thus reflect limited number of interacting partner proteins. Moreover, proteins, which belong to inner  $k$ -core/shell, are specific ones, highly interacting with each other and thus can be considered to be the most important ones. Decomposition of this core decomposes the network and thus makes this the innermost core.

After decomposition of the WhoG-PIN, we have obtained the inner core member proteins which are highly robust, central and thus highly interactive in nature [56]. We have arrived to the 154th core with a number of 2180 proteins (**Figure 3**; data not shown). An idea was to look in for the rank holder proteins of the AS-PIN obtained through the EC, DC, CC or BC measures. Interestingly, it was found that the top ranker PilL, across EC, DC and CC measures, belong to the 111th core and not the 154th core. On the contrary, the top ranking BC protein, BarA, was in the 154th core along with the closely ranked PilV in the 150th core. The only other protein, amongst the unanimous top rankers of AS-PIN, STY4521 had a position of 145 in  $k$ -core measures. Very strikingly, two proteins of BC top rankers were also in the 154th innermost core along with BarA. These were the RNA polymerase sigma factor, RpoS and the chaperone protein, SicA. On a note of comparison among the top ranking proteins of EC and BC analysed for AS-PIN, proteins of the latter group had higher ranks in the whole genome context, with STY4586, STY4644 and STY4664 having the same 154th innermost core measures. On the contrary, those from the former ranking group (EC) mostly moved around the core numbers 56–70. This reflected that proteins from the BC rankers were more important in their interaction with other proteins, forming a bridge amongst those and thereby rendering high betweenness.

In an earlier work by Lahiri et al., SicA was found to be in the group of innermost core of the interactome comprising the five most extensively worked out SPI of *S. Typhimurium* [11]. This core group had IacP, InvA, InvB, InvC, InvE, InvF, InvG, InvI, InvJ, OrgA, OrgB, OrgC, PrgH, PrgI, PrgJ, PrgK, SipA, SipB, SipC, SipD, SpaO, SpaQ, SpaS, SpiC, SptP, SsaJ, SseC, SseD and SseF as other members. Referring to the context of *S. Typhi*, IacP, InvE, invF, InvG, PrgK, SpaL (InvC in *S. Typhimurium*), SpaO, SpaS and SptP all shared the same innermost 154th core with a close contestant SsaJ in the 153rd core. Interestingly, all these proteins belong to the SPI-1 and SPI-2 group, which makes up the needle for injecting the virulence factors as delineated in the **Figure 4** of Lahiri et al. All these take us to the juncture where we

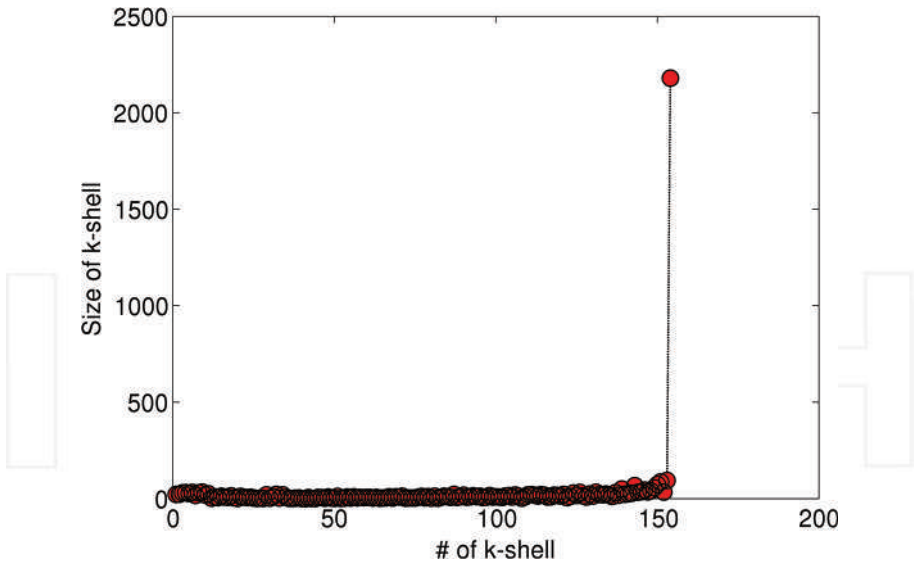


Figure 3. Distribution of the k-shell sizes for the set of proteins from the WhoG-PIN of *S. Typhi* CT18.

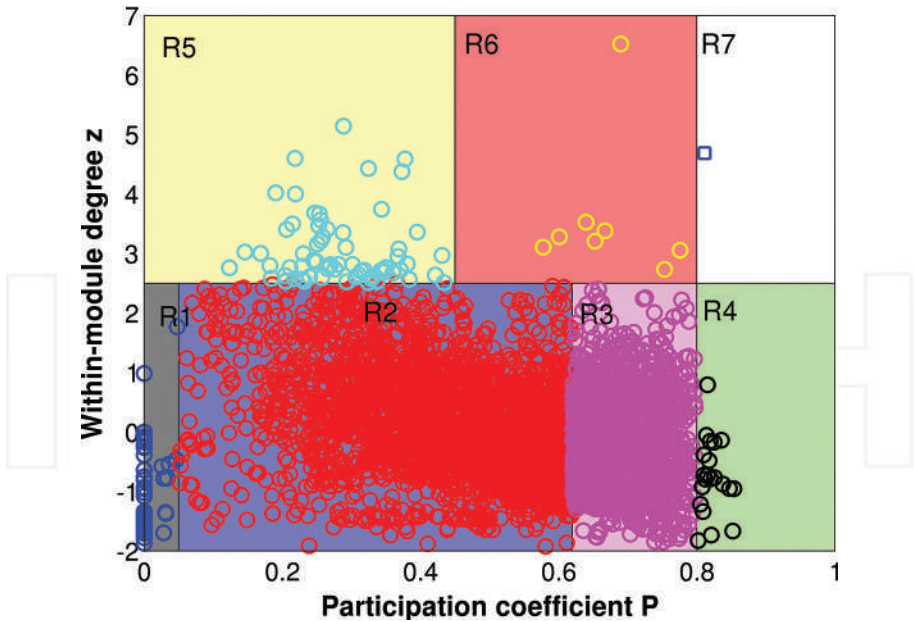


Figure 4. Cartographic representation for classification of proteins from the WhoG-PIN of *S. Typhi* CT18 based on its role and region in network space.

can foresee that the needle proteins are quite important virulence factors when it comes to search targets for drug. To top them all, SicA stands out as being one of the topmost rankers in BC measure of AS-PIN and in the innermost core of the WhoG-PIN. This is quite justified as SicA is a *Salmonella* type III secretion-associated invasin chaperone protein required for the stabilization of SipB and SipC to prevent their premature association which may lead to their targeting for degradation. Along with InvF, SicA is required for transcriptional activation of several virulence genes like *sigDE* (*sopB*, *pipC*), *sipBCDA* and *sopE*. [57].

## 7. Cartographic analyses of WhoG-PIN

For the purpose of classification of the proteins of *S. Typhi* CT18, based on their functional role and region in the network space, we have performed a cartographic analyses for the WhoG-PIN. As described earlier here, this is delineated by within module z-score of each node (protein) and its participation coefficient within and between other modules [20]. The within-module degree z-score measures how 'well connected' a node 'i' is to other nodes in the module, while the participation coefficient measures how the node 'i' is positioned in its own module and with respect to other modules. These measures are done based on the modules of the network, which are calculated by Rosval method [21]. The proteins are mainly divided into two major categories namely the hub nodes and the non-hub nodes.

As can be understood from the name itself, a hub is a connection point of many nodes. The category of non-hub nodes can be assigned four different roles namely, R1 comprising ultra-peripheral nodes, R2 of peripheral nodes, R3 of non-hub connector nodes and R4 having the non-hub kinless nodes. Likewise, the hub nodes can be assigned three different roles namely, R5 of provincial hubs, R6 of connector hubs and R7 of kinless hubs (**Figure 4**). The kinless hubs nodes are supposed to be important in terms of functionality, which has high connection within module as well as between modules. Accordingly, the ultra-peripheral nodes occupy the least connecting position in the network followed by the peripheral nodes. These nodes can be pruned easily without much affecting the whole network while decomposing it to reach the core (refer previous section for k-core). The non-hub connectors are expected to take part in only a small but fundamental set of interactions. This is just opposite to those of the provincial hubs class which have many within-module connections. The non-hub kinless nodes are those with links homogeneously distributed among all modules. The most conserved in terms of decomposition as well as evolution would be, however, those from the connector hubs with many links to most of the other modules. The system would try to retain these connections as essential ones for their very survival.

As can be perceived from the above classification of the connectors and the hubs, the proteins belonging to the R4, R6 and R7 role players are very crucial and can be regarded as potential drug targets. In the context of our WhoG-PIN, the only one R7 is a putative transposase, STY0115 and reminds of the Tn5 transposase, the enzyme that helps bacteria to share antibiotic resistance genes [58, 59]. This is closely followed by the plasmid transfer protein, TrhC

Protein name	R	Description of function
STY0115	7	Putative transposase
trhC	6	Plasmid transfer protein
gltB	6	Glutamate synthase (NADPH) large subunit
ptsG	6	PTS system glucose-specific transporter subunit IIBC
hemE	6	Uroporphyrinogen decarboxylase
nagE	6	PTS system N-acetylglucosamine-specific transporter subunit IIBC
STY3507	6	Aerobic respiration control sensor protein
t0287	6	PTS system sucrose-specific transporter subunit IIBC
treB	6	PTS system trehalose-specific transporter subunit IIBC
Cat	4	Chloramphenicol acetyltransferase
pspF	4	Phage shock protein operon transcriptional activator
STY4151	4	Acetyltransferase
STY4518	4	Acetyltransferase
STY4668	4	Hypothetical protein with Acetyltransf domain
STY4678	4	Integrase
STY4680	4	Integrase
STY0326	4	Hypothetical protein
STY3695	4	DNA-invertase
modB	4	Molybdenum transporter permease
STY4017	4	Putative transferase
modA	4	Periplasmic molybdenum-binding protein
sopE	4	Guanine nucleotide exchange factors
STY1020	4	Sequence-specific DNA binding
STY3193	4	Hypothetical protein
ugpB	4	Glycerol-3-phosphate-binding periplasmic protein
tviA	4	Flagellar regulator
hpaG	4	Isomerase/decarboxylase
STY4175	4	Hypothetical protein
ratA	4	C54 island protein
livG	4	High-affinity branched-chain amino acid transport ATP-binding protein LivG
STY0352	4	Periplasmic protein

**Table 2.** Functions of the R4, R6 and R7 Proteins from the WhoG-PIN cartographic analysis.

in R6 group. This could very well play a good target for drugs as plasmids are known to be powerhouse of the antibiotic resistance genes [60]. Uncoupling of phosphotransferase system could also be an effective way of getting targets for novel drugs as exemplified by PtsG, TreB, NagE and t0287 [61]. Inhibition of glutamate Synthase, GltB has already been utilized as target for *Mycobacterium tuberculosis* [62] as has been uroporphyrinogen decarboxylase, HemE, *albeit* in a different context [63]. Recently, bacterial GCN5-related N-acetyltransferases of the R4 group have been thought of as essential drug targets as well [64]. All the functions of R7, R6 and R4 are listed in **Table 2**.

## 8. Conclusion

This work schematically delineates a process of figuring out the most indispensable protein in a system of interacting proteins of *S. Typhi*. It deals with the computational framework of building of the theoretical networks comprising the 17 individual SPI-PINs along with the AS-PIN followed by the conventional parametric approach of identifying the most interacting protein connected to other important proteins in the concerned phenotype of virulence. This is reinforced by the analysis of disintegrating the WhoG-PIN to the innermost core of the proteins, essential for virulence. All these lead to the identification of SicA to be the most indispensable one amongst a group of other virulent proteins being benefitted through network centrality and decomposition analyses. A further investigation of the WhoG-PIN brought forth the proteins of important conserved class, potential enough to be the most important ones and thus indispensable among the barrage of other proteins of the whole genome of *S. Typhi* CT18.

## Acknowledgements

The authors wish to acknowledge the support of IMSc, Chennai and Dept. of Computer Applications at BSAU, Chennai for the provision of computational facilities. The personal contribution of Ong Su Yean for the SPI data and of Indrajeet Chakraborty for the formatting are highly appreciated and acknowledged.

## Author contributions

CL conceived the concepts, planned and designed the analyses. SP and MIA contributed equally for producing the data analysed by CL. Artwork was done by MIA and SP. CL primarily wrote and edited the manuscript aided by additional help from SP, MIA and KMM.

## Conflict of interest

The authors declare that they have no conflict of interest.

## Author details

Shrikant Pawar<sup>1#</sup>, Izhar Ashraf<sup>2,3#</sup>, Kondamudi Manobhai Mehata and Chandrajit Lahiri<sup>4\*</sup>

\*Address all correspondence to: [chandrajitl@sunway.edu.my](mailto:chandrajitl@sunway.edu.my)

1 Department of Biology and Department of Computer Science, Georgia State University, Atlanta, GA, USA

2 The Institute of Mathematical Sciences, Chennai, India

3 B.S. Abdur Rahman University, Vandalur, Chennai, India

4 Department of Biological Sciences, Sunway University, Selangor, Malaysia

<sup>#</sup>These authors contributed equally to this work.

## References

- [1] Tacket CO, Levine MM. Human typhoid vaccines - old and new. In: DAA Ala'Aldeen, CE Hormaeche, eds. *Molecular and Clinical Aspects of Bacterial Vaccine Development*. England: John Wiley and Sons, 1995; 119–78.
- [2] *Salmonella* (non-typhoidal) [Internet]. 2013. Available from: <http://www.who.int/media-centre/factsheets/fs139/en/> [Accessed: 2016-08-20].
- [3] Rowe B, Ward LR, Threlfall EJ Multidrug-resistant *Salmonella* Typhi: a worldwide epidemic. *Clinical Infectious Diseases*. 1997; 24:S106–S109.
- [4] Senthilkumar B, Prabakaran G. Multidrug resistant *Salmonella* typhi in asymptomatic typhoid carriers among food handlers in Namakkal district, Tamil Nadu. *Indian Journal of Medical Microbiology*. 2005; 23(2):92.
- [5] Chitnis V, Chitnis D, Verma S, Hemvani N. Multidrug resistant *Salmonella* Typhi in India. *The Lancet*. 1999; 354:514–15.
- [6] Eliopoulos GM, Maragakis LL, Perl TM. *Acinetobacter baumannii*: epidemiology, antimicrobial resistance, and treatment options. *Clinical Infectious Diseases*. 2008; 46(8): 1254–63.
- [7] Pan A, Lahiri C, Rajendiran A, Shanmugham B. Computational analysis of protein interaction networks for infectious diseases. *Briefings in Bioinformatics*. 2016; 17(3):517–26.
- [8] Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001; 411(6833):41–2.
- [9] Ong SY, Fui Ling N, Siti SB, Anton Y, Maqsudul A. Analysis and construction of pathogenicity island regulatory pathways in *Salmonella enterica* serovar Typhi. *Journal of Integrative Bioinformatics*. 2010; 7(1):145.



- [10] Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. *Clinical Microbiology Reviews*. 2004; 17(1):14–56.
- [11] Lahiri C, Pawar S, Sabarinathan R, Ashraf MI, Chand Y, Chakravortty D. Interactome analyses of *Salmonella* pathogenicity islands reveal SicA indispensable for virulence. *Journal of Theoretical Biology*. 2014; 363:188–97.
- [12] Seidman SB. Network structure and minimum degree. *Social Networks*. 1983; 5(3):269–87.
- [13] Guimera R, Amaral LA. Functional cartography of complex metabolic networks. *Nature*. 2005; 433(7028):895–900.
- [14] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*. 2014; 43(D1):D447–52.
- [15] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*. 2003; 13(11):2498–504.
- [16] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *ICWSM*. 2009; 8:361–2.
- [17] Diestel R. *Graph Theory*. Berlin and Heidelberg: Springer-Verlag, 2000.
- [18] Tang Y, Li M, Wang J, Pan Y, Wu FX. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems*. 2015; 127:67–72.
- [19] MATLAB and Statistics Toolbox Release. The MathWorks, Inc., Natick, Massachusetts, United States, 2010.
- [20] Guimera R, Amaral LA. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*. 2005; (02):P02001.
- [21] Rosvall M, Bergstrom CT. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PloS ONE*. 2011; 6(4):e18209.
- [22] Gerlach RG, Jäckel D, Stecher B, Wagner C, Lupas A, Hardt WD, Hensel M. *Salmonella* Pathogenicity Island 4 encodes a giant non-fimbrial adhesin and the cognate type 1 secretion system. *Cellular Microbiology*. 2007; 9(7):1834–50.
- [23] Hensel M. *Salmonella* pathogenicity island 2. *Molecular Microbiology*. 2000; 36(5):1015–23.
- [24] Blanc-Potard AB, Solomon F, Kayser J, Groisman EA. The SPI-3 pathogenicity island of *Salmonella enterica*. *Journal of Bacteriology*. 1999; 181(3):998–1004.
- [25] Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature*. 2001; 413(6858):848–52.
- [26] Lambert MA, Smith SG. The PagN protein of *Salmonella enterica* serovar Typhimurium is an adhesin and invasin. *BMC Microbiology*. 2008; 8(1):1.

- [27] Folkesson A, Advani A, Sukupolvi S, Pfeifer JD, Normark S, Löfdahl S. Multiple insertions of fimbrial operons correlate with the evolution of *Salmonella* serovars responsible for human disease. *Molecular Microbiology*. 1999; 33(3):612–22.
- [28] Townsend SM, Kramer NE, Edwards R, Baker S, Hamlin N, Simmonds M, Stevens K, Maloy S, Parkhill J, Dougan G, Bäumlér AJ. *Salmonella enterica* serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infection and Immunity*. 2001; 69(5):2894–901.
- [29] Porwollik S, McClelland M. Lateral gene transfer in *Salmonella*. *Microbes and Infection*. 2003; 5(11):977–89.
- [30] Pickard D, Wain J, Baker S, Line A, Chohan S, Fookes M, Barron A, Gaora PÓ, Chabalgoity JA, Thanky N, Scholes C. Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding *Salmonella enterica* pathogenicity island SPI-7. *Journal of Bacteriology*. 2003; 185(17):5055–65.
- [31] Bueno SM, Santiviago CA, Murillo AA, et al. Precise excision of the large pathogenicity island, SPI7, in *Salmonella enterica* serovar Typhi. *Journal of Bacteriology*. 2004; 186:3202–13.
- [32] Hansen-Wester I, Hensel M. Genome-based identification of chromosomal regions specific for *Salmonella* spp. *Infection and Immunity*. 2002; 70(5):2351–60.
- [33] Hashimoto Y, Li N, Yokoyama H, Ezaki T. Complete nucleotide sequence and molecular characterization of ViaB region encoding Vi antigen in *Salmonella* Typhi. *Journal of Bacteriology*. 1993; 175(14):4456–65.
- [34] Zhang XL, Tsui IS, Yip CM, Fung AW, Wong DK, Dai X, Yang Y, Hackett J, Morris C. *Salmonella enterica* serovar Typhi uses type IVB pili to enter human intestinal epithelial cells. *Infection and Immunity*. 2000; 68(6):3067–73.
- [35] Miroid S, Rabsch W, Rohde M, Stender S, Tschäpe H, Rüssmann H, Igwe E, Hardt WD. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella* Typhimurium strain. *Proceedings of the National Academy of Sciences*. 1999; 96(17):9845–50.
- [36] Morgan E, Campbell JD, Rowe SC, Bispham J, Stevens MP, Bowen AJ, Barrow PA, Maskell DJ, Wallis TS. Identification of host-specific colonization factors of *Salmonella enterica* serovar Typhimurium. *Molecular Microbiology*. 2004; 54(4):994–1010.
- [37] Edwards RA, Matlock BC, Heffernan BJ, Maloy SR. Genomic analysis and growth-phase-dependent regulation of the SEF14 fimbriae of *Salmonella enterica* serovar Enteritidis. *Microbiology*. 2001; 147(10):2705–15.
- [38] Thomson N, Baker S, Pickard D, Fookes M, Anjum M, Hamlin N, Wain J, House D, Bhutta Z, Chan K, Falkow S. The role of prophage-like elements in the diversity of *Salmonella enterica* serovars. *Journal of Molecular Biology*. 2004; 339(2):279–300.
- [39] Bishop AL, Baker S, Jenks S, Fookes M, Gaora PÓ, Pickard D, Anjum M, Farrar J, Hien TT, Ivens A, Dougan G. Analysis of the hypervariable region of the *Salmonella enterica* genome associated with tRNA<sup>leuX</sup>. *Journal of Bacteriology*. 2005; 187(7):2469–82.

- [40] Faucher SP, Forest C, Beland M, Daigle F. A novel PhoP-regulated locus encoding the cytolysin ClyA and the secreted invasin TaiA of *Salmonella enterica* serovar Typhi is involved in virulence. *Microbiology*. 2009; 155(2):477–88.
- [41] Miller SI, Kukral AM, Mekalanos JJ. A two-component regulatory system (phoP phoQ) controls *Salmonella* Typhimurium virulence. *Proceedings of the National Academy of Sciences*. 1989; 86(13):5054–8.
- [42] Gunn JS, Alpuche-Aranda CM, Loomis WP, Belden WJ, Miller SI. Characterization of the *Salmonella* Typhimurium pagC/pagD chromosomal region. *Journal of Bacteriology*. 1995; 177(17):5040–7.
- [43] Miao EA, Scherer CA, Tsolis RM, Kingsley RA, Adams LG, Bäumlner AJ, Miller SI. *Salmonella* Typhimurium leucine-rich repeat proteins are targeted to the SPI1 and SPI2 type III secretion systems. *Molecular Microbiology*. 1999; 34(4):850–64.
- [44] Shah DH, Lee MJ, Park JH, Lee JH, Eo SK, Kwon JT, Chae JS. Identification of *Salmonella* Gallinarum virulence genes in a chicken infection model using PCR-based signature-tagged mutagenesis. *Microbiology*. 2005; 151(12):3957–68.
- [45] Vernikos GS, Parkhill J. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics*. 2006; 22(18):2196–203.
- [46] Fuentes JA, Villagra N, Castillo-Ruiz M, Mora GC. The *Salmonella* Typhi hlyE gene plays a role in invasion of cultured epithelial cells and its functional transfer to *S. Typhimurium* promotes deep organ infection in mice. *Research in Microbiology*. 2008; 159(4):279–87.
- [47] Green J, Baldwin ML. The molecular basis for the differential regulation of the hlyE-encoded haemolysin of *Escherichia coli* by FNR and HlyX lies in the improved activating region 1 contact of HlyX. *Microbiology*. 1997; 143(12):3785–93.
- [48] Koschützki D, Schreiber F. Comparison of centralities for biological networks. In: Griegerich, R. Stoye, J. (eds.) *Proceedings of the German Conference on Bioinformatics (GCB2004)*, Lecture Notes in Informatics, 2004; P-53, pp.199–206.
- [49] Özgür A, Vu T, Erkan G, Radev DR. Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics*. 2008; 24(13):i277–85.
- [50] Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, Schneider R, Bagos PG. Using graph theory to analyze biological networks. *BioData Mining*. 2011; 4(1):1.
- [51] Altier C. Genetic and environmental control of *Salmonella* invasion. *Journal of Microbiology*. 2005; 43(Spec No):85–92.
- [52] Hueck CJ, Hantman MJ, Bajaj V, Johnston C, Lee CA, Miller SI. *Salmonella* Typhimurium secreted invasion determinants are homologous to *Shigella lpa* proteins. *Molecular Microbiology*. 1995; 18(3):479–90.

- [53] Erdős P, Rényi A. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*. 1960; 5:17–61.
- [54] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*. 1998; 393:440–2.
- [55] Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. *Nature*. 2000; 406:378–82.
- [56] Alvarez-Hamelin I, Dall'Asta L, Barrat A, Vespignani A. *Advances in Neural Information Processing Systems 18*. Cambridge, MA: MIT Press, 2006; pp. 41–50.
- [57] Darwin KH, Robinson LS, Miller VL. SigE is a chaperone for the *Salmonella enterica* serovar Typhimurium invasion protein SigD. *Journal Bacteriology*. 2001; 183:1452–4.
- [58] Meredith TC, Wang H, Beaulieu P, Gründling A, Roemer T. Harnessing the power of transposon mutagenesis for antibacterial target identification and evaluation. *Mobile Genetic Elements*. 2012; 2(4):171–8.
- [59] Devaud M, Kayser FH, Bächli B. Transposon-mediated multiple antibiotic resistance in *Acinetobacter* strains. *Antimicrobial Agents Chemotherapy*. 1982; 22(2):323–9.
- [60] Williams JJ, Hergenrother PJ. Exposing plasmids as the Achilles' heel of drug-resistant bacteria. *Current Opinion in Chemical Biology*. 2008; 12(4):389–99.
- [61] Garcia De Gonzalo CV, Denham EL, Mars RAT, Stülke J, van der Donk WA, van Dijl JM. The phosphoenolpyruvate:sugar phosphotransferase system is involved in sensitivity to the glucosylated bacteriocin sublancin. *Antimicrobial Agents Chemotherapy*. 2015; 59:6844–54.
- [62] Mowbray SL, Kathiravan MK, Pandey AA, Odell LR. Inhibition of glutamine synthetase: a potential drug target in *Mycobacterium tuberculosis*. *Molecules*. 2014; 19:13161–76.
- [63] Tsou Y-A, Chen K-C, Lin H-C, Chang S-S, Chen CY-C. Uroporphyrinogen decarboxylase as a potential target for specific components of traditional Chinese medicine: a virtual screening and molecular dynamics study. *PLoS ONE*. 2012; 7(11), e50087.
- [64] Favrot L, Blanchard JS, Vergnolle O. Bacterial GCN5-related N-acetyltransferases: from resistance to regulation. *Biochemistry*. 2016; 55:989–1002.

