Original Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

IDENTIFICATION OF NOVEL DRUG TARGETS AGAINST *CAMPYLOBACTER JEJUNI* USING METABOLIC NETWORK ANALYSIS

ANUKRITI VERMA, BHAWNA RATHI* AND SHIVANI SHARDA

Amity Institute of Biotechnology J-3 Block, Amity University Campus, Sector – 125 Noida - 201303 (U.P.)

ABSTRACT

Campylobacter jejuni is a gram- negative bacteria that is known to cause Guillain-Barre syndrome (GBS) that can lead to paralysis of the human body. It is necessary to find its essential drug targets in order to curb and control GBS. The large amount of genomic/proteomic data in collaboration with systems biology approach can lead to the identification of potential and putative pathogenic drug targets. The combination of subtractive proteomics approach and metabolic network reconstruction was employed for finding novel targets in *Campylobacter jejuni*. Our study revealed that this pathogenic microorganism consists of 228 essential genes present in database of essential genes that are non- paralogous. 151 sequences were non- homologous to Homo sapiens. The KO numbers of 94 essential genes could be retrieved out of which 36 genes were involved in metabolism. The metabolic network of these sequences was created and simulation analysis of this network was performed that produced 6 essential reactions including 2 proteins UDP-N-acetylmuramate dehydrogenase encoding murB gene and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase encoding ispH and lytB genes. These can be used as prospective drug targets.

KEYWORDS: ***Campylobacter jejuni, In-silico*** **subtractive proteomics, Metabolic network reconstruction, Flux balance analysis.**

BHAWNA RATHI
Amity Institute of Biotechnology J-3 Block, Amity University Campus, Sector – 125 Noida - 201303 (U.P.)

INTRODUCTION

Guillain-Barre syndrome (GBS) is an immune-mediated neurological disorder that causes abnormal myelination of peripheral nervous system. It can lead to areflexic paralysis and can be fatal also. Its annual incidence is 0.6–4 cases per 100,000 worldwide. It is reported that around 25%–40% of GBS patients experience *Campylobacter jejuni* infection 1–3 weeks preceding the illness [1]. The symptoms include numbness, paresthesia, weakness and pain in the limbs [2]. The diagnosis of Guillain-Barré syndrome is based upon the differential diagnosis pattern (clinical features and laboratory analysis) so the early diagnosis of GBS is difficult [3]. *Campylobacter jejuni* is also found to trigger reactive arthritis that is an autoimmune disorder which causes inflammation in the joints due to certain infections at distant sites in the body. The diagnosis of reactive arthritis is also based upon a differential diagnosis pattern and hence its early diagnosis becomes difficult that leads to the chronicity of the disease [4]. Therefore to generate improved diagnostic tools and therapeutics it is necessary to detect novel therapeutic drug targets. Of late the use of proteomics and bioinformatics data has become a potential tool to find new drug targets [5]. The amalgamation of subtractive proteomics techniques and reconstruction of metabolic networks aids to discover probable and plausible drug targets of the pathogenic bacteria [6, 7, 8, 9, 10] that has been incorporated in various studies such as for finding novel targets of *Streptococcus suis.* Recent work on the creation and analysis of metabolic network has been done on *Scheffersomyces stipitis* and *Pseudomonas stutzeri A1501* in order to analyse their metabolism [11, 12] The current study incorporates the use of subtractive proteomics approach and metabolic network reconstruction to decipher promising drug candidates in *Campylobacter jejuni* so as to improve the diagnostic criteria and treatment regime in future.

MATERIALS AND METHODS

The schematic representation of *In-silico* analysis of *Campylobacter jejuni* comprising of systematic unification of subtractive proteomics and metabolic network reconstruction approaches is depicted in figure 1.

(i) Subtractive proteomics approach

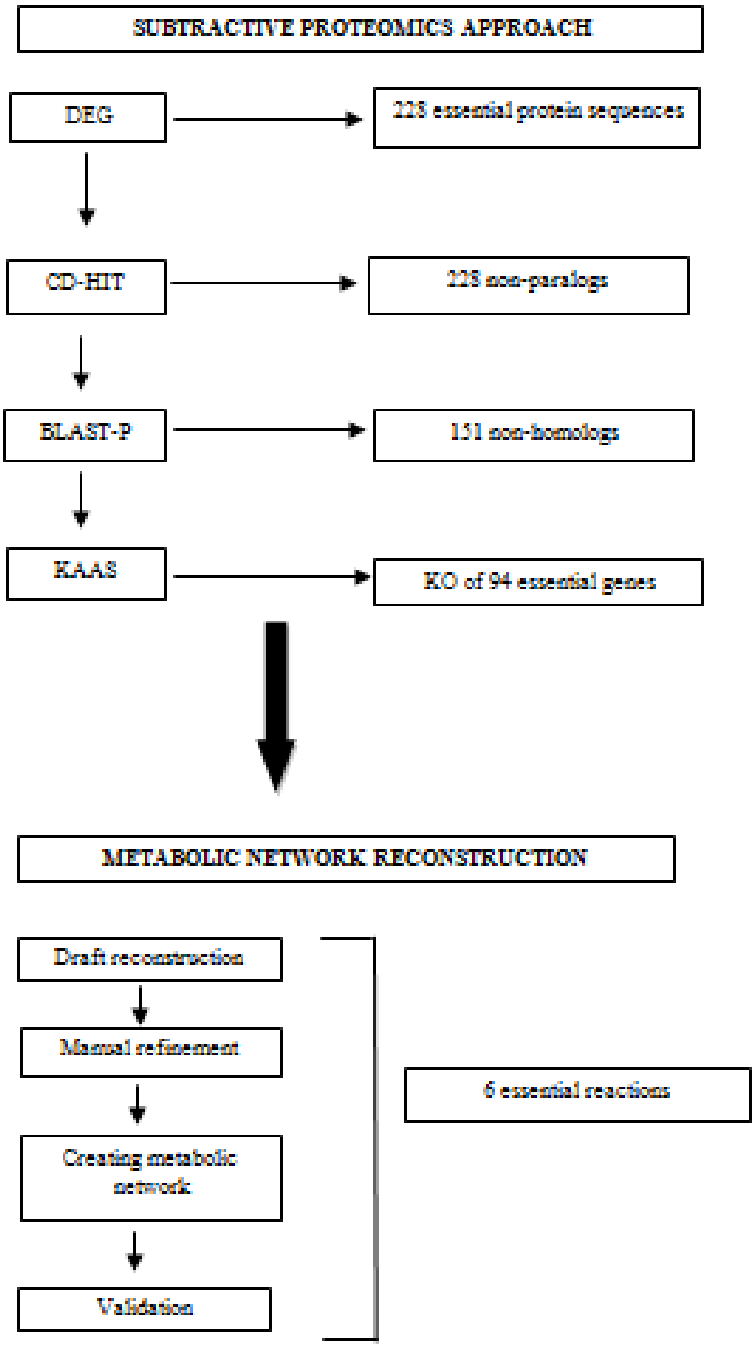
The essential protein sequences of *Campylobacter jejuni* were retrieved from Database of Essential Genes (DEG) (http://tubic.tju.edu.cn/deg/) [13] that includes exceptional targets required for antibacterial drugs. CD-HIT Suite was used for clustering and comparing DEG sequences of *Campylobacter jejuni* at 60% identity level (http://weizhong-lab.ucsd.edu/cdhit_suite/) [14]. The non-paralogous sequences were aligned using BLAST-P (Basic Local Alignment Search Tool) at E-value cutoff of 10^{-4} in order to infer homologous sequences against human database. KAAS (KEGG Automatic Annotation Server) (http://www.genome.jp/tools/kaas/) was used that performed BLAST of non-homologous protein sequences against a database of *Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819, Campylobacter jejuni subsp. jejuni 81-176, Campylobacter jejuni subsp. jejuni 81116, Campylobacter jejuni RM1221 and Campylobacter jejuni subsp. doylei 269.97* to automatically allocate KEGG orthology (KO) identifiers or K numbers to these sequences [15].

(ii) Metabolic network reconstruction

The draft reconstruct of the essential genes with assigned KO numbers was made using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database [16]. The information regarding protein name, gene name, EC number, reaction ID and pathways was obtained. There might be few incorrectly incorporated information or other missing links present, so the whole reconstructed draft is re-evaluated and refined using other databases. ClubSub-P was used for subcellular localization prediction of protein sequences involved in metabolism that links various prevailing subcellular localization prediction tools and eradicates the false-positive and false-negative predictions by providing an elevated precision output [17] (http://toolkit.tuebingen.mpg.de/clubsubp/).Metnetmaker software was used to create the network for the essential genes involved in metabolic pathways. It assigned flux to the reactions present within the constructed network [18] that was visualized using cytoscape software [19]. SurreyFBA was used for constraint-based modeling of FBA ready metabolic model obtained from metnetmaker [20]. It provides information regarding the number of metabolites and number of reactions present in the network and performs flux balance analysis (FBA), flux variability analysis (FVA), identifies live reactions, orphan metabolites, connected components and essential reactions present in the network.

Figure 1

Schematic representation of In-silico analysis of Campylobacter jejuni to identify putative drug targets



RESULTS AND DISCUSSION

In this paper the *In-silico* analysis of *Campylobacter jejuni* has been performed employing the subtractive proteomics approach and creating a metabolic network in order to recognize potential drug targets. 228 essential genes of *Campylobacter jejuni* was retrieved from database of essential genes (DEG) that contains the genes that play a vital role to carry and maintain the cellular activities [21]. These were subjected to CD-HIT suite that indicates the presence of non-paralogous sequences [14] from which 228 clusters (non-paralogs) were obtained. These non-paralogs were subjected to sequence alignment using BLAST-P that revealed 151 non-homologous sequences to the human proteome. The KO numbers of 94 essential genes were acquired using KAAS that directly links these KO numbers with the KEGG orthology database that connects the genomic information to the functional information through KEGG PATHWAY mapping and BRITE mapping [15] (table 1). KEGG database revealed the genes involved in various pathways. Out of these 36 genes are involved in metabolism including few genes involved in multiple pathways of metabolism that are global and overview maps, carbohydrate metabolism, energy metabolism, lipid metabolism, nucleotide metabolism, amino acid metabolism, metabolism of other amino acids, glycan biosynthesis and metabolism, metabolism of cofactors and vitamins, metabolism of terpenoids and polyketides and xenobiotics biodegradation and metabolism. Among these 2 genes did not have EC numbers and were removed and rest

34 genes were considered for further analysis. The subcellular localization information of proteins obtained from ClubSub-P postulates significant signs to their function (compartment) in a cell [17]. Metabolic network reconstruction of these 34 genes was done in order to analyze the metabolic pathways to comprehend the correlation between genome and physiology of an organism [22]. An FBA ready model by assigning maximum positive and negative fluxes and selecting all reactions as objective functions of the 34 essential genes whose KEGG pathways that are included in metnetmaker software was obtained and visualized using the cytoscape [19] (Figure 2 and 3). It revealed that 98 orphan metabolites are used by less than two reactions and 6 live reactions that carry a steady state flux are present. The network is divided into 7 components (one component consisting of 51 connected reactions, one component consisting of 4 connected reactions, 3 components consisting of 2 connected reactions and 2 components consisting of one single reaction). Surrey FBA was used to perform constrained based analysis to assess the metabolic capabilities under a set of constraints of the reconstructed metabolic networks [20]. Here the predefined constraints were maximum allowable fluxes provided by metnetmaker that defines a space or rate by which each reaction consumes or produces each metabolite [9]. FBA (Flux Balance Analysis) was performed which is used to calculate the metabolic flow through the metabolic network. It represents the metabolic reactions as a stoichiometric matrix with m compounds and n reactions and performs mass balance analysis at steady state to maximize or minimize an objective function consisting of predefined combination of linear fluxes. This utilizes linear programming and gives a particular steady-state flux distribution as an output [9]. The output resulted in 4 reactions carrying a steady state flux that determines that the flux that produces a metabolite that equals to flux that consumes a metabolite. Such reactions can be used further for essentiality or deletion studies in order to portray the metabolic efficiency of the network and identify drug targets [23]. The disadvantage of FBA is that it generates a single flux distribution to achieve the optimal objective function. There might be more solutions to such problems. These alternative solutions can be provided by FVA (Flux variability analysis). FVA was performed to find the minimum and maximum values of the fluxes of the reactions without hampering the optimal objective function value [24]. This resulted in 6 reactions out of which 4 coincide with the FBA results (table 2). The reactions are considered essential if their fluxes are made null (0 value) and this in turn makes the objective value null (in this case if the essential reactions are inactivated then they will inactivate other reactions also). The analysis of this network revealed 6 reactions that were similar to reactions found after performing FVA to be essential that consists of 17 metabolites and three genes namely murB encoding UDP-N-acetylmuramate dehydrogenase having EC number 1.3.1.98 involved in carbohydrate metabolism and glycan biosynthesis and metabolism and ispH and lytB encoding 4-hydroxy-3-methylbut-2-enyl diphosphate reductase having EC number 1.17.1.2 involved in metabolism of terpenoids and polyketides (table 3). Many latest reports have also incorporated the construction of a metabolic network and its analysis for microorganisms such as *Scheffersomyces stipitis* and *Pseudomonas stutzeri A1501* in order to enlighten their metabolism [11, 12] The *In- silico* structural characterization of UDP-N-acetylmuramate dehydrogenase and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase has already been performed by researchers for *Mycobacterium tuberculosis* and *Plasmodium falciparum* identifying essential drug targets against tuberculosis and malaria respectively [25, 26] Further structural analysis and identification of inhibitors of essential reactions obtained from our study can help to decipher putative drug targets in future [10, 27, 28].

Table 1
In-silico subtractive proteomics approach

Protein sequences in Database of essential genes (DEG)	228
Non- paralogous sequences	228
Paralogous sequences	0
on-homologous sequences	151
Homologous sequences	77
Sequences having KO identifiers	94

Figure 2
Reconstructed metabolic network of essential genes of Campylobacter jejuni

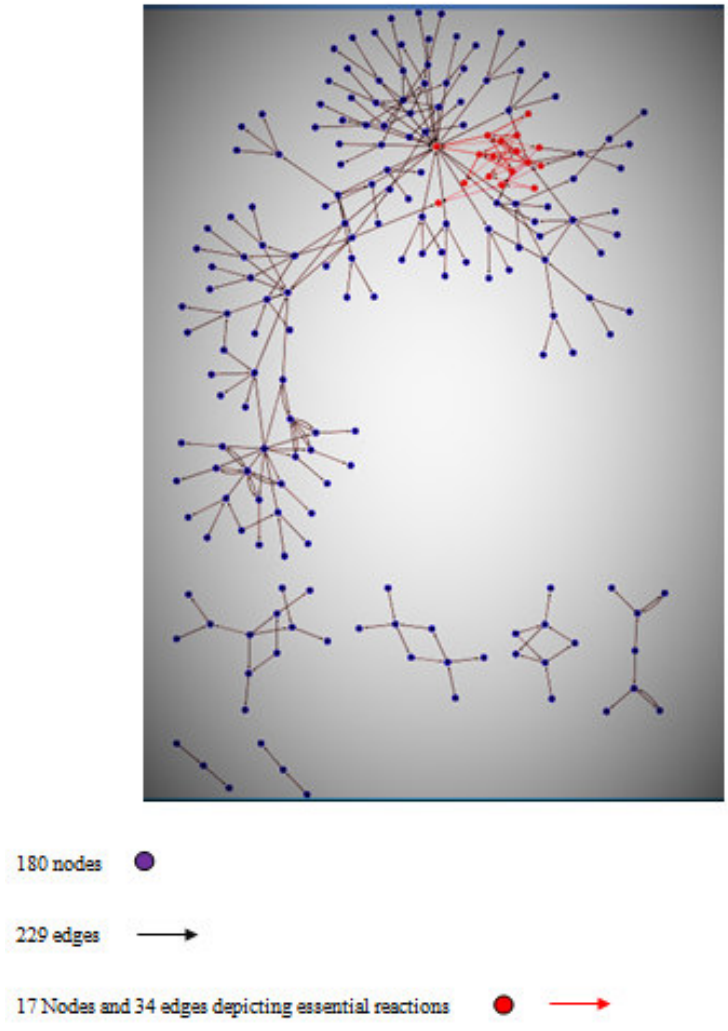


Table 2
Reactions obtained after performing constrained base analysis of reconstructed metabolic network

Reactions obtained after performing FBA	Reactions obtained after performing FVA	Live reactions	Essential reactions
R05884	R05884	R05884	R05884
R07219	R07219	R07219	R07219
R08209	R08209	R08209	R08209
R08210	R08210	R08210	R08210
	R03191	R03191	R03191
	R03192	R03192	R03192

Table 3
Essential reactions present in metabolic network

Essential reaction ID	DEG ID	KO Number	Protein name	Gene name	EC	Reaction
R03191	DEG10310215	K00075	UDP-N-acetylmuramate dehydrogenase	murB	1.3.1.98	UDP-N-acetylmuramate + NAD+ <=> UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine + NADH + H+
R03192	DEG10310215	K00075	UDP-N-acetylmuramate dehydrogenase	murB	1.3.1.98	UDP-N-acetylmuramate + NADP+ <=> UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine + NADPH + H+
R05884	DEG10310114	K03527	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	ispH, lytB	1.17.1.2	1-Hydroxy-2-methyl-2-butenyl 4-diphosphate + NADPH + H+ <=> Isopentenyl diphosphate + NADP+ + H2O
R07219	DEG10310114	K03527	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	ispH, lytB	1.17.1.2	Dimethylallyl diphosphate + NADP+ + H2O <=> 1-Hydroxy-2-methyl-2-butenyl 4-diphosphate + NADPH + H+
R08209	DEG10310114	K03527	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	ispH, lytB	1.17.1.2	Isopentenyl diphosphate + NAD+ + H2O <=> 1-Hydroxy-2-methyl-2-butenyl 4-diphosphate + NADH + H+
R08210	DEG10310114	K03527	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	ispH, lytB	1.17.1.2	Dimethylallyl diphosphate + NAD+ + H2O <=> 1-Hydroxy-2-methyl-2 butenyl 4-diphosphate + NADH + H+

CONCLUSION

Recently the practice of amalgamation of proteomics and bioinformatics data has become a promising mechanism to discover novel drug targets. Database of Essential Genes (DEG) consists of genes that are regarded to form the functional basis of life. In the present study the incorporation of *In-silico* subtractive proteomics approach and metabolic network reconstruction has been done to find putative drug targets in *Campylobacter jejuni* to improve the future treatment regime. 6 essential reactions of murB, ispH and lytB genes were deferred that can be used for molecular modelling, virtual screening and in experimental studies so as to discover potential therapeutic compounds against *Campylobacter jejuni* in future

REFERENCES

1. Nyati K, Nyati R. Role of *Campylobacter jejuni* Infection in the Pathogenesis of Guillain-Barré Syndrome: An Update. Biomed Res Int. 2013;2013:1-14.
2. Yuki N, Hartung H. Guillain–Barré Syndrome. New Engl J Med. 2012;366:2294-2304.
3. Fokke C, Berg B, Drenthen J, Walgaard C, Doorn P, Jacobs B. Diagnosis of Guillain-Barré syndrome and validation of Brighton Criteria. Brain. 2013;137:33-43.
4. Pope J, Krizova A, Garg A. *Campylobacter* Reactive Arthritis: A Systematic Review. Semin Arthritis Rheum. 2010;37:48-55.
5. Menna-Barreto RF, Belloze KT, Perales J, Silva-Jr FP. Proteomic and bioinformatic analysis of *Trypanosoma cruzi* chemotherapy and potential drug targets: new pieces for an old puzzle. Curr Drug Targets. 2014;15:255-71.
6. Shinde P, Patil K, Malavkar N, Dalvi A, Janve M *In Silico* Subtractive Proteomics Approach for Identification and Characterization of Drug Targets in *Streptococcus suis*. J Comput Biol. 2013;2:1-10.
7. Thiele I, Heinken A, Fleming R. A systems biology approach to studying the role of microbes in human health. Curr Opin Biotechnol. 2012;24:1-9.
8. Thiele I, Swainston N, Fleming R, Hoppe A, Sahoo S, Aurich S, et al. A community-driven global reconstruction of human metabolism. Nat. Biotechnol. 2013;31:415-427.
9. Orth J, Thiele I, Palsson B. What is flux balance analysis? Nat. Biotechnol.2010;28: 245–248.
10. Li Z, Wang R, Zhang X. Two-stage flux balance analysis of metabolic networks for drug target identification. BMC Syst Biol.2011; 5:S11.
11. Damiani A, He Q, Jeffries T, Wang J. Comprehensive evaluation of two genome-scale metabolic network models for *Scheffersomyces stipites.* Biotechnol. Bioeng.2015;112:1250-1262.
12. Babaei P, Marashi S, Asad S. Genome-scale reconstruction of the metabolic network in *Pseudomonas stutzeri A1501.* Mol. BioSyst. 2015;11:3022-3032.
13. Metris A, Reuter M, Gaskin DJ, Baranyi J, Vliet A. *In vivo* and *in silico* determination of essential genes of *Campylobacter jejuni.* BMC Genomics. 2011;12:535.
14. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 2010;26:680-682.
15. Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35:182-185.
16. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42:199-205.
17. Paramasivam N, Linke D. ClubSub-P: cluster-based subcellular localization prediction for

Gram-negative bacteria and archaea. Front Microbiol. 2011;2:1-14.
18. Forth T, McConkey G, Westhead D. MetNetMaker: a free and open-source tool for the creation of novel metabolic networks in SBML format. Bioinformatics.2010;26:2352–2353.
19. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, et al. Integration of biological networks and gene expression data using Cytoscape. Nature protocols.2007;2: 2366-2382.
20. Gevorgyan A, Bushell ME, Avignone-Rossa C, Kierzek A. SurreyFBA: a command line tool and graphics user interface for constraint-based modeling of genome-scale metabolic reaction networks. Bioinformatics.2011;27: 433-434.
21. Zhang R, Ou H, Zhang C. DEG: a database of essential genes. Nucleic Acids Res. 2004;32:271-272.
22. Francke C, Siezen RJ and Teusink B. Reconstructing the metabolic network of a bacterium from its genome. Trends Microbiol. 2005;13:550–558.
23. Kauffman K, P Prakash P and Edwards J. Advances in flux balance analysis. Curr Opin Biotech.2003;14:491–496.
24. Müller A and Bockmayr A. Fast thermodynamically constrained flux variability analysis. Bioinformatics.2013;29:903-909.
25. Babajan B, Anuradha CM, Chaitanya M, Gowsia D, Chitta S. *In silico* structural characterization of *Mycobacterium tuberculosis H37Rv* UDP-N-acetylmuramate dehydrogenase. Int J Integr Biol.2009;6:12-16.
26. Rekittke I, Olkhova E, Wiesner J, Demmer U, Warentin E, Jomaa H, Ermler U. Structure of the (E)-4-hydroxy-3-methyl-but-2-enyl- diphosphate reductase from *Plasmodium falciparum.* FEBS Letters.2013;587:3968–3972.
27. Raman K, Rajagopalan P and Chandra N. Flux Balance Analysis of Mycolic Acid Pathway: Targets for Anti-Tubercular Drugs. PLoS Comput Biol.2005;1:0349-0358.
28. Suguna L and Chitra P. Exploring biological networks for finding potential drug targets in methicillin resistant *Staphylococcus Aureus.* Int. J. Pharm. Biol.2011;2:544-551.