ISSN:0975-1459
Journal of Pharmaceutical Sciences and Research
www.jpsr.pharmainfo.in

Elucidating Vital Drug Targets of *Salmonella enterica* utilizing the Bioinformatic Approach

Anukriti Verma, Nikita Goel**, Vijaya Laxmi**, Bhawna Rathi*, Shivani Sharda

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

* brathi@amity.edu

** Authors contributed equally to this work

Abstract

Salmonella enterica, a rod shaped gram- negative bacterium is a major cause of salmonellosis worldwide inducing enteritis and typhoid fever. It is necessary to seek out its essential drug targets as a few strains have developed drug resistance to typhoid fever. The massive quantity of genomic/proteomic knowledge together with systems biology approach can result in the identification of potential drug targets. The combination of subtractive proteomics approach and metabolic network modelling was used for locating novel targets in *Salmonella enterica*. Our study revealed that this pathogenic bacteria consists of 1399 essential genes out of which 649 non- paralogs and 378 non- homologs are present. The KO numbers of 301 essential genes were obtained consisting of 143 genes related to metabolism. The metabolic network of *Salmonella enterica* was constructed and simulation studies were performed that disclosed 10 essential reactions as well as proteins UDP-N-acetylmuramate dehydrogenase, CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase, CDP-4-dehydro-6-deoxyglucose reductase, 4-hydroxy-tetrahydrodipicolinate reductase, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase encoded by murB, rfbH, ascD, ddhD, rfbI, dapB, ispH and lytB genes that can be considered as imminent drug targets. The structures of CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase, CDP-4-dehydro-6-deoxyglucose reductase and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase were detected and verified. The list of top 10 inhibitors were obtained that can be considered as potential drug targets in future.

Keywords

Salmonella enterica, Subtractive proteomics approach, Metabolic network modelling, Simulation studies, Docking.

INTRODUCTION

Salmonellosis is a foodborne sickness brought on by Salmonella (gram-negative microorganisms) prompting gastroenteritis and in a couple cases even typhoid fever that can be deadly moreover [1]. In 2000, just about 21.6 million individuals around the globe developed typhoid fever leading to 216,500 deaths [2]. Salmonellosis is caused by *S. enteritidis (16.9%), S.typhimurium (16%), S. enteritidis heidelberg (3.9%), S. enteritidis newport (10.4%), S. enterica serotype Javiana (5.5%), S. enterica serotype I 4,5,12:i: (5.7%) and S. enteritidis montevideo (3.4%)* [3]. The symptoms include stomach spasms, bloody stools, chills, diarrhea and even urolithiasis. Salmonellosis has often led to reactive arthritis [4]. Generally few salmonella strains have created drug resistance to typhoid fever [5]. Subsequently to produce enhanced indicative devices and therapeutics it is important to identify novel effective drug targets. Recently the utilization of proteomics and bioinformatics information has turned into a potential mechanism to discover new drug targets [6, 7, 8, 9, 10]. The present study combines the use of subtractive proteomics methodology, metabolic network reconstruction, simulation, structure identification and docking studies to decipher promising drug targets and identify inhibitors of *Salmonella enterica* in order to enhance the diagnosis criteria and treatment administration in future.

MATERIALS AND METHODS

The methodology to identify drug targets of *Salmonella enterica* is portrayed in Fig. 1.

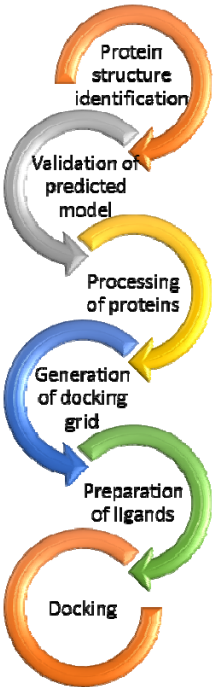


Fig. 1. *In-silico* analysis of *Salmonella enterica* to infer potential drug targets

Subtractive proteomics approach

Essential protein sequences of *Salmonella enterica* were recovered from Database of Essential Genes (DEG) (http://tubic.tju.edu.cn/deg) [11] that incorporates outstanding targets needed for antibacterial medications. CD-HIT Suite at 60% identity level (http://weizhong-lab.ucsd.edu/cdhit_suite/) [12] was utilized for clustering of protein sequences of *Salmonella enterica*. The non-paralogous sequences were subjected to BLAST-P at E-value cutoff of 10^{-4} in order to deduce homologous successions against human database. KAAS (KEGG Automatic Annotation Server) (http://www.genome.jp/tools/kaas/) was utilized that executed BLAST of non-homologous protein arrangements against database of *Salmonella enterica subsp. enterica serovar Typhi CT18, Salmonella enterica subsp. enterica serovar Typhi Ty2, Salmonella enterica subsp. enterica serovar Typhi Ty21a, Salmonella enterica subsp. enterica serovar Typhimurium LT2, Salmonella enterica subsp. enterica serovar Typhimurium U288, Salmonella enterica subsp. enterica serovar Paratyphi A ATCC9150, Salmonella enterica subsp. enterica serovar Paratyphi A AKU12601, Salmonella enterica subsp. enterica serovar Paratyphi B, Salmonella enterica subsp. enterica serovar Paratyphi, Salmonella enterica subsp. enterica serovar Choleraesuis, Salmonella enterica subsp. enterica serovar Heidelberg SL476, Salmonella enterica subsp. enterica serovar Newport SL254, Salmonella enterica subsp. enterica serovar Schwarzengrund, Salmonella enterica subsp. enterica serovar Agona SL483, Salmonella enterica subsp. enterica serovar Dublin, Salmonella enterica subsp. enterica serovar Gallinarum 287/91, Salmonella enterica subsp. enterica serovar Enteritidis P125109, Salmonella enterica subsp. enterica serovar Javiana, Salmonella enterica subsp. arizonae* to naturally assign KEGG orthology (KO) identifiers [13].

Metabolic network reconstruction

The draft reconstruct of the genes involved in metabolism was made using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database [14, 15]. The data in regards to protein name, gene name, EC number, reaction ID and pathways was acquired. There may be few erroneously fused data or other missing connections present, so the entire reproduced draft was re-assessed and refined utilizing different databases. Metnetmaker software was used to make a metabolic reconstruct of essential genes included in metabolic pathways. It appointed flux to the reactions present inside the reconstruct [16] and was visualized with the help of cytoscape [17]. Surrey FBA was applied for constraint based displaying of FBA prepared metabolic model [18] that gives data with respect to the quantity of metabolites and number of reactions present in the system and performs flux balance analysis (FBA), flux variability analysis (FVA), identifies live reactions, connected components and essential reactions present in the network.

Structure identification and docking studies

CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase, CDP-4-dehydro-6-deoxyglucose reductase and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase encoded by rfbH, ascD, ddhD, rfbI, ispH and lytB genes are found to be essential after performing the simulation studies on the metabolic network of *Salmonella enterica*. The structure prediction of these proteins was done using Schrödinger. The validation of the predicted structures were done using PROCHECK and these were processed using the Protein Preparation Wizard (PrepWizard) of the Maestro Schrödinger suite to convert the the raw state protein by creating disulfide bonds, adding hydrogen atoms calculating the protonation states of all ionizable groups, assigning bond orders and optimizing the orientation of hydroxyl groups, Asn, Gln and His residues [19]. The active sites of the proteins were predicted using SITEMAP that were further incorporated in the generation of a docking grid using the Receptor Grid Generation. Ligand preparation was done using the Ligprep. Docking studies were performed using the ZINC Drug Database that were exposed to Glide [20]. Probable ligands or protein inhibitors were identified out of which the top 10 inhibitors were selected according to the docking score.

RESULTS AND DISCUSSION

Subtractive proteomics approach

1399 essential genes of *Salmonella enterica* were retrieved from database of essential genes (DEG) that contains the genes that form a fundamental part to convey and support the cellular activities [21] CD-HIT suite resulted in 649 clusters (non-paralogs) [12] that were subjected to BLAST-P revealing 378 non-homologous sequences against the human protein database. The KO numbers of 301 essential genes of *Salmonella enterica* were procured utilizing KAAS that specifically interfaces these KO numbers with the KEGG orthology database that associates the genomic data to the functional data through KEGG PATHWAY mapping and BRITE mapping [13] (Table a). KEGG database uncovered the genes included in different pathways. There are 143 genes included in metabolic pathways 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 biosynthesis of other secondary metabolites. 13 genes did not have/ had partial EC numbers and were not considered for further investigation.

Metabolic network reconstruction

Metabolic network reconstruction of the rest essential genes involved in metabolic pathways was done to evaluate the metabolic pathways to fathom the relationship between genome and physiology of an organism [22]. An FBA prepared model by appointing maximum positive and negative fluxes and choosing all reactions as objective functions was acquired using metnetmaker software that was envisioned utilizing cytoscape [17] (Fig. 2 and 3). It uncovered 10 live reactions that carry a steady state flux.

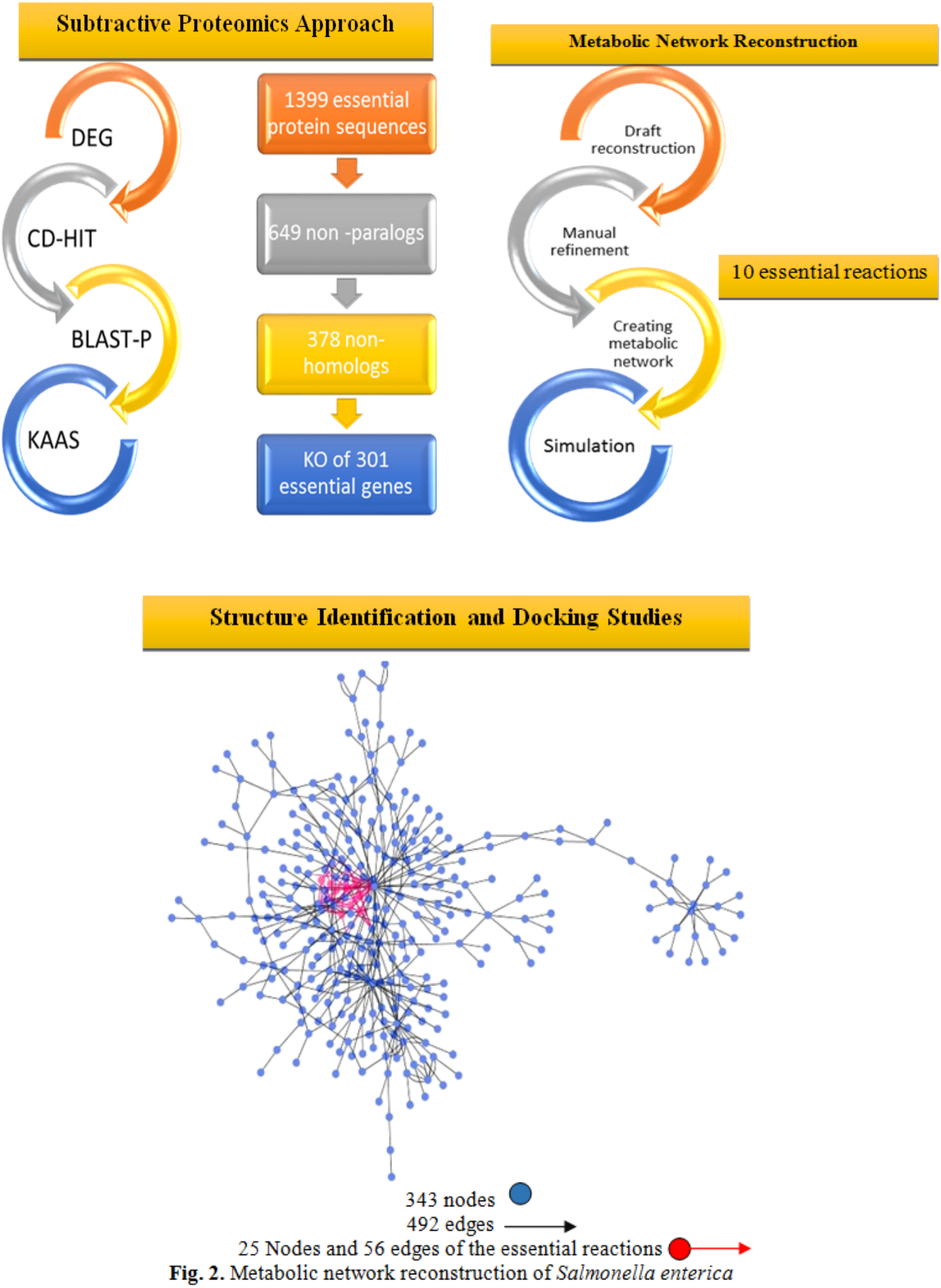


Fig. 2. Metabolic network reconstruction of *Salmonella enterica*

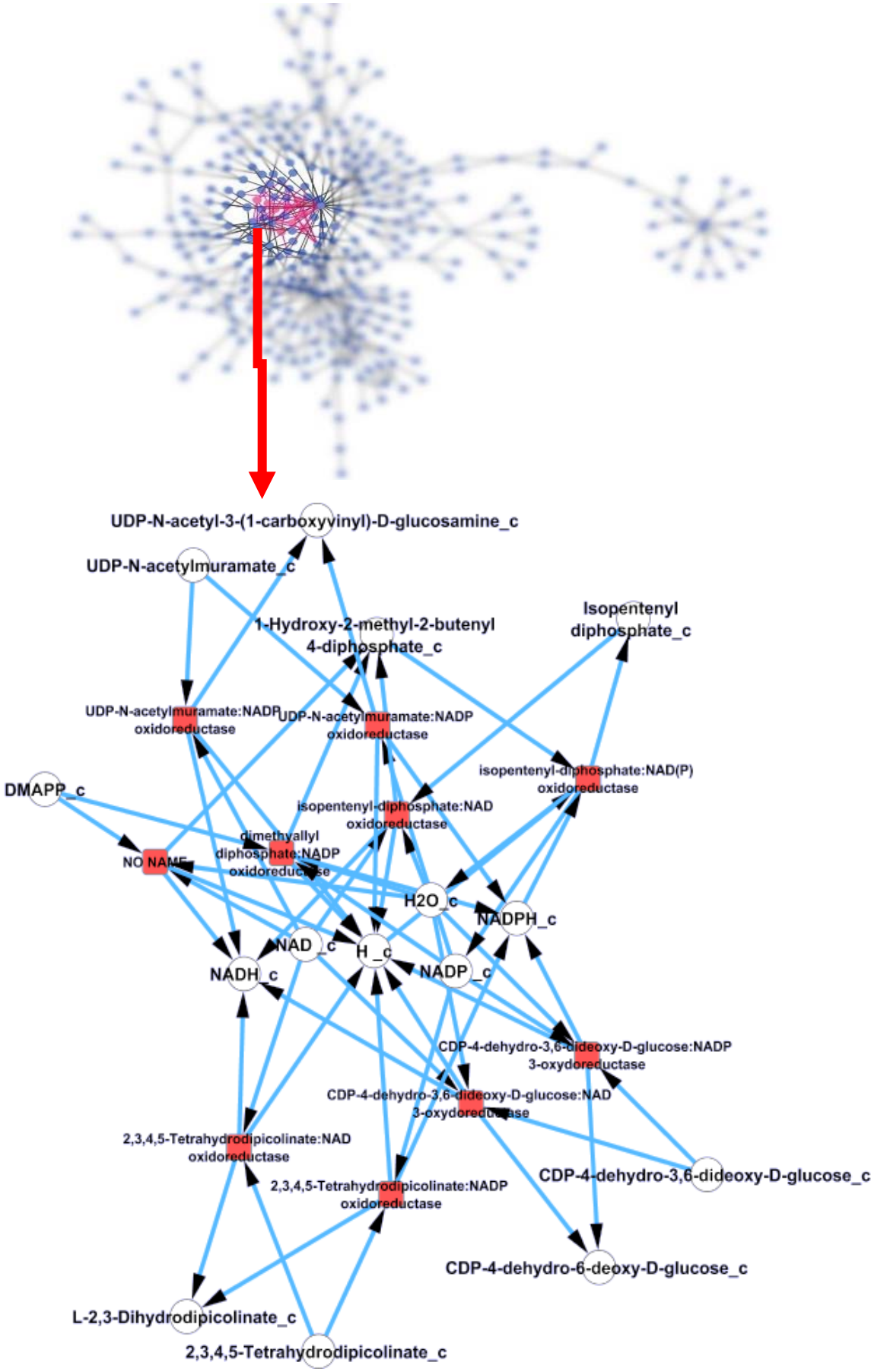


Fig. 3. Metabolic network reconstruction of essential reactions of *Salmonella enterica*

Simulation of metabolic network

SurreyFBA was utilized 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.

A. FBA (Flux Balance Analysis)

FBA was executed that is used to compute the metabolic course through the metabolic system. It signifies the metabolic reactions as a stoichiometric matrix with m compounds and n reactions. It helps in implementing the mass equilibrium studies at steady state to maximize or minimize an objective function, comprising of predefined blend of direct fluxes, using linear programming and providing specific steady-state flux distribution as an output. 8 reactions were obtained through FBA analysis carrying a steady state flux that verifies that the flux that creates a metabolite equivalents to flux that expends a metabolite. Such responses can be utilized further for essentiality or deletion studies to render the metabolic effectiveness of the network and recognize drug targets [23]. Its detriment is that it creates a solitary flux dissemination to accomplish the optimal objective function.

B. FVA (Flux variability analysis)

There may be more answers for such issues. These optional arrangements can be given by FVA (Flux variability analysis) that was carried out to locate the minimum and maximum values of the fluxes of the reactions, without debilitating the optimal objective function value [24]. This created 10 reactions out of which 8 concur with the FBA outcomes (Table b). The responses are viewed as vital if their fluxes are made invalid (0) and this thus makes the objective value null (for this situation if the essential reactions are inactivated then they will inactivate other reactions also).

Essential reactions

The examination of this network revealed 10 reactions that were coinciding with FVA results to be essential that comprise eight genes namely murB encoding UDP-N acetylmuramate dehydrogenase having EC number 1.3.1.98 involved in carbohydrate metabolism and glycan biosynthesis and metabolism; rfbH, ascD, ddhD and rfbI encoding CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase having EC number 1.17.1.1 involved in carbohydrate metabolism; dapB encoding 4-hydroxy-tetrahydrodipicolinate reductase having EC number 1.17.1.8 involved in global and overview maps, amino acid metabolism and biosynthesis of other secondary metabolites and ispH, 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 c).

Structure identification and docking studies

The structures of CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase, CDP-4-dehydro-6-deoxyglucose reductase and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase were predicted using Schrödinger (Table d). CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase, CDP-4-dehydro-6-deoxyglucose reductase and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase resulted in 91.5%, 91.5% and 94.5 % structure validation respectively using PROCHECK. Protein Preparation Wizard (PrepWizard) of the Maestro Schrödinger suite assigned the bond orders to the workspace and skipped the residues with existing double and triple bonds for preparation of protein. Hydrogens were added and metals were treated. Disulfide bonds were created and waters were deleted farther than 5 angstroms from hets. Optimization of H-bonds was performed. RMSD: 0.3 was used for restrained minimization. The active sites of the proteins were obtained. Zinc Drug Database (Zdd) contain commercially accessible pure compounds that are approved to be used as drugs. The size of this database is 3001. The ligands were prepared for molecular docking against the Zdd. This database generated 6289 ligands. For CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase 6001 inhibitors were obtained, for CDP-4-dehydro-6-deoxyglucose reductase 5922 inhibitors and for 4-hydroxy-3-methylbut-2-enyl diphosphate reductase 5375 inhibitors after performing docking. The list of top 10 inhibitors was obtained according to the docking score (Table e). These can be used for pharmacophore modelling or ADME/Tox property identification in future [25].

Protein sequences in Database of essential genes (DEG)	1399
Non- paralogous sequences	649
Paralogous sequences	750
Non-homologous sequences	378
Homologous sequences	271
Sequences having KO identifiers	301

Table a. Subtractive proteomics approach used for *Salmonella enterica*

Reactions obtained after performing FBA	Reactions obtained after performing FVA	Live reactions	Essential reactions
R03391	R03191	R03191	R03191
R03392	R03192	R03192	R03192
R04198	R03391	R03391	R03391
R04199	R03392	R03392	R03392
R05884	R04198	R04198	R04198
R07219	R04199	R04199	R04199
R08209	R05884	R05884	R05884
R08210	R07219	R07219	R07219
	R08209	R08209	R08209
	R08210	R08210	R08210

Table b. Significant reactions involved in the metabolic network of *Salmonella enterica*

Essential reaction ID	DEG ID	KO Number	Protein name	Gene name	EC	Reaction
R03191	DEG10330269	K00075	UDP-N-acetylmuramate dehydrogenase	murB	1.3.1.98	UDP-N-acetylmuramate + NAD+ <=> UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine +
R03192	DEG10330269	K00075	UDP-N-acetylmuramate dehydrogenase	murB	1.3.1.98	UDP-N-acetylmuramate + NADP+ <=> UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine + NADPH + H+
R03391	DEG10260042	K12452	CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase	rfbH	1.17.1.1	CDP-4-dehydro-3,6-dideoxy-D-glucose + NAD+ + H2O <=> CDP-4-dehydro-6-deoxy-D-glucose + NADH + H+
	DEG10260043	K00523	CDP-4-dehydro-6-deoxyglucose reductase	ascD, ddhD, rfbI	1.17.1.1	CDP-4-dehydro-3,6-dideoxy-D-glucose + NAD+ + H2O <=> CDP-4-dehydro-6-deoxy-D-glucose + NADH + H+
R03392	DEG10260042	K12452	CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase	rfbH	1.17.1.1	CDP-4-dehydro-3,6-dideoxy-D-glucose + NAD+ + H2O <=> CDP-4-dehydro-6-deoxy-D-glucose + NADH + H+
	DEG10260043	K00523	CDP-4-dehydro-6-deoxyglucose reductase	ascD, ddhD, rfbI	1.17.1.1	CDP-4-dehydro-3,6-dideoxy-D-glucose + NAD+ + H2O <=> CDP-4-dehydro-6-deoxy-D-glucose + NADH + H+
R04198	DEG10330006	K00215	4-hydroxy-tetrahydrodipicolinate reductase	dapB	1.17.1.8	2,3,4,5-Tetrahydrodipicolinate + NAD+ + H2O <=> (2S,4S)-4-Hydroxy-2,3,4,5-tetrahydrodipicolinate + NADH + H+
R04199	DEG10330006	K00215	4-hydroxy-tetrahydrodipicolinate reductase	dapB	1.17.1.8	2,3,4,5-Tetrahydrodipicolinate + NADP+ + H2O <=> (2S,4S)-4-Hydroxy-2,3,4,5-tetrahydrodipicolinate + NADPH + H+
R05884	DEG10330005	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	DEG10330005	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	DEG10330005	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	DEG10330005	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+

Table c. Essential reactions of *Salmonella enterica* obtained after simulation studies

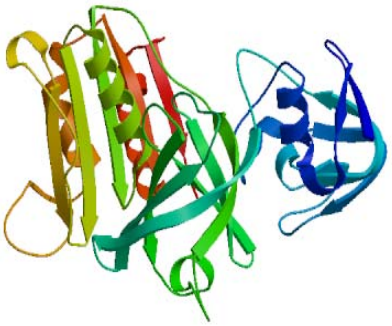
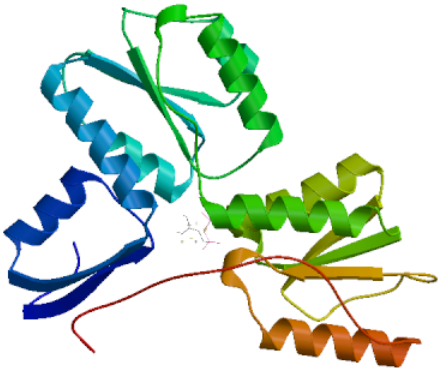
Protein name	PROCHECK score	Structure predicted
CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase	91.5%	
CDP-4-dehydro-6-deoxyglucose reductase	91.5%	
4-hydroxy-3-methylbut-2-enyl diphosphate	94.5 %	

Table d. Structure prediction and verification of CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase, CDP-4-dehydro-6-deoxyglucose reductase and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase of *Salmonella enterica*

Protein name	Total Inhibitors	Top 10 inhibitors
CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase	6001	ZINC06507052 ZINC12503116 ZINC03794794 ZINC00020252 ZINC00057358 ZINC20444132 ZINC03830609 ZINC00057147 ZINC03830609 ZINC00057347
CDP-4-dehydro-6-deoxyglucose reductase	5922	ZINC03831425 ZINC08551105 ZINC08551107 ZINC03831427 ZINC03831426 ZINC03831428 ZINC03831426 ZINC03201893 ZINC08551106 ZINC08551106
4-hydroxy-3-methylbut-2-enyl diphosphate	5375	ZINC03831235 ZINC14879972 ZINC01612996 ZINC15668997 ZINC13523524 ZINC14879972 ZINC03874185 ZINC12504524 ZINC00643046 ZINC19702309

Table e. Top 10 hits (inhibitors) obtained after docking for CDP-6-deoxy-D-xylo-4-hexulose-3-dehydrase, CDP-4-dehydro-6-deoxyglucose reductase and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase of *Salmonella enterica*

CONCLUSION

In the current study the consolidation of subtractive proteomics methodology, metabolic network reconstruction, structural prediction and docking has been performed to discover possible drug targets in *Salmonella enterica* to enhance the future treatment administration. As of late the act of incorporation of proteomics and bioinformatics information has turned into a promising tool to find novel drug targets. Database of Essential Genes (DEG) that comprises of the genes that are respected to shape useful premise of life was used. 10 key reactions of murB, rfbH, ascD, ddhD, rfbI, dapB, ispH, lytB genes and their inhibitors were conceded that can be utilized for the identification of the ADME/Tox properties, pharmacophore modelling and in experimental analysis so as to find potential drug targets of *Salmonella enterica* in future.

REFERENCES

[1] Crum-Cianflone, N.F., *Curr Gastroenterol Rep.* 2008, 10, 424–431.
[2] Crump, J.A., Luby, S.P., Mintz, E.D., *Bull World Health Organ.* 2004, 82, 346-353.
[3] Murray, P.R., Baron, E.J., Jorgensen, J.H., Landry, M.L., Pfaller, M.A. (Eds.), *Manual of Clinical Microbiology* (9th ed.), ASM Press, Washington, D.C 2007.
[4] Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.A., Roy, S.L., Jones, J.L., Griffin, P.M., *Emerg Infect Dis.* 2011, 17, 7–15.
[5] CDC: Antibiotic Resistance Increasing in Certain Salmonella, Food Safety News, 09 Jun 2015.
[6] Li, Z., Wang, R., Zhang, X. *BMC Syst Biol.* 2011, 5, S11.
[7] Orth, J., Thiele, I., Palsson, B., *Nat. Biotechnol.* 2010, 28, 245–248.
[8] Thiele, I., Swainston, N., Fleming, R., et al., *Nat. Biotechnol.* 2013, 31, 415-427.
[9] Thiele, I., Heinken, A., Fleming, R., *Curr Opin Biotechnol*. 2012, 24, 1-9.
[10] Shinde, P., Patil, K., Malavkar, N., Dalvi, A., Janve, M., *J Comput Biol*. 2013, 2, 1-10.
[11] Hao, L., Yan, L., Feng, G., Chun-Ting Zhang, C., Zhang, R., *Nucleic Acids Res.* 2013, 1-7.
[12] Rubinstein, R., Ramagopal, U.A., Nathenson, S.G., Almo, S.C., Fiser, A., *Structure.* 2013, 21, 766–776.
[13] Chung, W.C., Chen, L.L., Lo, W.S., Kuo, P.A., Tu, J., Kuo, C.H., *Genome Announc.* 2013, 1, e00126-13.
[14] Kanehisa, M., Goto, S., Kawashima, S., Nakaya, A., *Nucleic Acids Res.* 2002, 30, 42-46.
[15] Wrzodek, C., Büchel, F., Ruff, M., Dräger, A., Zell, A., *BMC Systems Biology*. 2013, 7, 15.
[16] Forth, T., McConkey, G., Westhead, D., *Bioinformatics*. 2010, 26, 2352–2353.
[17] Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J, Wang, P.L., Lotia, S., Pico, A.R., Bader, G.D., Ideker, T., *Nat Methods.* 2012, 9, 1069–1076.
[18] Gevorgyan, A., Bushell, M.E., Avignone-Rossa, C., Kierzek, A.M., *Bioinformatics.* 2011, 27, 433-434.
[19] Unger M, Eichhoff A, Schumacher L, Strysio M, Menzel S, Carsten S, Vanina A, Vanesa Z, Seman M, Brander J, Rohde H, Zhu K, Haag F, Mittrücker H, Goldbaum F, Aktories K, Koch-Nolte F.. *Sci Rep*. 2015, 5, 1-10.
[20] Tamilvanan T, Hopper W. *Bioinformation*. 2013, 9, 286-92.
[21] Zhang, R., Ou, H., Zhang, C. *Nucleic Acids Res.* 2004, 32, 271-272.
[22] Ma, H., Zeng, A. *Bioinformatics.* 2003, 19, 270-277.
[23] Kauffman, K., Prakash, P., Edwards, J., *Curr Opin Biotech.* 2003, 14, 491–496.
[24] Müller, A., Bockmayr, A., *Bioinformatics.* 2013, 29, 903-909.
[25] Vijayakumar B, Umamaheswari A, Puratchikody A, Velmurugan D., *Bioinformation*. 2011, 7, 134-41.