Research Article

Subtractive genomics approach in identifying polysacharide biosynthesis protein as novel drug target against *Eubacterium nodatum*

Shilpa S. Shiragannavar, Arun K. Shettar, Shivakumar B. Madagi, Sunanda Sarawad[1*]

Department of Studies in Bioinformatics and Biotechnology, Akkamahadevi Women's University Vijayapura-586108, Karnataka, India

Received: 5 November 2018 Revised: 28 November 2018 Accepted: 17 December 2018

Abstract

Objective: The present investigation was carried out to find the therapeutic drug targets against *Eubacterium nodatum*. **Materials and Methods:** Based on the CD-HITS 1627 proteins were selected among the total proteome count of 1690 proteins. BLASTP were used for sequence analysis which revealed around 1498 non homologues proteins in human genome. Database of Essential Gene (DEG) was used to study the high stringency analysis of the non homologues/remaining proteins which revealed around 807 essential proteins of *Eubacterium nodatum*. **Results:** Metabolic pathway analysis of human host and pathogen was performed by KEGG server. The KEGG results sorted around 132 proteins among selected 807 proteins. Further among the132 human non –homologues proteins 20 unique non homologues essential proteins were considered through Psortb, Cello and SOSUI which shows that, the identified drug targets are exposed to have high potential for designing novel drug against *Eubacterium nodatum*. The Jpred and Phyre2 server were used to identify 2D and 3D structures of 16 membrane associated proteins and validation was done with RAMPAGE and Procheck. Further study investigated around 200 selective drugs using the drug bank, PubChem databases which revealed Cinnamoyl echinadiol as a single potent drug according to the FDA approved drug bank database with diverse ADMET, virtual screening and drug-likeliness property. **Conclusion:** Molecular docking analysis shows that, Polysaccharide biosynthesis protein as a novel drug target with Cinnamoyl echinadiol drug against pathogen *Eubacterium nodatum*.

***Keywords*:** *Eubacterium nodatum,* BLASTP, DEG, KEGG, Cinnamoyl echinadiol, Polysacharide biosynthesis protein

Introduction

In silico subtractive genomic analysis is a very fast and efficient method for identifying proteins in pathogenic species that area sent in the host. These proteins could serve as potential drug targets against the pathogens infecting the host tissues. Another important condition is the essentiality of the pathogen-specific proteins. In silico subtractive genomics approach is a powerful-method to identify the specific genes, which are present in the pathogen but absent in the host. This helps in the identification of novel organism specific genes, which can be used as drug targets (Gupta et al., 2010). Essential proteins are those which are believed to be critical for the survival of a cell. Although the essentiality of a gene is dependent on specific environment and cellular conditions, in general, the essentiality of a protein target is a positive indicator for drug ability of the target. Recent progress in the field of computational biology and bioinformatics has generated various in silico analysis and drug designing approaches, eliminating the time and cost involved in the trial and error experiment at ions that go into drug development (Barh et al., 2011). These methods serve to shortlist the potential drug targets that will subsequently be used for laboratory testing. Subtractive genomics is one such insilico approach used for drug target identification based on determination of essential and nonhomologous proteins within the pathogenic organism (Barh et al., 2011; Hosen et al., 2014). Drug target identification is an important step involved in computer aided drug design process. The need for a rapid search for small molecules that binds to target of biological interest is of

Address for Corresponding Author:
Ms. Sunanda Sarawad
Faculty of Science
Dept of Bioinformatics & Biotechnology,
Akkamahadevi Women's University, Vijayapura-586108, Karnataka, India
Email: sunandasarwad25@gmail.com

DOI: https://doi.org/10.31024/ajpp.2019.5.2.24

vital importance in the drug discovery process. Macromolecular modeling by docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex (Munster et al., 2009; Fraser et al., 2009; Rout et al., 2000).

Till date the complete genome sequence of about 863 bacteria has been determined and 1653 bacterial genome projects are currently in progress. Availability of genome sequences of pathogens has provided a tremendous amount of information that can be useful in drug target and vaccine target identification. The differences in the proteins of the host and the pathogen can be effectively used for designing a drug specifically targeting the pathogen. This approach has been used successfully to identify essential genes in *Pseudomonas aeruginosa* (Sakharkar et al., 2004; Perumal et al., 2007).

Ideally, a drug target should be nonhomologous with host proteins as this would decrease the chances of non specific interactions with host proteins and associated side-effects. It is also advantageous if the target protein is known to be "essential" for bacterial survival; any disruption in the functioning of such a protein would lead to death of the bacterial cell (Chhabra et al., 2010). An additional resource that has aided the in silico identification of essential genes in pathogenic organisms is the Database of Essential Genes (DEG) (Zhang et al., 2004).This database contains records for all the essential genes that are currently known and the records are updated as new essential genes are identified and characterized.

In the past decade, complete genome sequences of several microbes are worked out (De Groot et al., 2002) and comparative genomics and subtractive genomics approaches have been used to retrieve valuable information for finding the treatment of various infections caused by the pathogens (Galperin & Koonin, 1999). The critical genes crucial for the survival of pathogen and absent in the host (Koonin et al., 1998) are also identified using the subtractive genomics approach. The chances of cross-reactivity and side-effects (Barh et al., 2011) are minimized by selecting such non-homologous proteins which are not present in humans. The genes and their products which can be used as potential drug targets are also identified by analyzing these genes with the KEGG pathway database (Moriya et al., 2007).

Infectious diseases are major dreadful threat to human life, spread through many causative agents like bacteria, fungus, virus, etc. Therefore, development of potential drug is highly essential to fight against infectious diseases. Despite rapid advances in the diagnosis of bacterial infections and the availability of effective antibiotics, meningococcal disease continues to represent a substantial public health problem for most countries (Tzeng et al., 2000).

Eubacterium nodatum is one of the predominant bacteria found in peri-implant sites (Tamura et al., 2013). *E. nodatum* is a species of *Eubacterium* that has commonly been found in patients with chronic periodontal disease. *E. nodatum* is obligately anaerobic, and a Gram-positive microbe (Holdeman et al., 1980). Although this bacterium was initially identified in the late 1980s, it has been difficult to understand its metabolic pathways since it is challenging to isolate, has a slow growth process, and is structurally complex (Uematsu & Hoshino.,1996; Noble et al., 2014). According to the Centers for Disease Control and Prevention (CDC), 47.2% of American adults suffer from periodontitis, a bacterial infection that forms on oral surfaces and damages the tissues that surround teeth in peri-implant patients (Eke et al., 2012). Because *E. nodatum* is a periodontal pathogen, research has been focused on learning about oral microbiota, and how changes in this microbiota affect and is affected by different diseases.

E. nodatum causes pelvic inflammatory disease when it colonizes the female genital tract. In the 1980s, before aseptic techniques were practiced, *E. nodatum* colonized the female genital tract through contaminated intrauterine devices for contraception, eventually causing pelvic inflammatory disease (Hill, 1992). *E. nodatum* is very similar to *Actinomyces israelii* in its morphology as well as its colonizing strategies (Hill, 1992). *A. israelii* commonly infects livestock with a disease called lumpy jaw, but rarely infects humans (Valour et al., 2014). *E. nodatum* infects humans through the same strategy causing severe periodontitis (Hill, 1992). It is generally unknown the pathologic biomechanisms used by *E. nodatum* in relation to other microorganisms in the biofilm due to their uncultivable nature (Ireland et al., 2013).

E. nodatum are branched, filamentous, Gram-positive bacteria that are non motile, and do not produce spores. When colonized, the rod-shaped bacteria clump together in broth cultures. When they form these clumps under favorable conditions, a biofilm is created. After incubation, cells form circular and raspberry-shaped colonies that are cream-colored (Hill et al., 1987).

There are a total of 1.83 million base pairs that make up the *E. nodatum* genome. *E. nodatum* has a total of 1661 genes and 1543 of which code for proteins involved in amino acid adenylation, antiporter transportation and proteins which play the role of kinases, transferases, polymerases, and transcription and translation factors. The genome has a guanine-cytosine percentage of 38.1. The genetic sequence of *E. nodatum* is 93.6% similar to *Eubacterium tardum*, a species in the same genius (Jousimies-Somer, 1997). In this

context, the present study is planned to identify a potential drug target against infectious diseases caused.

Materials and Methods

Subtractive genomics approach is implemented for the selection and characterization of potential drug targets for *Eubacterium nodatum* was done by using different step wise manner which is summarized in figure1.

Retrieval of whole Proteome of the *Eubacterium nodatum*

The complete proteome of *Eubacterium nodatum* strain was retrieved from the Uniprot database in the FASTA format www.uniprot.org (Magrane, 2011). The complete genome sequence data of *Eubacterium nodatum* was achieved with use of sequencing techniques.

Identification of essential proteins in *Eubacterium nodatum*

Identification of paralogous protein in *Eubacterium nodatum* proteome was eliminated at 60%using CD-HIT suite (Fu et al., 2012; Li ad Godzik, 2006). The paralogous were excluded and the remaining sets of proteins (non paralogous) were subjected to BlastP (http://www.ncbi.nim.nih.gov/blastp) against *Homo sapiens* protein sequences using threshold expectation value <=10^{-5}, sequence similarity >35%, bit score > 100 others as default parameters to find out the non-human homologous proteins of *Eubacterium nodatum (*Hosen et al., 2014). Blast P analysis was performed for the non homologous protein sequences of *Eubacterium nodatum* against *DEG* with random expectation value (E-value) $10^{-5.}$ A minimum bit score cut off 100 was used to screen out the proteins that appeared to represent essential proteins (Rathi et al., 2009; Zhang et al., 2008).

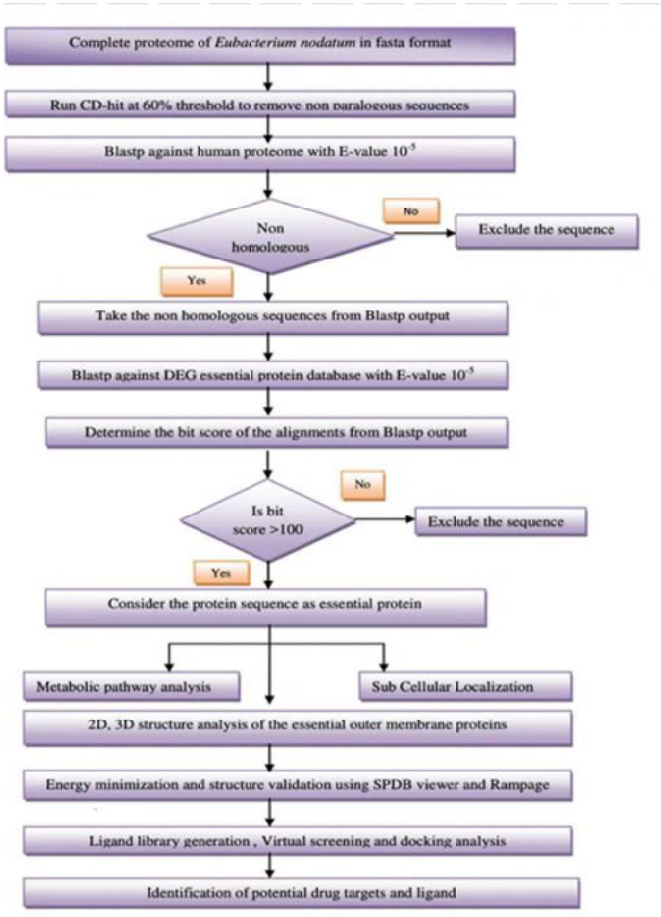

Figure 1. Schematic flow chart describing the detailed methodology for drug and putative drug targets identification

Analysis of metabolic pathway

Metabolic pathway analysis of the essential proteins of *Eubacterium nodatum* screened for the identification of the unique potential drug targets was performed using Kyoto Encyclopedia of Genes and Genomes (KEGG) Database (http://www.genome.jp/kegg/pathway.html) (Kanehisa et al., 2000).

Sub cellular location prediction of protein targets

PSORTb [www.psort.org/psortb/] was used for the Protein localization which is used to predict the protein function and genome annotation (Yu et al., 2014). SOUSI [harrier.nagahama-i-bio.ac.jp/sosui/sosui_submit.html] and CELLO [cello.life.nctu.edu.tw] were used to cross-check the data obtained from PSORTb and proteins were sorted according to their sub-cellular localization (Hossain et al., 2017; Yu et al., 2010).

2D and 3D model development

J-Pred4 server (http://www.compbio.dundee.ac.uk/jpred4) was used for prediction of secondary structure of the unique protein targets (Drozdetski et al., 2015). The modeling of three dimensional structure of identified target proteins was done using Phyre2 (Kelley et al., 2015) and CPH servers (http://www.cbs.dtu.dk/services/CPHmodels) (Nielsen et al., 2010). The models generated by both the servers were validated to check the structural quality and reliability and the best scoring models were used for further analysis.

3D model validation of target protein

The 3D models, structural quality, stereo chemical analysis, and reliability of refined models were evaluated using RAMPAGE and PROCHECK web tools (Monadal et al., 2015). Backbone conformation was evaluated by the inspection of the Psi/Phi Ramachandran plot obtained from PROCHECK analysis (Laskowski et al., 1993). Models constructed from the Pyre2 and CHP models were subjected to energy minimization (Kelley et al., 2015; Nielsen et al., 2010). The energetically stable structure was obtained through structural refinement and energy minimization using Swiss PDB Viewer (Ragno et al., 2016).

Building of compound library

A compound library of around 200 antibacterial

phytochemicals were collected and virtually generated for the study. The natural antibacterial compounds of different plant varieties were collected using literature survey and databases. The structures of the selected antibacterial chemical compounds were obtained from Pub Chem Database and the information of the compounds was retrieved from Drug bank Database (Wishart et al., 2017). The resultant sdf files of the compounds were converted to .mol files using Open babble (O'Boyle et al., 2011).

Virtual screening and Molecular docking studies for identification of novel drug against the bacterium

Virtual screening approach was used to identify the bactericidal potentiality of new herbal drug against the *Eubacterium nodatum*. The selected antibacterial compound should satisfy the Lipinski rule of five and drug likeliness to become potent drug (Lipinski et al., 1997, 2000). Over 50% of the compounds fail due to deficiencies of ADME/Tox during drug development. The *in- vitro* ADMET screenings have been implemented to avoid failure at the development stage with the aim of discarding compounds in the initial discovery phase (Tareq et al., 2010; Kŏster et al., 2011). The drug likeliness, toxicity and mutagenicity of the natural compounds were evaluated using preADMET server (Kovačević et al., 2014).

Preparation of the target proteins and ligand for docking analysis

The targets obtained from the subtractive genomics study were prepared for docking using SwissPDB viewer. To get the stable state of the target structures energy minimization was done prior to autodocking (Sastry et al., 2013). The ligands from .sdf files were converted to .pdb files for Autodock studies using Openbabble.

Molecular Docking studies using Autodock tool

All the five target protein structures were individually docked with the ligand using Autodock software. Many computational tools are available such as MetaPocket which was used for predicting binding sites for all the targets proteins. MetaPocket is available at http://metapocket.eml.org (Zhu et al., 2013). The interactions between the proteins and ligand were visualized using Molegro molecular viewer.

Results and discussion

Subtractive genome analysis is a strategy, which is used for the identification of a suitable drug targets among the essential proteins within the *Eubacterium nodatum.* Subtractive genomic has been reported as an innovative and power full method for identifying unique sequence as potential therapeutic targets (Hosen et al., 2014; Uddin et al., 2014).

Identification of non homologous essential proteins

The results obtained through computational analysis revealed that, the total proteome count of the *Eubacterium nodatum* was found to be around 1690. The non-redundant or duplicate sequences from the whole proteome were analyzed by the CD-HIT suite, which shows that 63 proteins were found to be paralogous or duplicate with 60% identity and 1627 proteins were non-paralogous (Wadood et al., 2018). Further these selected non-paralogous proteins were analyzed to check the homology against human using PDB to get the non-human homologous protein and the paralogous sequences were eliminated from the dataset. Among the tested sequences, around 123 sequences were found to be human homologous, that were excluded and 1498 non-human homologous sequences were taken for the further analysis. The non-human homologous sequences were focused to reduce the cross reactivity between human and pathogen (Rathi et al., 2009; Sakharkar et al., 2004). The essential genes important for the survival of bacteria were identified using DEG Database of Essential Genes for 1498 non-homologous protein sequences which show around 807 essential proteins (Zhang et al., 2004). The results are depicted in the table 1.

Table 1. Subtractive genomics and metabolic pathway analysis results for *Eubacterium nodatum*

Features of *Eubacterium nodatum*	Number
Total proteins	1690
CD-hits	1627
Blast P- Homologous	123
Blast P–Non homologous	1498
DEG	807
KEGG pathways	132
Sub cellular localization	20

Metabolic pathways analysis

Metabolic pathway construction of the essential proteins was done using KEGG for 807 proteins. Among the out of 807,132 protein sequences were involved in the metabolic pathway which is indispensible for the propagation of the pathogen. Each pathway was analyzed individually for their involvement in unique pathways and it was found that 20 proteins were present in the pathogen and termed as pathogen specific metabolic pathways and the remaining 112 pathogen pathways were defined as common host-pathogen metabolic pathways (Mondal et al., 2015).

The two component system of the bacteria represents the primary signal transduction in prokaryotic organisms, ABC3 transport family protein couple with ATP hydrolysis and active transport of a wide variety of substrates such as ions, sugars, lipids, sterols, peptides, proteins, and drugs. The beta-lactam proteins are the widely used group of

antibiotics, which exert their effect by interfering with the structural cross linking of peptidoglycans in bacterial cell walls. Peptidoglycan represents bacterial cell wall biosynthesis inhibitor. Mate domain protein represents major mechanism used by bacteria for uptake of carbohydrates. Transglycosylase forms the cell wall in bacteria surrounding the cytoplasmic membrane. PIN domain protein represents the correct processing, quality control & turnover of cellular RNA molecules which are critical to many aspects in the expression of genetic information. Polysaccharide biosynthesis protein is a regulatory system that allows bacteria to share information about cell density and adjust gene expression accordingly.

Mate efflux family protein enables bacteria to sense, respond, and adapt to changes in their environment or in their intracellular state. Peptidoglycans binding domain protein controls the auto-phosphorylation of the cheA histidine kinase, signal peptidase and transport newly synthesized protein into or across the cell membrane. Biotin metabolism is the essential cofactor of biotin-dependent carboxylases, such as pyruvate carboxylase and acetyl-CoA carboxylase. These essential proteins were selected to be potential drug targets based on the pathways (Kanehisa et al., 2009).

Sub cellular localization prediction

Localization of the Membrane proteins in the cell is an important factor for identification of suitable and effective drug targets. These cytoplasmic membrane bounded proteins are more favorable as drug targets and SOSUI server was used for localization predictions (Duffield et al., 2010). As the result of SOUSI 83 soluble proteins and 49 membrane proteins were identified. The 49 proteins were again cross checked for their reliability to be extracellular or membrane proteins using Psortb and CELLO. The comparative study was done to get the result accuracy. The membrane proteins were concentrated for further analysis as the membrane proteins have high therapeutic interventions (Rahmanet et al., 2014). Out of 132 proteins, Poly sacharide biosynthetic protein was selected based on the selective parameter.

2D and 3D structure prediction and validation

The secondary structure prediction was followed by tertiary structure prediction for Poly sacharide biosynthetic protein was doneusingJ-Pred4 and Phyre2 server. J-Pred4 features higher accuracy with a bind three-state (α-helix, β-strand and coil) secondary structure prediction accuracy of 82.0% (Drozdetski et al., 2015). The Phyre2 web portal is used for automated protein modeling, structure prediction and analysis. The server uses distantly related proteins as template for the structure prediction. Hence, the protein targets can be modeled with reasonable high accuracy (Kelley et al., 2009).

The structure validation was predicted by Phyre2 with the help of Procheck and RAMPAGE, a ramachandran plot analysis server (Zhou et al., 2011; Laskowski et al., 1993). RAMPAGE shows the most favored regions of 99.4% residues. The validated structures of the models were considered for the energy minimization using SwissPDB viewer to get better refined model & stable state of the compound (Guex et al., 1997).

Table 2. Protein Structure validation using Procheck & RAMPAGE

Uniprot ID	Target Proteins	Validation
W2BWJ1	Aminoacid carrier protein	95.5%
W2BWE5	Mate efflux protein	97%
W2BYV5	**Polysaccharide biosynthesis protein**	**97.1%**
W2C007	Abc3 transport family protein	94.9%
W2BZG5	Betalactamase	91.3%
W2BZD8	Mate domain protein	98.0%
W2BZJ6	Transglycosylase	91.7%
W2C1H9	Abc transporter ATP protein binding	94.3%
W2C1Q4	Cobalt transport protein	94.3%
W2BY52	Amino acid carrier	95.0%
W2BYC1	Domain protein	96.2%
W2BYC8	Signal peptidase	94.8%
W2BYV3	Peptidoglycon binding protein	92.4%
W2BXG7	Collagen adhesion family protein	93.9%
W2BXF2	Permease family protein	93.3%
W2BVZ4	Putative membrane protein	97.7%

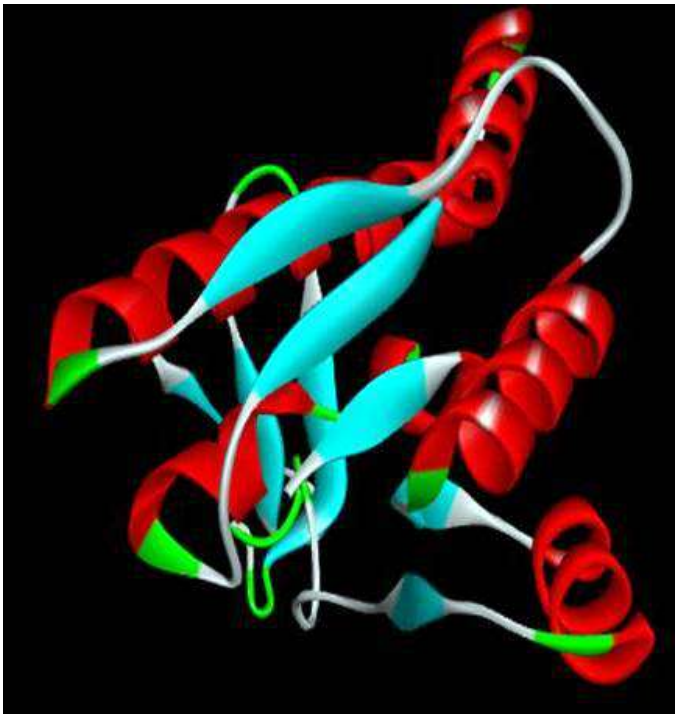

Figure 2. 3D structure of Polysaccharide biosynthesis protein

Building of compound library

A compound library of around200 antibacterial phytochemicals were collected and virtually generated for the study. The natural antibacterial compounds were collected from various plant varieties. The compounds were individually analyzed for Lipinski's Rule of five (Lipinski et al., 1997). The ligandcinn amoyl echinadiol was selected from the compound library based on the Lipinski rule of five with parameters such as the Molecular Weight is 384.516 g/mol, Hydrogen Bond Donor 4, Hydrogen Bond acceptor 1, X Logp3 value 4.6 and is non toxic and rodent non mutagenic (Petit et al., 2012). The compounds obeying Lipinski's rule of five were further screened using preADMET.

Virtual screening and Molecular docking studies

In this study, virtual screening approach was used to identify the bactericidal potentiality of new herbal drug against the *Eubacterium nodatum*. The drug likeliness, toxicity and mutagenicity of the natural compounds were screened using preADMET server (Köster et al., 2011; Kovačević et al., 2014). The compound Cinnamoyl echinadiol from *Echinacea purpurea* extract was selected for docking studies. The chemical structure of Cinnamoyl echinadiol was obtained from PubChem Database (https://pubchem.ncbi.nlm.nih.gov/compound/54608033). The results of preADMET and Drug likeliness predictions are given in the table 3.

Ames test, Carcino Rat and Carcino mouse are done for

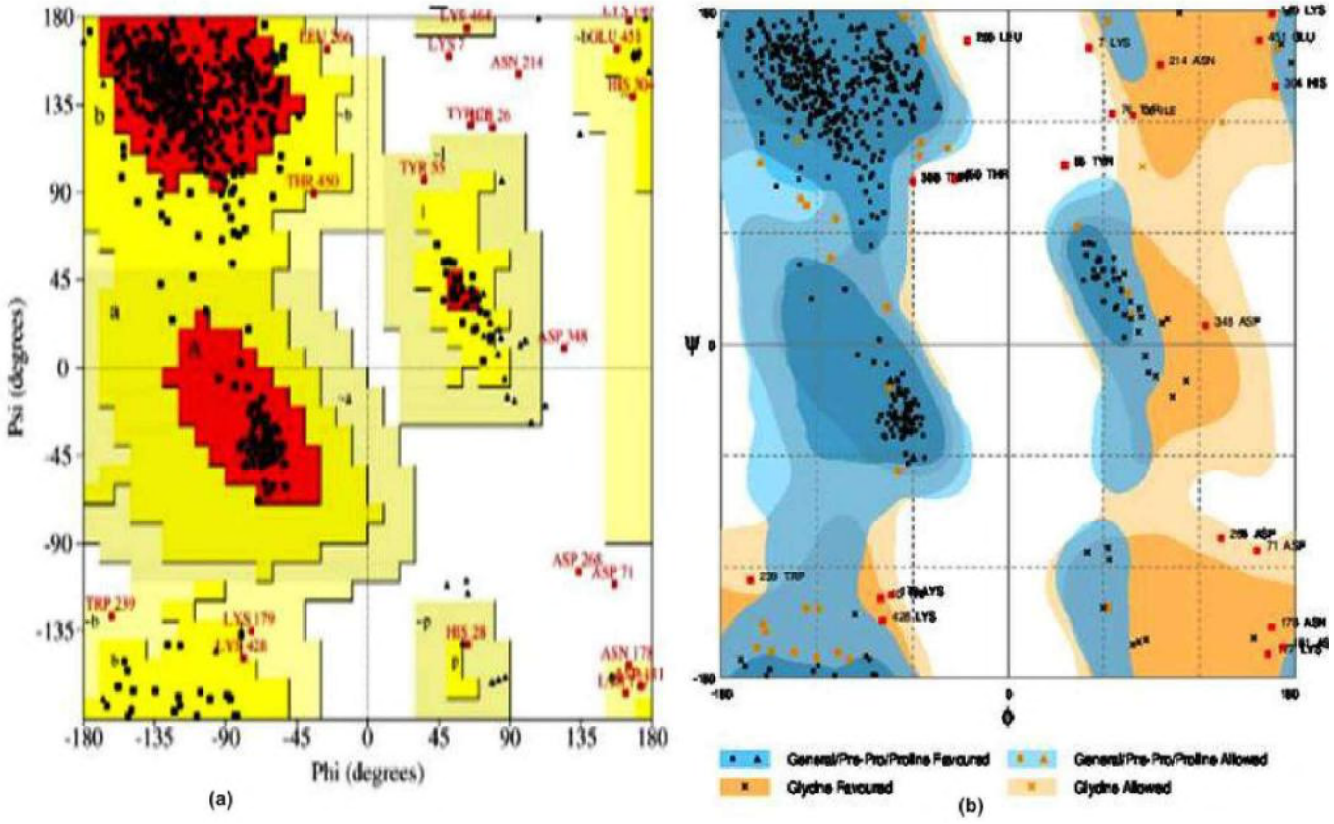

Figure 3. (a) PROCHEK and (b) RAMPAGE showing the amino acids residues in favored region of Polysaccharide biosynthesis protein

Table 3. Pre Admet toxicity and drug likeliness prediction

Predictions	Id	Value
Toxicity Prediction	Ames_Test	Non-Mutagen
	Carcino_Mouse	Negative
	Carcino_Rat	Negative
ADME Prediction	Blood brain barrier	1.85502
	Buffer_solubility_mg_L	16.8761
	Caco2	52.1139
	HIA	95.950489
	MDCK	0.246603
	Plasma_Protein_Binding	92.460532
	Skin_Permeability	-1.63615
Drug Likeliness	CMC_like_Rule	Qualified
	CMC_like_Rule_Violations	0
	Rule_of_Five	Suitable
	WDI_like_Rule	Out of 90% cutoff

checking the mutagenicity and carcinogenicity of the drug. As the rodent genome is related to human the carcinogenicity is validated using mouse models. MDCK (Madin Darby Canine Kidney) and

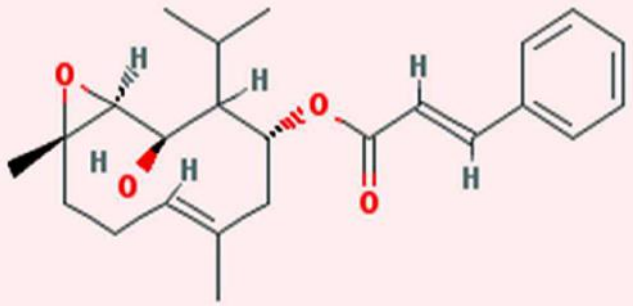

Figure 4. Structure of Cinnamoylechinadiol

CaCo2 tests are used to predict the rate drug absorption, while HIA (Human Intestinal Absorption) is done to check the potentiality drug for oral delivery and skin permeability. BBB (Blood Brain Barrier) on cell models are done to check the penetration of drug in central nervous system. Plasma protein binding test is done to check the disposition and efficacy of drug. CMC (Comprehensive Medicinal Chemistry database) test is done to know the effective range of physico chemical properties, lipophilicity and binding affinity. WDI (World Drug Index) computational approach to differentiate drug like and non-drug like molecules.

Table 4. Binding energies for ligand Cinnamoyl echinadiol with target proteins

Protein Targets	X	Y	Z	Binding energy for Cinnamoyl echinadiol
Aminoacid carrier protein	14.86	22.269	28.56	-8.44
Polysacharide biosynthesis protein	**-53.275**	**21.47**	**6.749**	**-11.71**
Transglycosylase protein	7.097	28.01	9.659	-7.06
ABC3 transport family protein	-8.511	16.742	26.638	-9.63
Cobalt transport protein	78.129	34.158	105.452	-9.24

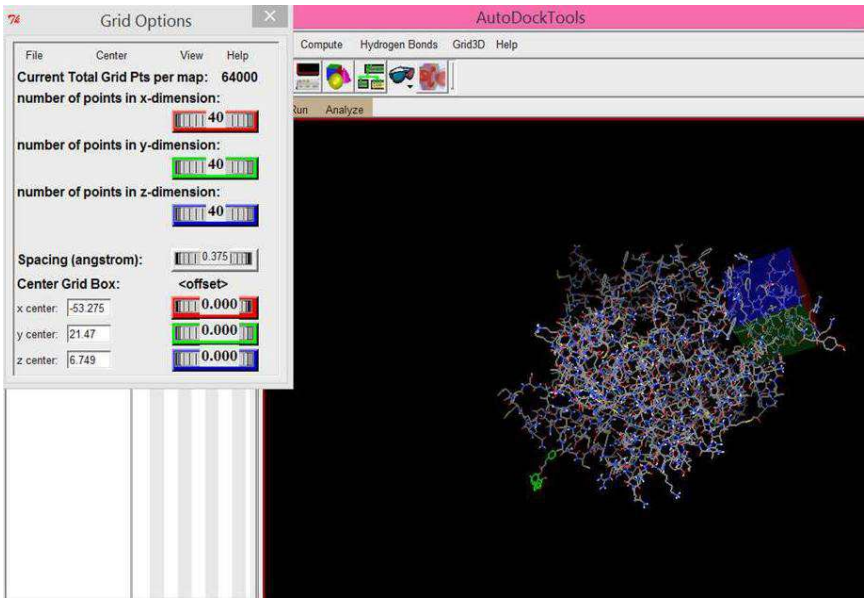

Figure 5. Molecular docking of Polysaccharide biosynthesis protein and Cinnamoyl echinadiol using Autodock

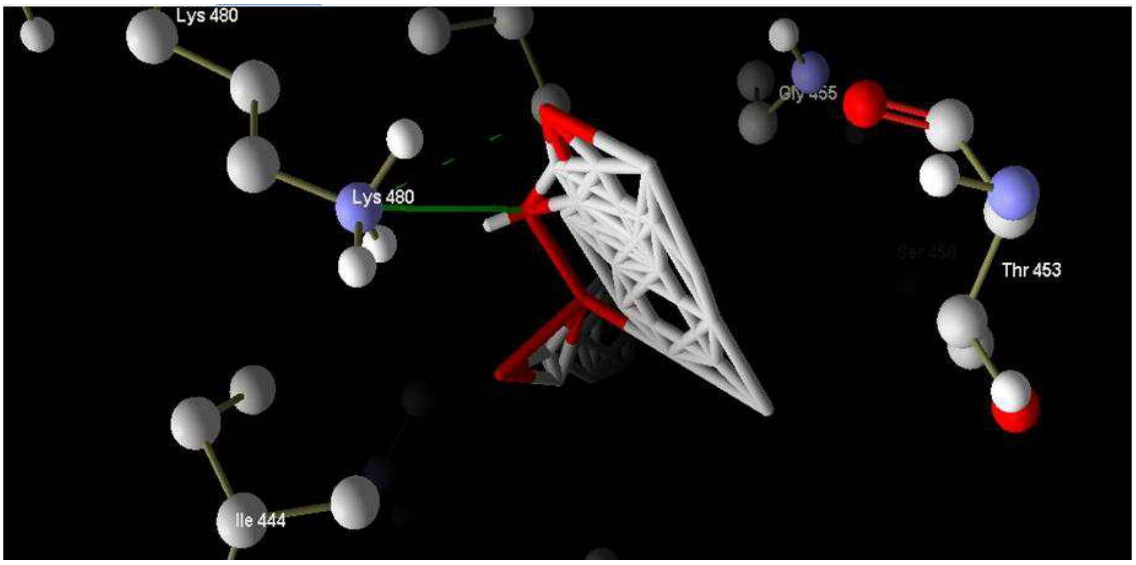

Figure 6. Hydrogen Bond interactions between Polysaccharide biosynthesis protein and Cinnamoyl echinadiol viewed by Molegro molecular viewer

The active binding site residues in the targets were predicted by using Metapocket (Huang et al., 2009). The XYZ co-ordinates were recorded for all the five targets. Molecular Docking of the Novel drug gable proteins with ligand for finding binding score & binding affinity was performed using Autodock. The XYZ - 53.275, 21.47, 6.749 co-ordinates were used for Autodock during Grid box generation. The binding energies for all the five target proteins with the ligand cinnamoyl echinadiol were tabulated. All the target proteins show good binding energy with the ligand. Polysaccharide biosynthesis protein has the highest binding energy with inhibition constant 2.62nM, hence the protein can be considered as potential drug target against *Eubacterium nodatum*.

Conclusion

The present work was carried out to find the natural/herbal drug against *Eubacterium nodatum*. In the present study subtractive genomics was used to identify the targeted protein and later the Virtual screening approach is used to identify new herbal drug for the pathogenic organism. Total of 1627 proteins were selected among the total proteome count of 1690 proteins based on CD hits. BLASTP revealed around 1498 non homologues proteins in human genome whereas essential proteins of *Eubacterium nodatum* were identified by Database of Essential Gene (DEG). Further the KEGG was used in the identification of selective proteins based on the unique pathway. Further the identified drug targets were exposed to have high potential for designing novel drug against *Eubacterium nodatum*. The overall subtractive genomics study concluded that poly sacharide biosynthetic protein found to be significant. The Jpred and Phyre2 server were used to identify 2D and 3D structures of poly sacharide biosynthetic protein and validation was done with RAMPAGE and Procheck. Virtual screening is a computational technique used in drug discovery research to determine the best drug that binds at appropriate sites with target and enhance the rate of reaction. Further study investigated around 200 selective drugs using the drug bank, PubChem databases which revealed Cinnamoyl echinadiol as a single potent drug according to the FDA approved drug bank database with diverse ADMET, virtual screening and drug-likeliness property. Molecular docking analysis shows that, Polysaccharide biosynthesis protein as anovel drug target with Cinnamoyl echinadiol drug against pathogen *Eubacterium nodatum*. Overall study concluded that, subtractive genomics approach can be effectively applied to identify drug targets in pathogenic organisms for specific host organism.

Conflict of interest

We wish to confirm that there are no known conflicts of interest associated with this publication.

References

Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, Azevedo V, Kumar A. 2011. *In silico* subtractive genomics for target identification in human bacterial pathogens. Drug Development Research, 72 (2):162-177.

Chhabra G, Sharma P, Anant A, Deshmukh S, Kaushik H, Gopal K, Srivastava N, Sharma N, Garg LC. 2010. Identification and modeling of a drug target for *Clostridium perfringens* SM101. Bioinformation, 4(7):278.

De Groot AS, Sbai H, Aubin CS, McMurray J, Martin W. 2002. Immuno-informatics-mining the genomes for vaccine components. Immunology And Cell Biology, 80:255-269.

Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. Nucleic acids research, 43(1):389-394.

Duffield M, Cooper I, McAlister E, Bayliss M, Ford D, Oyston P. 2010. Predicting conserved essential genes in bacteria: in silico identification of putative drug targets. Molecular BioSystems, 6(12): 2482-2489.

Eke PI, Dye BA, Wei L, Thornton Evans GO, Genco RJ. 2012. Prevalence of periodontitis in adults in the United States: 2009 and 2010. Journal of Dental Research, 91(10):914–920.

Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, Griffin J, Baggaley RF, Jenkins HE, Lyons EJ, Jombart T. 2009. Pandemic potential of a strain of influenza A (H1N1): early findings. Science, 324(5934):1557-1561.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics, 28(23):3150-3152.

Galperin MY, Koonin EV. 1999. Searching for drug targets in microbial genomes. Current Opinion in Biotechnology, 10 (1999):571-578.

Guex N, Peitsch MC. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis, 18(15): 2714-2723.

Gupta SK, Singh S, Gupta MK, Pant KK, Seth PK. 2010. Definition of potential targets in *Mycoplasma pneumoniae* through subtractive genome analysis. Journal of Antivirals and Antiretrovirals, 2:38-42.

Harikishan Reddy EH, Satpathy GR. 2009.Identification of potential targets and lead molecules for designing inhibitory drugs against *Chlamydophila pneumonia*. Online Journal of Bioinformatics, 10(1):14-28.

Hill GB, Ayers OM, Kohan AP. 1987. Characteristics and sites of infection of *Eubacterium nodatum, Eubacterium timidum*, *Eubacterium brachy* and other asaccharolytic eubacteria. Journal of Clinical Microbiology, 25(8):1540–1545.

Holdeman LV, Cato EP, Burmeister JA, Moore CE. 1980. Descriptions of *Eubacterium timidum* sp. nov., *Eubacterium branchy* sp. nov., and *Eubacterium nodatum* sp. nov. Isolated from human Periodontitis. International Journal of Systematic Bacteriology, 30(1):163-169.

Hill GB. 1992. *Eubacterium nodatum* mimics Actinomyces in intrauterine device-associated infections and other settings within the female genital tract. Obstetrics & Gynecology, 79(4):534-538.

Hosen MI, Tanmoy AM, Mahbuba DA, Salma U, Nazim M, Islam MT, Akhteruzzaman S. 2014. Application of a subtractive genomics approach for in silico identification and characterization of novel drug targets in *Mycobacterium tuberculosis F11*. Interdisciplinary Sciences: Computational Life Sciences, 6(1): 48-56.

Hosen MI, Tanmoy AM, Mahbuba DA, Salma U, Nazim M, Islam MT, Akhteruzzaman S 2014. Application of a subtractive genomics approach for in silico identification and characterization of novel drug targets in *Mycobacterium tuberculosis* F11. Interdisciplinary Sciences: Computational Life Sciences, 6(1):48-56.

Hossain T, Kamruzzaman M, Choudhury TZ, Mahmood HN, Nabi AHM, Hosen M. 2017. Application of the Subtractive Genomics and Molecular Docking Analysis for the Identification of Novel Putative Drug Targets against *Salmonella enterica subsp. entericaserovar Poona*. BioMed Research International, 2017:1-9.

Huang B. 2009. MetaPocket: a Meta approach to improve protein ligand binding site prediction. OMICS A Journal of Integrative Biology, 13(4):325-330.

Ireland AJ, Soro V, Sprague SV, Harradine NWT, Day C, Al-Anezi S, Jenkinson HF, Sherriff M, Dymock D, Sandy JR. 2013. The effects of different orthodontic appliances upon microbial communities. Orthodontics & Craniofacial Research, 17: 115-123.

Jousimies-Somer H. 1997. Recently described clinically important anaerobic bacteria: taxonomic aspects and update. Clinical Infectious Diseases, 25(2):78–87.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research, 28(1): 27-30.

Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. 2009.KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Research, 38(1): 355-360.

Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nature Protocols, 4(3): 363.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols, 10(6): 845.

Koonin EV, Tatusov RL, Galperin MY. 1998. Beyond complete genomes: From sequence to structure and function. Current Opinion in Structural Biology, 8:355-363.

Köster H, Craan T, Brass S, Herhaus C, Zentgraf M, Neumann L, Klebe G. 2011. A small nonrule of 3 compatible fragment library provides high hit rate of endothia pepsin crystal structures with various fragment

chemotypes. Journal of Medicinal Chemistry, 54(22):7784-7796.

Kovačević SZ, Jevrić LR, Kuzmanović SOP, Lončar ES. 2014. Prediction of In-silico ADME Properties of 1, 2-O-Isopropylidene Aldohexose Derivatives. Iranian Journal of Pharmaceutical Research, 13(3):899.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26(2): 283-291.

Lipinski CA. 2000. Drug-like properties and the causes of poor solubility and poor permeability. Journal of Pharmacological and Toxicological Methods, 44(1): 235-249.

Magrane M. 2011. UniProt Knowledgebase: a hub of integrated protein data. Database.

Mondal SI, Ferdous S, Jewel NA, Akter A, Mahmud Z, Islam MM, Karim N. 2015. Identification of potential drug targets by subtractive genome analysis of *Escherichia coli* O157: H7: an in silico approach. Advances and Applications in Bioinformatics and Chemistry, 8:49.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M, KAAS. 2007. An automatic genome annotation and pathway reconstruction server. Nucleic Acids Research, 35:182-185.

Munster VJ, de Wit E, van den Brand JM, Herfst S, Schrauwen EJ, Bestebroer TM, van de Vijver D, Boucher CA, Koopmans M, Rimmelzwaan GF, Kuiken T. 2009. Pathogenesis and transmission of swine-origin 2009 A (H1N1) influenza virus in ferrets. Science, 325(5939):481-483.

Narayan Sarangi A, Aggarwal R, Rahman Q, Trivedi N. 2009. Subtractive genomics approach for in silico identification and characterization of novel drug targets in *Neisseria meningitidissero group B*. Journal of Computer Science and Systems Biology, 2: 255-8.

Nielsen M, Lundegaard C, Lund O, Petersen TN. 2010. CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles. Nucleic Acids Research, 38(2): 576-581.

Noble JM, Scarmeas N, Celenti RS, Elkind MSV, Wright CB, Schupf N, Papapanou PN. 2014. Serum IgG antibody levels to periodontal microbiota are associated with incident Alzheimer disease. PLoS ONE, 9(12).

O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. 2011. Open Babel: An open chemical toolbox. Journal of cheminformatics, 3(1): 33.

Perumal D, Lim CS, Sakharkar KR, Sakharkar MK. 2007. Differential genome analyses of metabolic enzymes in *Pseudomonas aeruginosa*for drug target identification. In Silico Biology, 7: 0032.

Petit J, Meurice N, Kaiser C, Maggiora G. 2012. Softening the rule of five—where to draw the line?. Bioorganic & medicinal chemistry, 20(18):5343-5351.

Ragno R, Marshall GR, Ballante F. 2016. U.S. Patent Application No. 14/901,924.

Rahman MA, Noore MS, Hasan MA, Ullah MR, Rahman MH, Hossain MA, Islam MS. 2014. Identification of potential drug targets by subtractive genome analysis of *Bacillus anthracis A0248*: An in silico approach. Computational Biology and Chemistry, 52:66-72.

Rathi B, Sarangi AN, Trivedi N. 2009.Genome subtraction for novel target definition in *Salmonella typhi*. Bioinformation, 4(4):143.

Rout GR, Samantaray S, Das P. 2000. *In-vitro* manipulation and propagation of medicinal plants. Biotechnology Advances. 18(2): 91-120.

Sakharkar KR, Sakharkar MK, Chow VT. 2004. A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. In Silico Biology, 4(3): 355-360.

Sakharkar KR, Sakharkar MK, Chow VTK. 2004. A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. In Silico Biology, 4: 0028.

Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. Journal of Computer-Aided Molecular Design, 27(3):221-234.

Tamura N, Ochi M, Miyakawa H, Nakazawa F. 2013. Analysis of bacterial flora associated with peri-implantitis using obligate anaerobic culture technique and 16S rDNA gene sequence. The International Journal of Oral & Maxillofacial Implants, 28(6):1521–1529.

Tareq Hassan Khan M. 2010. Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. Current Drug Metabolism, 11(4): 285-295.

Tzeng Y, Stephens DS. 2000. Epidemiology and pathogenesis of *Neisseria meningitidis*. Microbes Infection, 2: 687-700.

Uddin R, Saeed K. 2014.Identification and characterization of potential drug targets by subtractive genome analyses of methicillin resistant *Staphylococcus aureus*. Computational Biology and

Chemistry, 48: 55-63.

Uematsu H, Hoshino E. 1996.Degradation of arginine and other amino acids by *Eubacterium nodatum ATCC 33099*. Microbial Ecology in Health and Disease, 9: 305-311.

Valour F, Sénéchal A, Dupieux C, Karsenty J, Lustig S, Breton P, Ferry T. 2014. Actinomycosis: etiology, clinical features, diagnosis, treatment, and management. Infection and Drug Resistance, 7: 183–197.

Wadood A, Jamal A, Riaz M, Khan A, Uddin R, Jelani M, Azam SS. 2018. Subtractive genome analysis for in silico identification and characterization of novel drug targets in *Streptococcus pneumonia strain JJA*. Microbial Pathogenesis, 115: 194-198.

Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Assempour N. 2017. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Research, 46:1074-1082.

Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, Lu CH. 2014. CELLO2GO: a web server for protein subCELlularLOcalization prediction with functional gene ontology annotation. PloS One, 9(6): e99368.

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Brinkman FS. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics, 26(13): 1608-1615.

Zhang R, Lin Y. 2008. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic Acids Research, 37: 455-458.

Zhang R, Ou HY, Zhang CT. 2004. DEG: a database of essential genes. Nucleic Acids Research, 32: 271-272.

Zhang R, Ou HY, Zhang CT. 2004. DEG: a database of essential genes. Nucleic Acids Research, 32(1): 271– 272.

Zhou AQ, O'hern CS, Regan L. 2011. Reply to: Comment on Revisiting the Ramachandran plot from a new angle. Protein Science, 20(11): 1774-1774.

Zhu M, Lu M, Zhang J, Zhao F, Li G, Aihua Y, Zhao M. 2013. Carbon isotope chemostratigraphy and sedimentary facies evolution of the Ediacaran Doushantuo Formation in western Hubei, South China. Precambrian Research, 225:7-28.