" Identification of drug target against Bacteroides fragilis 638R: through Insilico genome analysis"

A Thesis Submitted in Partial Fulfillment of The

Requirement for the Degree in

Bachelor of Technology

In Biotechnology

Submitted by:-

Ajeet Singh

111BT0590

Under the Supervision of:-

Dr. Nandini Sarkar

Assistant Professor

Department of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela

Odisha-769008

Declaration

We Hereby Declare That This Thesis Is Our Own Work And Effort. Throughout This Documentation Wherever Contributions Of Others Are Involved, Every Endeavour Was Made To Acknowledge This Clearly With Due Reference To Literature. This Work Is Being Submitted For Meeting The Partial Fulfilment For The Degree Of Bachelor Of Technology In Biotechnology At National Institute Of Technology, Rourkela For The Academic Session 2011 – 2015.

Ajeet Singh

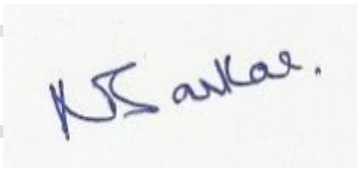
111BT0590

NATIONAL INSTITUTE OF TECHNOLOGY

ROURKELA 769008, INDIA

CERTIFICATE

This is to certify that the thesis entitled "**Identification of drug target against Bacteroides fragilis 638R: through Insilico genome analysis**" submitted to the National Institute of Technology, Rourkela by **AJEET SINGH**, Roll Number **111BT0590** for the award of the Degree of Bachelor of Technology In Biotechnology is a record of bona fide research work carried out by then under my supervision and guidance. The results presented in this thesis has not been, to the best of my knowledge, submitted to any other University or Institute for the award of any degree or diploma. The thesis, in my opinion, has reached the standards fulfilling the requirement for the award of the degree of Bachelor of technology in accordance with regulations of the Institute.



Date- 8th May -2015

Dr. Nandini Sarkar

Assistant Professor

Department of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela, Odisha-769008

Acknowledgement

If words are reflected as symbols of appreciation and token of Acknowledgement, then words play the role of thanks to exhibit the deeply

Embedded feeling of gratitude. I owe special dept. of gratitude to my guide

Dr. Nandini Sarkar, Department of Biotechnology and Medical Engineering

For her constant support and guidance throughout the course of the work.

It is an immense pleasure to thank my friends **Om Bikash ku das** and all others for their persistent encouragement and support.

I sincerely thank all my faculty members for their blessings.

Ajeet Singh

CONTENTS

Abstract:-

Bacteroides fragilis is a Gram-negative, rod shaped bacterium and the most standard anaerobic bacterium creating bacteremia in people. It is a part of the normal endogenous flora in human body and is normally commensal, but can cause infection if displaced into the bloodstream or surrounding tissue following trauma, surgery or disease. In this approach we have found certain target protein which can give rise to novel drug for the B.fragilis disease 638R. All the metabolic pathways which are present in the pathogen but not present in the human are taken as unique metabolic pathways. Here there are five pathways which unique and present only in bacteria. Whole genome sequence of the human pathogen **Bacteroides fragilis 638R** was explored to identify drugs targets. 526 Total number of protein coding genes were studied from B.fragilis, and 74 gene were having greater than 100 AA(amino acids) in there coding sequence were identified because of less than 100 amino acids in length were most unlikely to represent essential protein, we found 30 genes were identified human non-homologs. These human nonhomologs genes and there encoding protein were categorized on basis of the metabolic pathways involved in the basic survival mechanisms of the bacterium. After that we found **15** human non-homologous essential genes. Among all the human non-homologous essential genes **BF638R_1443** , having EC no: 5.1.3.20 is showing best Blast P result. This gene is present in the Cytoplasm and involves in the biological process like Carbohydrate metabolism process , Lipopolysaccharide biosynthesis. This in-silico genome analysis provides rapid and potential approach for identification of drug target and designing of drug.

Keywords: - Bacteroides fragilis 638R, drug targets, MSA, homology modeling, drug design.

List of tables:-

List of figures:-

ABBREVIATIONS:-

BLAST: - Basic Local Alignment Search Tool

NCBI: -National Center for Biotechnology Information

CEG:- Cluster of essential genes

B. fragilis: - Bacteroides fragilis

CASTp:-computed atlas surface topography of protein

DEG: - Database Essential Genes

PDB: - Protein Data Bank

DNA: - Deoxyribonucleic acid

EB: - Elementary body.

KEGG: - Kyoto encyclopedia of Gene & Genomes

KO: - KEGG Orthology

MSA: - Multiple sequence alignment.

PASS: -Putative Active Sites with Spheres

E-value: - expectation value

KAAS: - KEGG Automatic annotation server

1. Introduction:-

Bacteroides fragilis is a Gram-negative, rod shaped bacterium and the most standard anaerobic bacterium creating bacteremia in people. It is a part of the normal endogenous flora in human body and is normally commensal, but can cause infection if displaced into the bloodstream or surrounding tissue following trauma, surgery or disease[5]. It is foremost genus in the lower human intestinal nerve tract, as evidenced by its proliferation in a product of this faeces ,open-ended culture system, B fragilis is a part of the Bacteroidaceae family, involves B fragilis (causes the clinical infections), Bacteroides ovatus, distasonis, , vulgatus and thetaiotaomicron. These bacteria are immune to penicillin's, normally through the yield of beta-lactamase. This is the part of the GI florae and prevail in intra-abdominal and infections that beget from florae (eg, decubitus ulcers, perirectal abscesses). ETBF (Enter toxigenic Bacteroides fragilis) is too a potential cause of the diarrhea[7]. B fragilis are endogenous organic entity of GI tract. Bacteroides particles are anaerobic bacteria which is most important elements of bacterial florae of tacky membranes and are forasmuch a average cause of endogenous infections. Bacteroides infections can formulate in all body sites, allowing in the head, the CNS, the chest, the neck, the abdomen, the skin, the pelvis, and the pulpy tissues. Spread of strains amid patients is not known, in spite of that fact of this theme has not been next to good studied. In the Genetic background for antimicrobial refusal to accept in B. fragilis is varied with a some qualities obliging IS (insertion grouping) part embedded upstream for extended announcement. Inadequate therapy resistant to these anaerobic bacteria might lead to clinical failure[2]. Because their finickiest, they are oft overlooked and hard to isolate. Their isolation needs appropriate methods of cultivation, storage, and transportation of specimens. Treatment is complicated by three factors: slow growth, rising resistance to antimicrobial agent and polymicrobial synergistic character of infection[3]. Bacteroides fragilis, which represents just 0.005(0.5%) of the humans flora is the ordinarily segregated anaerobic bacteremia pathogen due, partially, to its strong envy variables. Components of the class Bacteroides has the best anti-microbial resistance cracy and the rates of higher resistance of all anaerobic bacteremia pathogens. Clinically, Bacteroides station have displayed rising imperviousness to a few anti-infection agents, including

metronidazole, cefoxitin, carbapenems, metronidazole, fluoroquinolon and clindamycin (e.g., moxifloxacin, levofloxacin and gatifloxacin) [16, 21]. The human colon in the body has the biggest people of microscopic organisms (1011 organic entities/gram of wet wt.) & the significant piece of body are most of anaerobes; ~25% types of Bacteroides. This review will abbreviate present state Bacteroides species, the most important anaerobes in gut. The facet of these organisms that will covered include their role as commensal organisms, their participation in human disease, and information about their resistance, metabolism and physiology mechanisms as well as a Brevity review of clinical features. Bacteroidetes are one of the biggest line ages of bacteria during their evolutionary process. Bacteroide particle are anaerobic bacterium, non spore forming, bile-resistant, gram-negative rods [18]

Several research work are going on logically to build up the effective drugs by genetic or genomic approaches. Novel drug target are outline to protect against anti-toxin touchy microorganisms. New compelling strategy has been created in bioinformatics for discovering sorted out focuses from prior from unexplored cell capacities and to identify internal organic procedure of pathogen[32,29]. The complete genome data is likewise vital for determination of precise way to deal with check vitality and selectivity benchmark of the organism. The focus of the methodology ought to be substantive encoded quality for the replication, development and survival of pathogen. The complete genome data is additionally urgent for determination of precise way to deal with check vitality and selectivity benchmark of the organism. The focus of the methodology ought to be substantive encoded quality for the replication, development and survival of pathogen. This target ought not to make any cytotoxicity harm to host. The qualities called as "key qualities" that are introduce in distinctive saved space of genome and key for the survival of the organic entity. These fundamental qualities can't continue inactivation through the transformation process. The unforeseen poisonous mutants help to adjust the status of these qualities [1, 6]. Ending the capacity of key qualities results passing managing constitution inside microbes. So it will be not useless by tending to these medications as "super shot" against pathogen. This won't just help to deflect cost additionally simple to distinguish destructive inhibitors by perceiving broaden drug targets[13].

Now a days it's very easy to recognize the targets by insilico-genomic approaches. "Differential genome display" is one of the beneficial approach for the anticipation of likely drug targets. This technique offers point of interest genomic data of pathogens i.e. how the complete arrangement of qualities and protein are encoded inside the little genome. The qualities which introduce in pathogen, yet truant in human are called non-homologous qualities[2,5]. These are most key parts for in silico genome examination. Utilizing bioinformatics instruments and procedures, the medication targets can be perceived so effortlessly from these qualities. The genes which are in charge of the establishment of life are known as the key genes. The function of key genes are normal to all cells. For the sustainment of diseases is based to work out for hostile to microbial specialists against microorganisms[7]. The portrayal of specific fundamental qualities for particular pathogen can be utilized as medication focus as a part of a few ration space of that microscopic organisms. DEGfuses the rundown of essential genes of some restricted pathogen. It is anything but difficult to experience the vitality of genes after the fruitful advancement and implementation of human genome project databases. So it is keeping an eye on one stage ahead improvement for novel medication target approaches. Anti- bacterial medication targets could be possible by perceiving the particular essential genes by "subtractive genome approaches"[1]. Subtractive genome methodologies is fruitful executed in this research paper to distinguish the potential drug target for Bacteriodus fragilis. The essential genes for Bacteriodus fragilis 638R are recorded effectively by aiding Database of Essential Genes (DEG) against human genome. The genes present in Bacteriodus fragilis 638R, nearly identified with human genome are called as homologous genes and these genes are discarded[11].

The potential drug targets are viably utilized as a part of immunization purposes. Immunization gives get safety to the avoidance of particular disease. Antibody contains operators, which are a piece of a living being utilized to execute that organic entity. Immunizations may be poisons, surface proteins or dormant piece of the living being which triggers the resistant framework to decimate the pathogen by distinguishing and recording the danger. Agent immunizations can be produced by focusing on the qualities show in cell divider or plasma layer[15].

Kyoto Encyclopedia of Genes and Genomes database gives interesting metabolic pathway map Bacteriodus fragilis 638R. As we are focusing on the genes situated in cell divider or plasma layer, four essential metabolic pathways are selected like c5-Branched dibasic acid metabolism pathways in Carbohydrate metabolism system, Methane metabolism pathways in energy metabolism, Lipopolysaccharide biosynthesis and Peptidoglycan biosynthesis. Protein accomplishes useful adaptation by communicating with diverse particles like ligand, substrate, DNA and different proteins[34]. It's exceptionally essential to acquire the particular three-dimensional protein structure for the ID of fitting cooperation by picturing the shape, physical, synthetic and organic properties. By the establishment of protein surface portrayal help to dissect determination of tying, catalyst system and look at for transformation. Another vital methodology is by visualizing action of protein utilizing structure-based drug design (SBDD). The substrate binding site of protein aides in conformational changes and chemical modification. This particular binding site of protein help to trigger actualizing the therapeutics approach for interruption in biological processes of pathogen[19].

2. Objective

The objective of this project is to identify a gene of unique metabolism pathways which can act a potential drug target for **Bacteroides fragilis 638R.** That gene should be non-homologous to host (human).

3. LITERATURE REVIEW:-

3.1 Bacteroides fragilis:-

Bacteroides fragilis is a Gram-negative, rod shaped bacterium and the most standard anaerobic bacterium creating bacteremia in people. It is a part of the normal endogenous flora in human body and is normally commensal, but can cause infection if displaced into the bloodstream or surrounding tissue following trauma, surgery or disease. Bacteroides partake in intra-stomach abscesses when intestinal substance spill into the peritoneum Arrangement of abscesses is a synergistic procedure including anaerobes & facultative microbes B. fragilis has uncommon ability to endure oxygen and to prompt canker arrangement through its CPC Different Bacteroides-like anaerobes are included in polymicrobial dental, lung, or pelvic diseases[23]

Bacteroides species are huge clinical pathogens and found in most of anaerobic contaminations, with a related mortality of greater than 19%. The microorganisms keep up a complex and by and large gainful association with the host when held in the gut, however when they get away from this environment they can bring about critical pathology, including bacteremia and sore development in numerous body destinations. Genomic and proteomic examinations have unfathomably added to our comprehension of the way in which Bacteroides species adjust to, and flourish in, the human gut.It has a complex capsular polysaccharide that is fundamental for sore arrangement -Made out of no less than 8 polysaccharides, Each is fit for transcriptional stage variety and Polysaccharide An is fundamental for abscesses in creature models and is zwitterionic [6, 13].

B. fragilis are more aerotolerant than different species and more impervious to responsive oxygen species

- Possesses a superoxide dismutase (SOD)
- Possesses catalase (CAT)

The outer film LPS (lipid An) is changed to be less noxious than that of E. coli Considers host resistance of immense amounts of living creatures without let[1,15].

Bacteroides spp are non-spore shaping gram-negative bacilli that are a piece of the human occupant flora. Microbiologically, they are recognized from other genera by development in 20% bile. At present, the Bacteroides fragilis gathering comprises of ten species: B. fragilis, B. thetaiotaomicron, B. merade B. eggerrthii, B. merdae, B. ovatus, B. stercoris, and B. caccae. Since 1990, numerous living beings beforehand assigned as Bacteroides have been renamed .Bacteroides, the prevalent variety in the human digestive system, are imperative in various metabolic exercises and may give some level of assurance from intrusive pathogens. Each of the 10 species are generally segregated from the colon, despite the fact that contaminations brought about by or connected with them can incorporate practically any organ.

Bacteroides spp are Obligate anaerobes and 25% of all colonic microbes . Normally included in diseases coming about because of aperture of a stomach viscus

- Diverticulitis
- ruptured appendix
- post-operation after inside surgery anddehiscence of a surgical anastamos

Any Bacteroides spp. may be involved in a polymicrobial infection, but most abscesses contain B. fragilis.

Anti-microbial resistance in Bacteroides- Most convey a beta-lactamase quality (impervious to penicillin, ampicillin, and 1st gen. ceph.), Harbors conjugative transposons, Can trade qualities with different Bacteroides and with different species ex. clindamycin resistance (just ~60% touchy at this point)

3.2 Infection:-

Infection with intra-stomach origin: Bacteroides sites contaminations in kids reflect those found in grown-ups. As in grown-ups, Bacteroides confines are most transcendent in diseases that have an intra-stomach cause; regularly exhibit in the GI tract creatures shall enter the peritoneal depression because of an unsettling influence, for example, puncturing, impediment, or direct injury[10]. A couple of studies assessing the micro-biology of peritoneal depression and postoperative injuries in kids taking after punctured reference section in pediaric patient found Bacteraides species were recouped from 0.93 of peritoneal liquids, alongside enteric gram-ve microorganisms and enteracocci . Inconveniences taking after peritonitis may

incorporate subphrenic, hepatic, (which may happen auxiliary to an infected appendix), necrotize enter-colitis, pelvic provocative infection, tubo-ovarian contamination, surgery, or injury [16]. B. fragilis are the most widely recognized an-aerobe found in postsurgical injury contaminations in wounds identifying with the gut verdure . Not surprisingly, wounds and different subcutaneous tissue contaminations in the rectal territory, or that generally began from the gut verdure, are regularly polymicrobial and frequently associated Bacteroides species [18].

Bone and joint infection in children: Anaerobes have once in a while been accounted for as a reason for joint and bone contaminations in kids. In the event that found, anaerobic contaminations in joint inflammation ordinarily include a solitary detach; the segregates discovered incorporate anaerobic gram-negative bacilli, Clostridium spp.and Peptostraptococcus spp. Anaerobic joint inflammation is by and large auxiliary to hematogenous spread. Anaerobic osteomyelitis will for the most part happen because of an anaerobic disease somewhere else in the body and may include more than one living being. Some of these contaminations may bring about positive blood societies, and the creatures recouped are like those from the contaminated locales [30].

3.3 Diagnosis:-

B.fragilis has been concentrated on widely and it is known to have been the cause for some intestinal aggravations and it's additionally the primary driver of acute and chronic diarrheal infection in people and animal. B. fragilis is additionally the reason for abscess arrangement. Because of the development at the disease site, canker arrangement needs to be cleared as quickly as time permits generally death rate can increment. The B fragilis that was extricated contains a component that is profoundly connected with sicknesses [31]. The enterotoxins were distinguished through a directed study in which tests from patients stool was gathered and procedures, for example, PCR, measures, distinctive blotching methods, and so forth

were utilized to distinguish the enterotoxins. Treatment of diarrheal maladies in youngsters is generally simple, the utilization of rehydration to reduce the effect of sickness [7].

A few types of Bacteroides produce succinic corrosive as a destructiveness variable (ex. B fragilis). Bfragilis produce polysaccharides container high in succinic acid. Once Bfragilis is discharged from the case, it incapacitates the relocation of leukocytes, which is needed for the site of recuperating. At the point when the B fragilis slaughters off the diseases developed and if left untreated then the demise rate is high 60%. Succinic corrosive was utilized to test the neutrophil capacity and they have discovered that the succinic corrosive improves the destructiveness element of Bacteroides. Something else they've found is that the harmfulness variable increments with lower pH and in micro-environment with high contaminations [29].

3.4 Treatment:

The understanding's recuperation from anaerobic disease relies on upon brief and proper administration as per the accompanying 3 standards:

- Poisons delivered by anaerobes must be killed.
- Nature must be changed to counteract nearby bacterial multiplication.
- The spread of microorganisms must be restricted

The earth is controlled by debriding necrotic tissue, depleting discharge, enhancing dissemination, easing obstacle, and expanding tissue oxygenation. Certain sorts of adjunctive treatment, for example, hyperbaric oxygen treatment, may be valuable yet stay doubtful. Much of the time, antimicrobial treatment is the main manifestation of treatment needed, yet it can likewise be utilized as a subordinate to a surgical methodology. Since anaerobic microorganisms are for the most part recouped blended with high-impact life forms, the proper decision for antimicrobial operators ought to give satisfactory treatment of both gatherings of pathogens [23].

A few antimicrobials have a constrained scope of action. For instance, metronidazole is dynamic just against anaerobes and can't be directed as a solitary operators in blended diseases. Others, for example, imipenem,

have wide spectra of movement against aerobes and anaerobes. Since society results are regularly not accessible, numerous patients are dealt with experimentally [8].

Recent reports of multidrug-safe B fragilis, underscores the requirement for enhanced anti-infection stewardship. Despite the fact that B fragilis has long been considered dependably vulnerable to various wide range ant anaerobic drugs,[3] these cases recommend clinicians ought to no more depend on total defenselessness information from overviews alone to direct treatment and ought to consider asking for helplessness testing when treating genuine contaminations brought on by B fragilis[2].

3.5 Current Research:

Current research of B fragilis incorporates: improvement of anti-biotics agents, its diseases on embryo and mother, the formative examination of well-disposed business items utilizing by-results of Bacteroides fragilis and significantly more. As a result of the one of a kind attributes of gram-negative bacteria and its resistance to anti-biotics agents, numerous studies have been done to attempt and comprehend its components of activity. Bacteroides fragilis is extremely remarkable it might be said that it has its own specific manner of getting to be impervious to anti-toxins thus far, no studies have possessed the capacity to distinguish their components of activity. Another anti-toxin under study, Trospectomycin, is utilized as a part of treatment against Bacteroides fragilis and results proposed that it’s a decent anti-infection against Bacteroides fragilis. Trospectomycin beat anti-microbials like ampicillin, doxycycline, it acts by tying to the 30s subunit and hinders protein combination. It's demonstrated amazing against numerous gram-ve microbes. Still, more research are being the performed to comprehend and to create anti-toxins that will treat diseases from the enterotoxigenic Bacteroides fragilis [13 19].

Another case that has been concentrated on is the relationship b/w B fragilis and its pathogenesis on the human genital tract. Despite the fact that Bacteroides fragilis was confined, it may not be the sole reason for the genital contaminations. A study has been done to see the connection between Bacteroides fragilis and vaginal contamination amongst pregnant ladies. Albeit both enterotoxigenic and the non-enter toxigenic

strains of Bacteroides fragilis were discovered when they did the study, the transmission system is still to a great extent obscure and its part have not been completely caught on. Another huge hypothesis being dead set is the connection between the improvement of the hatchling and the Bacteroides fragilis [11].

Succinic corrosive is a by-item created from Bacteroides species from the deciding result of their vitality digestion system Succinic corrosive is critical in numerous business items because of its modern applications in sustenance, pharmaceuticals, makeup and numerous others. Up to this point, numerous succinic acids have been made synthetically and have raised major ecological concerns amongst individuals. Another strategy is to deliver succinic corrosive by microbial aging; its straightforward and natural cordial. Then again, its advancement is still under study [12].

3.6 **Application to biotechnology:**

Through the EBI site, four chemicals were found for Choloyglycine hydrolaseBacteroides fragilis Fragilysin, 7-alpha-hydroxysteroid hydrogenase, and phosphoenolpyruvate. These compounds assume a part in catalyzing responses that happen in the pathway. Some likewise obliged a cofactor (Zinc ,Magnesium,) and others used (ex. NADP+, NAD+). Case in point, Fragilysin obliges Zinc, which upsets the tight intersections of the intestinal divider. This is thought to be the protein that helps Bacteroides fragilis break to different parts of the body where it discharges its poison to taint the site of activities. In any case, this creature does not add to any bio-technology known but rather examination is still under study.

4. TOOLS FOR THE STUDY

4.1 Tools used for sequence alignment:-

Sequence alignment gives a powerful way to compare novel sequence, Sequence alignment tools are used for comparability of amino acid sequences and characterized query genes. Basic Local Alignment Search Tool to compare and quick search of protein and nucleotide sequences from database. BLAST provides both local and global search alignment algorithm facilities to find the similarities from conserved domains of sequences. BLAST provides much faster alignment process implementing Smith–Waterman algorithm. There are five different version of BLAST like BLASTn, BLASTp, BLASTx, tBLASTn, tBLASTx. BLASTn assists to compare nucleotide sequences nucleotide databases. BLASTp assists to compare amino acid sequences from protein databases. BLASTx is used to compare six entrapped transcription product of a nucleotide sequences vs protein sequences. tBLASTx is used to compare six entrapped translation nucleotide sequence vs 6 entrapped sequence of nucleotide from database. tBLASTn is used to compare compare six entrapped translation nucleotide sequence vs six protein sequences from database [25, 32].

4.2 Online software used for Homology Modelling

Homology modelling:-

Model-fabricating by homology is a valuable system when 1st needs to anticipate the str of a target protein of known sequence when the target protein is identified with minimum another protein of known arrangement and str. If the proteins are nearly related, the known protein structures ‘called the folks’, can serve as the premise for a model of the target. Despite the fact that the nature of the model will rely on upon the level of likeness of the groupings, it’s conceivable to allocate this quality before test testing [33].

Steps in Homology modelling

1. Adjust the AA (amino acid) sequence of the target protein of known structure. It will normally be observe that insertion and erasures lie inside the loop regions b/w helices and sheets.

2. Define main chain fragments to describe to the districts insertions or erasures. Sewing these districts into main chain of known protein makes a model for the entire the target protein mainchain.

3. sidechains Replace of buildups that have been transformed. For relic that have not changed, hold the sidechain conformity. Relics that have changed have a tendency to keep the same sidechain conformational edges, and could modelled on this premise. Methods of Computational are presently accessible to search conceivable mixes of sidechain conformities.

4. Look at the model both by eye and by program to distinguish any genuine crashes between atoms. Mitigate these impacts, beyond what many would consider possible by manual manipulation.

5. Retouch the model by constrained minimization of energy. The part of this step is to fixed up the definite geometrical connections at spots where areas of mainchain has been joined together in agreeable situation. The impact is really corrective - energy refinement won't alter genuine blunders in such a model.

Swiss model (www.expasys.ch/swiss-mod/ SWISSMODEL.html) is used for automated homology modelling. It has a first approach mode that aides performs Homology Modeling. The user needs to enter his/ her email id and information the protein arrangement in Fasta position. It permits the user to pick as far as possible for format choice. It can seek the pdb document from the pdb database with the user giving the name of the pdb record or the client can transfer his/ her own pdb document. The yield record is a pdb document that is come back to the user's email address. The outcome can be sent by Swiss Model to PHD Secondary structure forecast at Columbia University furthermore, Fold Recognition Server (3D-pssm) of the ICRF.

Geno3D:-

(http://geno3d-pbil.ibcp.fr) performs Comparative protein structure modeling by spatial limitations (separations and dihedral) fulfillment. Geno3D is most habitually utilized for Homology or Comparative protein structure Modeling.Geno3d acknowledges information like Fasta organize yet just the one letter code

must be utilized. The outcome is gotten in the pdb design that can be seen in any Molecular Modeling software.Geno3d offers numerous other highlights, it permits the user to choose PDB entrances as formats for Molecular Modeling after a 3 stage iterative PSI BLAST. It exhibits the yield for every layout, alongside the optional structure forecast, shows percent of assention in auxiliary structure and repartition of data from format on inquiry succession. The yield connection is sent to the user's email address. It likewise informs the client when its server starts the Homology Displaying. It has an alternative where the user can choose what number of models to create. The fundamental thought behind having more than one model created is that the client may have a superior adaptability and comprehension. It likewise gives back a superimposed pdb document which has the models superimposed on one another. This is one of the great focuses in Geno3d as it permits us to think about the different models created in one window. All the outcomes acquired can be downloaded as an archive.tar.Z that can be opened in WinZip in windows and in UNIX or Linux stages. So the user does not need to spare results in site page impact or in an archive record. It likewise shows the Ramachandran plot in the outcome.

CPHmodels :-

Computerized neural-system based protein demonstrating server (Http://www.cbs.dtu.dk/administrations/CPHmodels/). CPHmodels is a gathering of databases and what's more, routines created to anticipate protein structure. It performs expectation of protein structure utilizing Comparative Modeling. It doesn't acknowledge more than 900 amino acids in the data succession. The arrangements are kept classified and are erased in the wake of preparing. This system did not issue me fitting results. The error it showed was like the one showed by Swiss Model.

4.3 Structure Analysis and Verification Server

DOPE: The DOPE model score is intended for selecting the best structure from an accumulation of models assembled by Modeler. DOPE utilizes the standard Modeler Function capacity.

PROCHEK: Checks the stereo chemical nature of a protein structure by examining deposit by-buildup geometry and general structure geometry. It is tell about: Covalent geometry, Planarity, Dihedral points, Chirality, Non-reinforced cooperation's.

WHAT_CHEK: Derived from a subset of protein check apparatuses from the WHATIF program; this does far reaching checking of numerous sterochemical parameters of the buildups in the model.

ERRAT: ERRAT is a protein structure confirmation calculation that is particularly appropriate for assessing the advancement of crystallographical model refinement and build. The system works by dissecting the insights of non-reinforced cooperation's between distinctive molecule sorts. A solitary yield plot is created that gives the estimation of the slip capacity versus position of a 9-buildup sliding window. By comparison with statistics from profoundly refined structures, the Error value have been aligned to give certainty limits. ERRAT will give a "general quality element" and in the event that it is a high 90% range protein structure is great. This is to a great degree valuable in settling on choices about unwavering quality [9].

VERIFY_3D decides the similarity of a nuclear model 3D with its own AA (amino acid) arrangement (1D) by allocated an auxiliary class taking into account its area and environment (circle, polar, alpha, beta, nonpolar and so forth) and contrasting the outcomes with great structures. At that point a database created from considered great structures is utilized to get a score for each of the 20 amino acids in this basic class. For every deposit, the scores of a sliding 21-buildup window (from -10 to +10) are included and plotted [9].

4.4 Protein active site prediction tools:-

The stereo-chemical approval of model structures of proteins is a critical part of the near atomic modeling methodology. 1st the determination of high caliber structures for incorporation in loop lexicons is critical for the straightforward reason that these direction sets will be utilized to assemble future models. 2nd Also, the auxiliary assessment of relative modeling yield must be utilized to distinguish conceivable dangerous locales. There is a few estimations are great indicators of stereo chemical quality; these incorporate planarity; unsatisfied contributors: chirality; phi/psi inclinations; non-reinforced contact separations; chi points and acceptors. Protein performs its capacity through communication with different particles, for

example, substrate, DNA ligand and different spaces of proteins. Auxiliary data of protein surface areas empowers definite investigations of the relationship of protein structure and capacity. In particular, portrayal of protein surface districts serves to investigate catalyst system, to focus tying specificity and to arrange change studies. It can likewise help to distinguish the natural parts of recently illuminated protein structures with an obscure function [10].

CASTp:-

Figured Atlas of Surface Topography of proteins (http: //cast.enger.uic.edu.) gives an online asset to finding, outlining and measuring curved surface locales on 3D str of proteins. These incorporate pockets situated on surfaces of protein and voids covered in the inside of protein.

PASS:-

Putative Active Sites with Spheres is a straightforward computational device that employments geometry to portray locales of covered volume in proteins and to distinguish positions prone to speak to tying locales based upon the size, shape, and entombment degree of these volumes. Utility as a prescient device for tying site distinguishing proof is tried by foreseeing known tying destinations of proteins in the PDB utilizing both complexed macromolecules and their relating epo-protein structures.

4.5 Structure Analysis and active site prediction:-

1. **Ramachandran plot.** **2. Sidechains**

Ramachandran plot: - The Sasisekharan-Ramakrishna-Ramachandran plot depicts permitted mainchain conformation. A Ramachandran plot is an approach to envision dihedral angles φ against ψ of amino corrosive deposits in protein structure. It demonstrates the conceivable adaptations of φ and ψ plots for a polypeptide.

A section of the direct polypeptide chain normal to all the protein str is indicated in fig. Rotation is allowed around the Cα-C and N-Cα single obligations of all buildups (one special case: proline). The edges ψ and φ around these bonds, and the edge of pivot near the peptide bond, ω, characterize the compliance of a residue.

The peptide bond it-self has a tendency to planar with two permitted states: cis, ω ≈ 0° (infrequently) and Trans, ω ≈ 180° (ordinarily) the sequence of φ, ψ and ω angles of all deposits in proteins characterizes the back-bone conformation.

The principle that two particles can't possess the same space restrains estimations of angles of conformational. The permitted ranges of φ & ψ for ω = 180°, fall into characterized areas in a chart called Sasisekharan- Ramakrishna-Ramachandran plot - as a rule abbreviated to 'Ramachandran plot' Fig. in the figure the Solid lines delimit energetically prefer areas of φ and ψ; softened lines up the figure delimit satirically-prohibited regions. Fig.2 demonstrates the average dispersion of residue compliances in an all-around decided protein structure. Most residue fall in or close to the permitted locales, in spite of the fact that a couple are constrained by the collapsing into enthusiastically less-positive states.

The allowed areas create standard conformities. A stretch of continuous relics in the conformation creates a α-helix. Rehashing the β conformation creates an expanded β-strand. More than β-strands can connect along the side to β-sheets. Sheets and helices are standard or "pre-assembled" basic pieces frame parts of the compliances of mostly proteins. They are settled by generally feeble communications, hydrogen securities, between main chain molecules. In a few stringy proteins the greater part of the buildups have a place with one of these sorts of structure-fleece contains silk β sheet, α helices.

4.6 Deep view software:-

Swiss-PdbViewer is an application that gives an easy to use interface permitting to break down a few proteins in the meantime. The proteins can be superimposed in request to derive basic arrangements and look at their dynamic destinations or some other important parts. Amino corrosive changes, H-bonds, angles and separations between particles are anything but difficult to get because of the natural realistic and menu interface. Deep View - Swiss-PdbViewer has been developed by Nicolas Guex (GlaxoSmithKline R&D). Swiss-PdbViewer is hard connected to SWISS-MODEL, a robotized homology modelling server created inside the Swiss Institute of Bioinformatics (SIB) at the Structural Bioinformatics Group at the Biozentrum in Basel

NCBI

NCBI (National Center for Biotechnology Information), it is a part of the United States National Library of Medicine. The NCBI progresses science and health by giving access to biomedical and genomic data. It contains the tools such as PubMed, GenBank (provide the nucleotide sequence), and BLAST (sequence similarity),

BLASTP

Basic Local Alignment Search Tool (BLAST). It is a device used to contrast a query sequence of nucleotides and a database of sequence to distinguish the closeness between the two sequences above a threshold. BLASTP: Here the amino acid sequence is compared. It gives clear data about the closeness of genes and which is related with some protein**.**

KEGG:-

Kyoto Encyclopedia of Genes and Genomes(KEGG) is a collecting of databases managing genome, drugs, biological pathway, chemical substances, diseases, Used for analysis of genomics, metagenomics and metabolomics It is an amassing of pathway maps fusing various substances including qualities, proteins, RNAs, substance blends, glycan, and compound reactions, and also illness qualities and targets, which are secured individual doorways into alternate databases of Kyoto Encyclo-pedia of Gene & Genomes. The KEGG database of work was launched by Minoru Kanehisa in 1995 .Anticipating the requirement for an electronic asset that can be utilized for natural elucidation of genome arrangement information, he began building up the KEGG PATHWAYs database.

DEG:

Database of Essential Genes, is a database and give apparatus to investigation the essentiality of the gene. Essential genes are those genes of a living being that are thought to be basic for its survival of the organism.DEG hosts records of presently accessible key genomic components, for example, protein-coding genes and non-coding RNAs, among , eukaryotes, microorganisms and archaea . Essential genes in a bacterium constitute an insignificant genome, shaping an arrangement of utilitarian modules, which assume key parts in the developing field, manufactured science.

UNIPROT

It provides the information of the gene about the sequence, location and function in the cell. UniProt

Knowledgebase is a protein database halfway curated by specialists, comprising of two areas: UniProtKB/Swiss -Port and UniProtKB/ TrEMBL.

PSORTb:-Computational prediction of subcellular restriction of proteins is a profitable instrument for genome examination and annotation, since a protein's subcellular limitation can give educates in regards to its capacity an organic entity. For bacterial pathogens, the forecast of proteins on the cell surface is exceptionally compelling because of the capability of such proteins to be essential medication or antibody targets. A few algorithms have been produced to dissect single highlights, for example, however the PSORT group of projects breaks down a few highlights immediately, utilizing data got from every investigation to produce a normal expectation of confinement site. Grown in 1991 by Kenta Nakai, PSORT is a calculation which allots a plausible confinement site to a protein given an amino acid sequence.

CEG:-CEG (Cluster of essential genes) are obtained from database DEG, which has been illumined in NAR in 2009 and 2004. CEG is a database containing clusters of orthologous key genes grew by CEFG Group in UESTC. Various from DEG, CEG database store crucial genes as orthologous gatherings and not in a single genes. All essentials genes in distinctive bacterial creatures are contained in one CEG bunch on the off chance that they have the same capacities. With this adjustment from DEG, clients could without much of a stretch choose that a vital quality is moderated in various bacterial pathogens or species-particular.

5. MATERIALS & METHODS:

5.1 *Methodology*

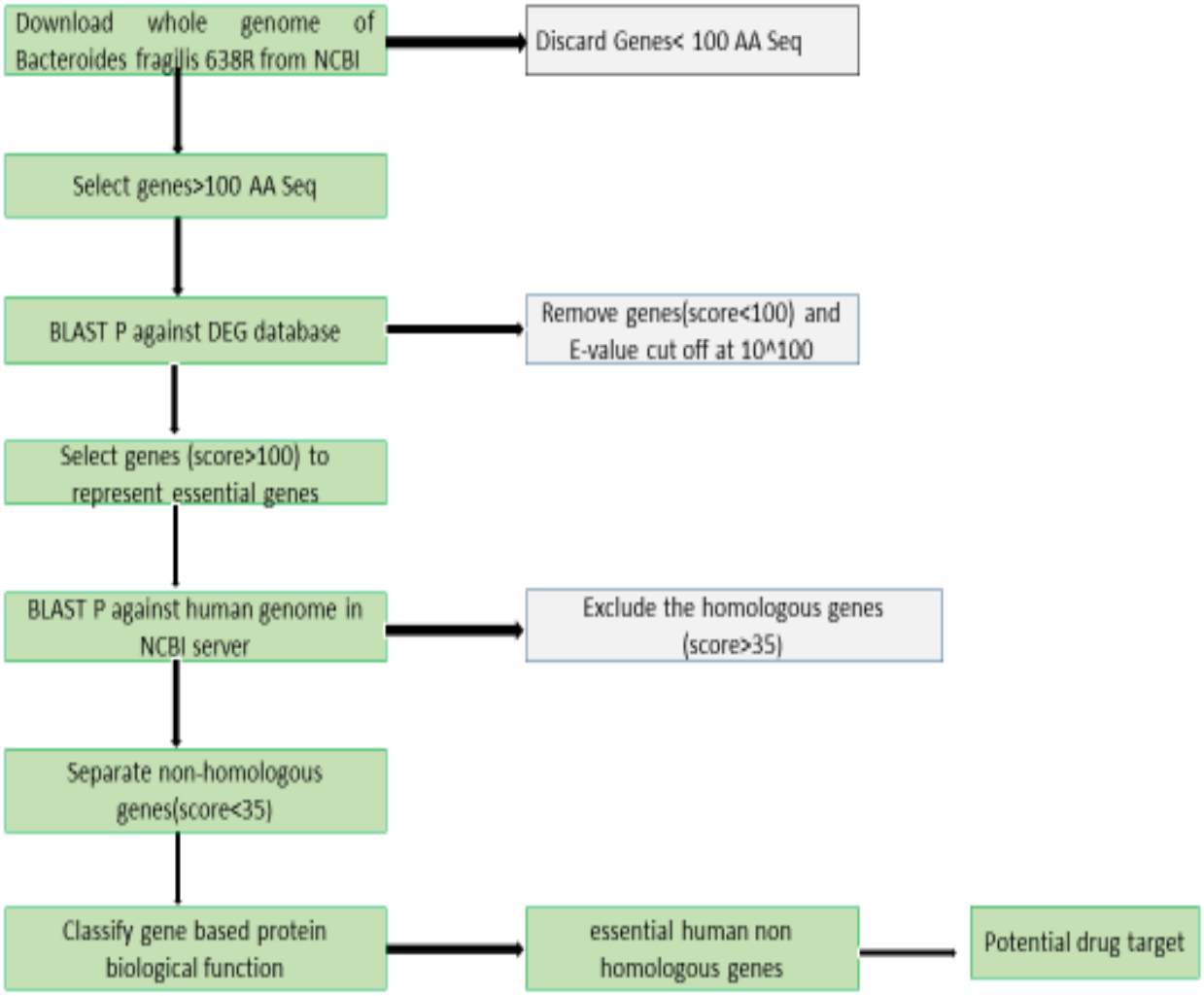


Fig: 2 methodology for identification of novel drug targets.

5.2 Retrieval of proteome from NCBI:-

The complete set of protein (proteome) is revived from NCBI. The sequence less than 100 amino acid sequence are consider to be paralog or duplicate protein. The non-paralog proteins are selected and paralog are eliminated.

5.3 Identification of Unique metabolic pathways:-

KEGG database is used for the selection of unique metabolic pathways from Bacteroides fragilis 638R and human. Some unique metabolic pathways are selected to identify appropriate genes .Identification gene of genes in unique metabolic pathways.

1. Search KEGG website
2. click on metabolism pathway
3. select the pathway listed above
4. dropdown box search the Bacteroides fragilis 638R
5. click on pathway entry
6. Then select metabolism from drop down box.

5.4 Identification of essential genes involved in unique metabolic pathway

- The AA sequence of pathogen are recovered from NCBI. For the choice of non-paralog protein, the outcome is submitted for CD-Hit suit.
- The proteins having seq. lengths under 100 amino acids were excluded as they are less inclined to represent to essential genes.
- The Bacteroides fragilis 638R proteins were investigated utilizing CD-HIT to distinguish the paralogues or copy proteins. Seq identity cut-off was kept at 0.6 and worldwide seq identity algorithm was chosen for arrangement of the amino acids; a transmission capacity of 20 amino acids furthermore, default parameters for alignment scope were selected.

5.5 Identification of non-homologous genes:-

Using BLASTP tool homologous and non-homologous genes can be differentiated. Homologous genes are present in both human and pathogen. Elimination of homologous genes are necessary, because these genes involves in the common biological processes and vaccination will be not effective. For the selection of

essential non-homologous genes the identity is considered below 35 % and expected threshold value is set at 0.005. We are targeting the most conserved bacterial to get best result for multi resistant strain pathogen.

5.6 Homology modelling of identified potential drug target:-

Model-constructing by homology is an important system when one needs to anticipate the structure of a target protein of known succession, when the target protein is connected with no less than one other protein of known grouping and structure. In the event that the proteins are almost related, the known protein structures - called the folks - can serve as the reason for a model of the target.

Homology Modelling using MODELLER 9.14:-

Modeler is utilized for homology or similar displaying of protein three dimensional (3D) structures. It will runs on python script file commands. Modeler is most regularly utilized for homology or near protein structure modeling. Modeler aides determine the spatial restraints from the templates. It produces various 3D models of the succession you submit fulfilling the template limitations. These incorporate NMR tests, fluorescence spectroscopy, cross-connecting trials, standards of auxiliary structure pressing ,image reproduction in electron microscopy, homologous structures (similar modeling),site-coordinated mutagenesis, deposit buildup and atom-atom potentials of mean force, and so forth. Modeler is not a mechanized homology modelling device. It is an exceptionally program and it permits the client to indicate what he needs at last result. Modeler runs on stages like Win XP, Macintosh. Linux and Sun Solaris.

Methods:-

The first 4 best result having PDB accession No. from are taken from NCBI-BLAST P result. The PDB file of 4 are downloaded from rcsb.org. The 4 PDB files are named as tseq 1, 2, 3&4. The ALI file is created and named as qseq. 5 python script are created and run with Modeller 9.14. The result will be generated in .txt format. We got 5 templet PDB files from Modeller and among them qseq.B99990003.pdb is showing the higher GA341 score. So this file is taken for further experiment.

5.7 Active site prediction for protein:

Active sites of the target protein was anticipated by utilizing tools like PASS, CASTp which would be the key variable for the adaptable docking. This gives asset for locating, depicting and measuring concave surface regions on three-dimensional structures of proteins. These incorporate pockets situated on protein surfaces and voids covered in the inside of proteins that are frequently connected with binding events. In addition, it gauges the measure of mouth openings of individual pockets, for better accessibility of binding sites to different ligands and substrates.

5.8 Deep view software:-

Swiss-PdbViewer is an application that gives an easy to understand interface permitting to examine a few proteins in the meantime. The proteins can be superimposed to find auxiliary arrangements and think about their dynamic locales or whatever other important parts. Amino acids mutation, angle, H-bonds, and separations between atoms are easy to get because of the instinctive realistic and menu interface.

Method:-

- Go to “Swiss Model” in “Swiss pdb viewer”.
- Select “load raw sequence from amino acids”.
- Load the input sequence of amino acid saved as .txt format.
- Load 5 pdb layers, which we got from modeler.
- Go to “fit” Manu and select “Magic Fit”.
- Save the alignment structure of the Protein.
- Go to “windows” and select alignment.
- The alignment of all pdb file will be showing in a new window and save it.
- Go to tools and select “Energy Minimization” or press Ctrl+N.
- The data of residues, bonds, angels, torsion, and improper, nonBonded, electrostatic, constraint along with total will be shown in new window after processing.

6. Result and Description:-

6.1 Identification of unique metabolic pathways:

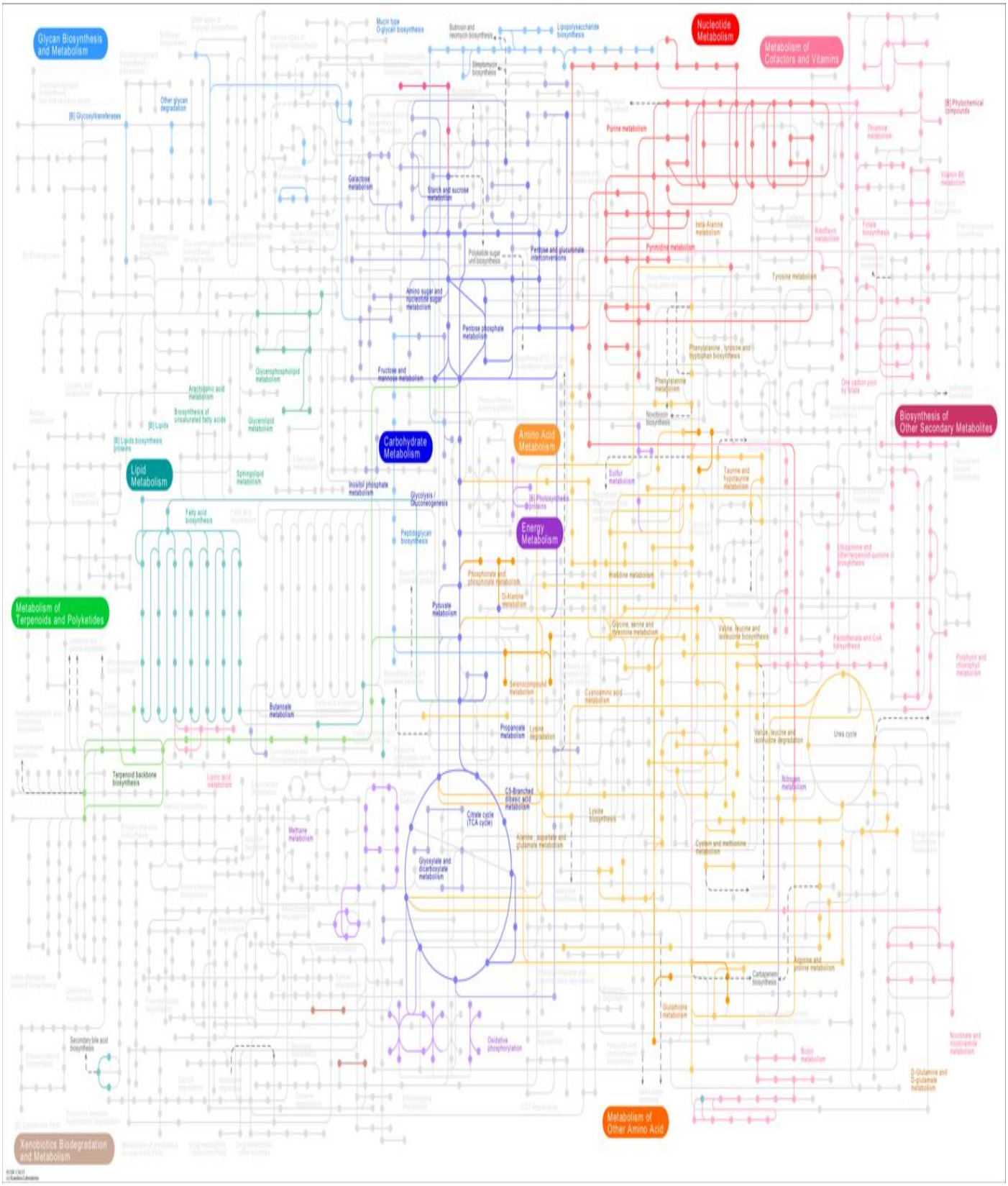


Fig1. (Unique metabolic pathways of bacteroides fragilis 638R)

6.2 Identified unique pathways:

Identified unique pathways are listed below

1 Carbohydrate metabolism

(a) C5-Branched dibasic acid metabolism pathway map.

2 Energy metabolism

(a) Carbon fixation pathway map in prokaryotes.

(b) Methane metabolism pathway map.

3 Lipopolysaccharide biosynthesis pathway map.

4 Peptidoglycan biosynthesis pathway map.

These five unique metabolic pathways are most important for vaccination purposes. Carbon fixation pathway map in prokaryotes is excluded because we are targeting the genes present in the location of cell wall / plasma membrane.

6.3 List of metabolic pathways and figure:

The unique pathway of KEGG website are as follows:-

6.3. 1:- the unique pathway of c5-Branched dibasic acid metabolism:-

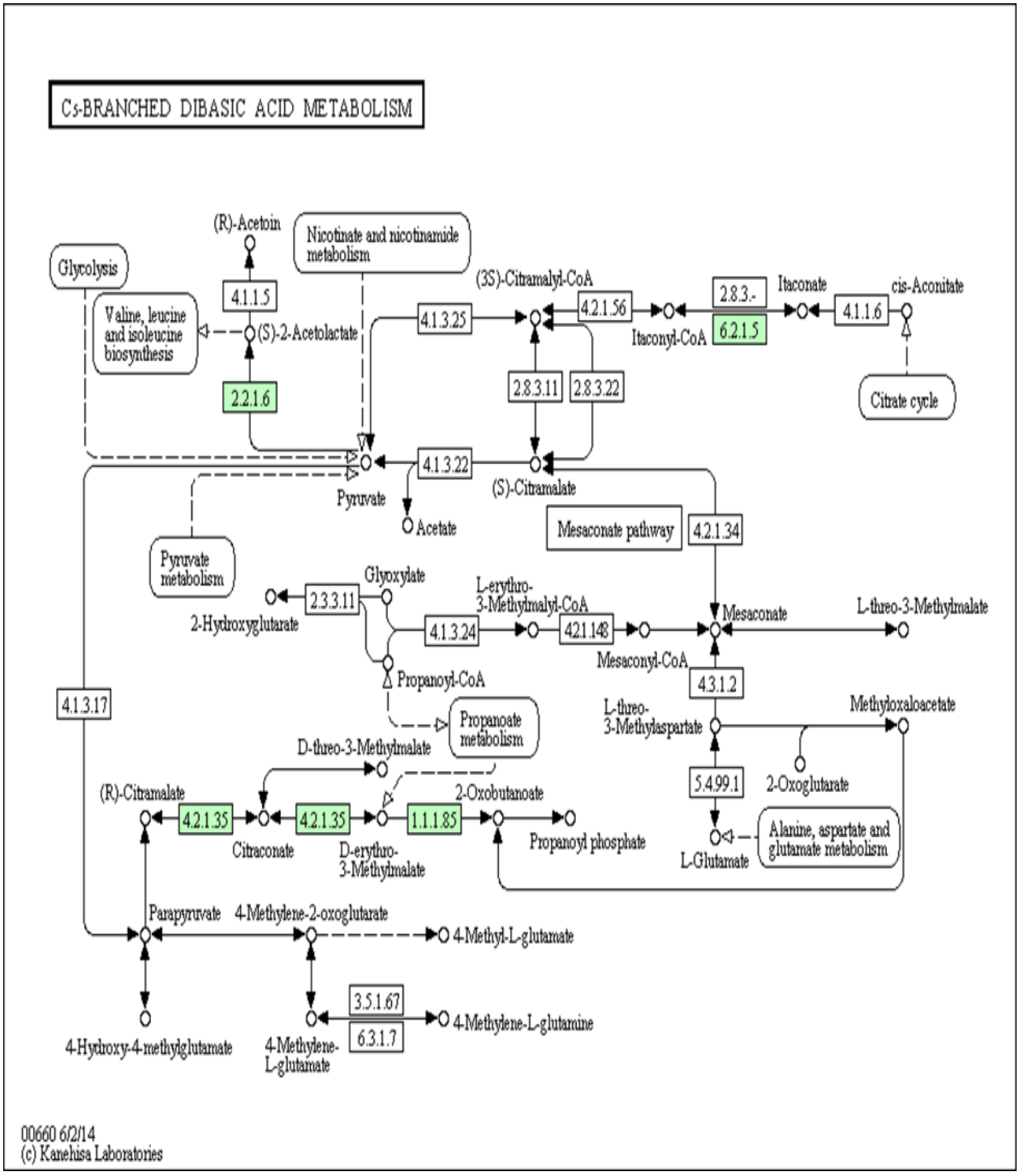


Fig.2 :-(unique pathway of c5-Branched dibasic acid metabolism)

List of gene from c5-Branched dibasic acid metabolism:

1.bfg:BF638R_2360	sucD; Probable succinyl-CoA synthetase (alpha chain); K01902 succinyl-CoA synthetase (alpha subunit) [EC:6.2.1.5]
2.bfg:BF638R_2361	sucC; Probable succinic-CoA syntheses (beta chain); K01903 succinyl-CoA synthetase (beta subunit) [EC:6.2.1.5]
3.bfg:BF638R_2394	Probable putative acetoacetate synthase /II/III (large subunit)[EC:2.2.1.6]
4.bfg:BF638R_3298	leuB; Probable putative 3-isopropylmalate dehydrogenase; K00252 3- isopropylenemalate dehydrogenase [EC:1.1.2.85]
5.bfg:BF638R_3300	leuD; Probable putative 3-isopropylmalate dehydrogenase; K00252 3- isopropylenemalate dehydrogenase
6.bfg:BF638R_3301	leuC; putative 3-isopropylmalate dehydratase large subunit; K01703 3-isopropylmalate/(R)-2-methylmal
7.bfg:BF638R_3576	Probable putative acetohydroxyacid synthase (small subunit); K01653 acetolactate synthase I/III (small subunit)
8.bfg:BF638R_3576	ilvB; Probable putative acetolactate synthase (large subunit); K01652 acetolactate synthase I/ II /III large subu

6.3.2. **The unique pathway of Methane metabolism:-**

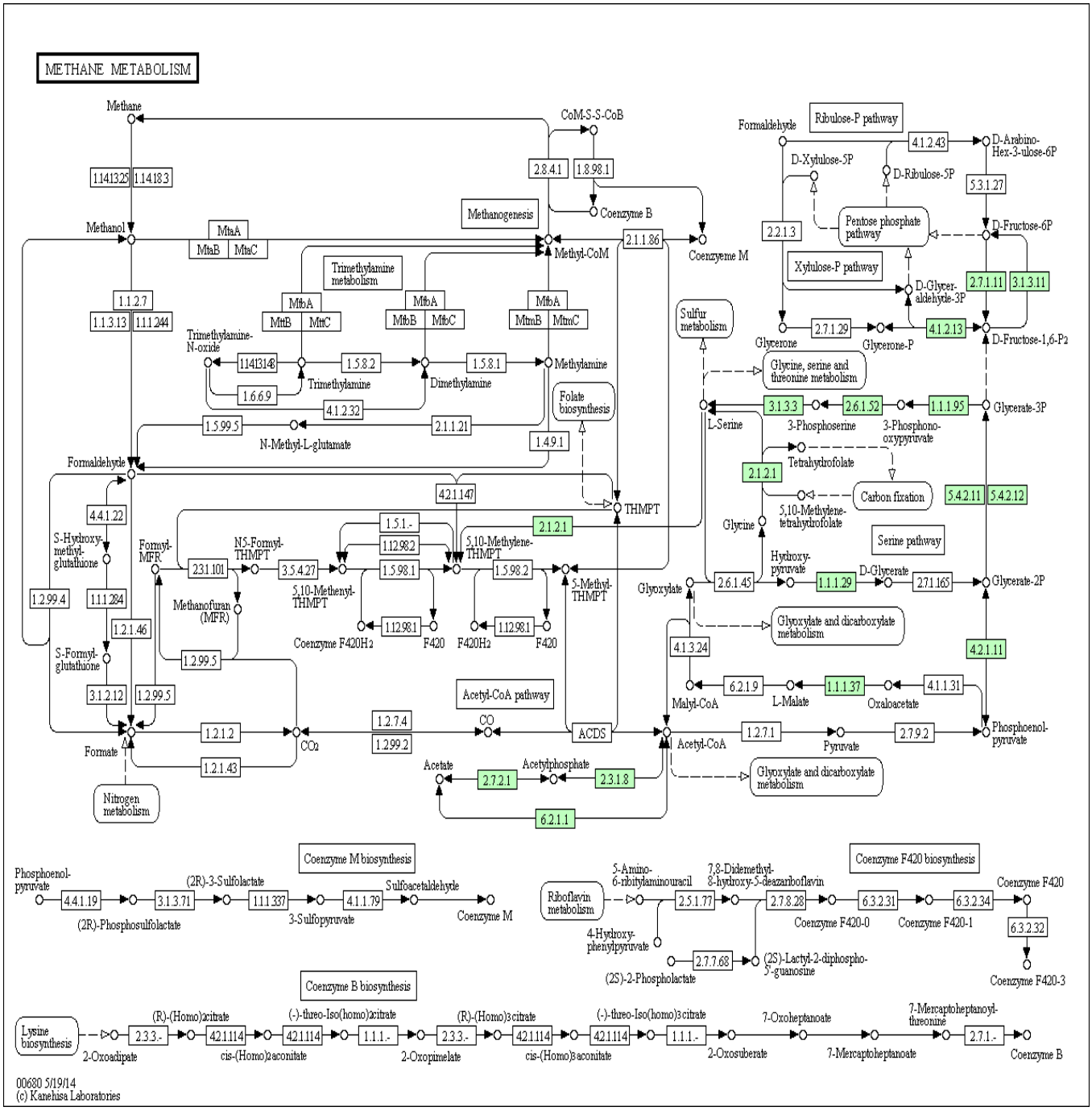


Fig.4:- (The unique pathway of Methane metabolism)

List of gene from Methane metabolism:

1. BF638R_2316	glyA; probable serine hydroxymethyltransferase GlyA[KO:K00600] [EC:2.1.2.1]
2..BF638R_1916	Possible putative glycerate dehydrogenase; K00018 glycerate dehydrogenase [EC:1.1.1.29]

	[KO:K00018] [EC:1.1.1.29]
3..BF638R_1202	eno; probablev putative heme-binding enolase EnO[KO:K01689] [EC:4.2.1.11]
4.BF638R_0537	mdh; probable putative malate-dehydrogenase MdH[KO:K00024] [EC:1.1.2.37]
5.BF638R_3828	mdh; probable putative malate dehydrogenase MdH[KO:K00024] [EC:1.1.2.37]
6.BF638R_3116	fbaB; probable probable putative fructose- aldolase class I FbaB[KO:K11645] [EC:4.1.2.13]
7. BF638R_3162	fba; probable putative fructose bisphosphate aldolase FbA [KO:K01624] [EC:4.1.2.13]
8. BF638R_1853	probable hypothetical protein; K04041 fructose-1,7-bisphosphatase III [EC:3.1.2.11] [KO:K04041] [EC:3.1.2.11]
9. BF638R_3563	pfkA1; probable putative 6-phosphofructokinase 1PfkA1 [KO:K00850] [EC:2.7.1.11]
10. BF638R_0166	pfkA2; probable putative 6-phosphofructokinase 2PfkA2 [KO:K00850] [EC:2.7.1.11]
11. BF638R_0490	ackA; probable putative acetate kinase AckA[KO:K00925] [EC:2.7.2.1]
12. BF638R_0489	ptaA; probable putative phosphate- acetyltransferase PtA[KO:K00625] [EC:2.3.1.8]
13. BF638R_0525	Probable putative Acyl-CoA synthetase; acetyl-CoA synthetase [KO:K01995] [EC:6.2.3.1]
14. BF638R_3117	gpmA; Probable probable putative 6-phosphofructokinase 1 GpmA[KO:K01834] [EC:5.6.2.11]
15. BF638R_0293	gpmI; Probable putative 2,3-bisphosphoglycerate-Independent glycerate mutase GpmL[KO:K15633] [EC:5.4.2.12]
16. BF638R_2333	Possible hypothetical protein; K15634 probable phosphoglycerate mutase [KO:K15634] [EC:5.4.2.12]
17. BF638R_0606	apgM; Probable putative phosphoglycerate mutase ApgM [KO:K15635] [EC:5.4.2.12]
18. BF638R_2144	serA; probable putative 6-phosphofructokinase SerA [KO:K00058] [EC:1.1.1.95]
19. BF638R_2143	serC; Probable putative phosphoserine aminotransferase SerC[KO:K02831] [EC:2.6.1.32]
20. BF638R_2401	Possible putative phosphoserin phosphatase; [KO:K01079] [EC:3.1.3.3]

6.3.3The unique pathway of Lipopolysaccharide biosynthesis metabolism:-

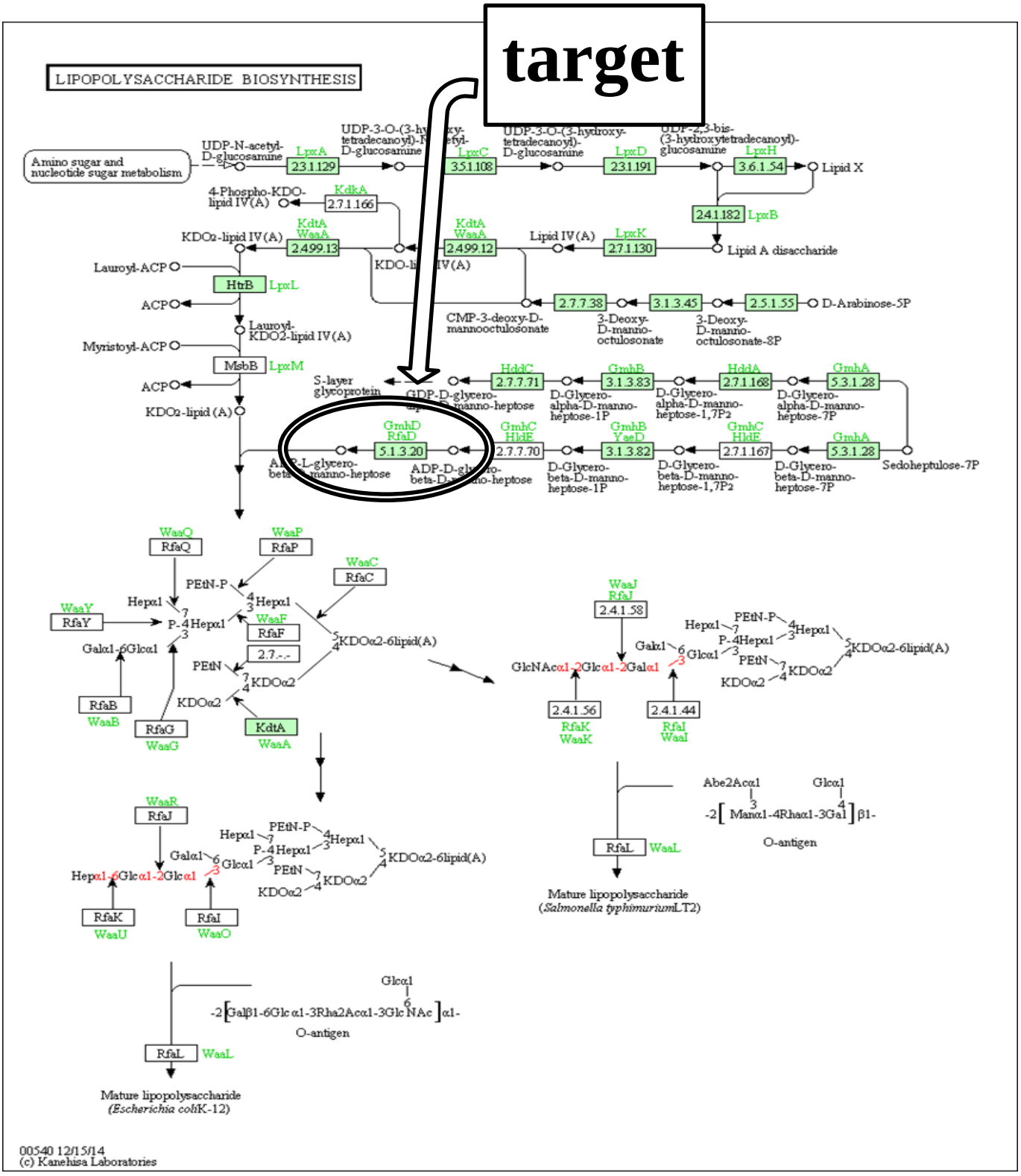


Fig.5:(unique pathway of Lipopolysaccharide biosynthesis metabolism)

List of gene from Lipopolysaccharide biosynthesis metabolism:-

1. BF638R_0141	Possible putative acyl--UDP-N-acetylglucosamine O-acyltransferase; [KO:K00677] [EC:2.3.1.129]
2.BF638R_0886	lpxC; probable UDP-3-O-N-acetylglucosamine deacetylase LoxC[KO:K16363] [EC:4.2.1.59 3.5.1.108]
3. BF638R_0887	lpxD; probable probable putative UDP-3-O-glucosamine N-acyltransferase LpxD[KO:K02536] [EC:2.3.1.191]
4. BF638R_0493	probable tetraacyldisaccharide kinase [KO:K03269] [EC:3.6.1.54]
5. BF638R_0742	lpxB; probable putative lipid A disaccharide- synthase [KO:K00748] [EC:2.4.1.182]
6. BF638R_3307	probable putative tetraacyldisaccharide kinase (lipid A biosynthesis) related protein; [KO:K00912] [EC:2.7.1.130]
7. BF638R_1000	kdsA; probable putative 2-dehydro-5- tetraacyldisaccharide kinase e aldolase KdsA[KO:K01627] [EC:2.5.1.55]
8. BF638R_3146	probable putative lipopolysaccharide biosynthesis-related protein [KO:K03270] [EC:3.1.3.45]
9. BF638R_2334	probable putative lipopolysaccharide biosynthesis; K00979 3-deoxy-manno-sonate cytidylyl transferase) [EC:2.7.7.38] [KO:K00979] [EC:2.7.7.38]
10. BF638R_4066	kdtA; probable 3-deoxy-D-manno-octulosonic-acid transferase KdtA[KO:K02527] [EC:2.5.99.15 2.5.99.14 2.4-5.99.13 2.4.99.12]
11. BF638R_3699	putative lipd A biosynthesis related exported protein; [EC:2.3.1.-] [KO:K02517] [EC:2.3.1.-]
12. BF638R_1445	probable putative D-sedoheptulose 7-phosphate isomerase [EC:5.3.1.28] [KO:K03271] [EC:5.3.1.28]
13. BF638R_1444	Possible probable putative histidine biosynthesis protein; [EC:3.1.3.823.1.3.83] [KO:K03273] [EC:3.1.3.83 3.1.3.82]
14. BF638R_1443	probable putative Nucleoside diphosphate sugar epimerase; K03274 ADP-N-

	glycero-C-manno-heptose 6-epimerase [KO:K03274] [EC:5.1.3.20]
15. BF638R_1442	probable putative GHMP kinase; K07031 D-glycero- -D-manno-heptose-7-phosphate kinase [EC:2.7.1.168] [KO:K07031] [EC:2.7.1.168]
16. BF638R_1446	Possible putative nucleotidyl transferase; D-glycero-alpha-D-manno-heptose 1-phosphate guanylyltransferase [KO:K15669] [EC:2.7.7.71]

6.3.4. The unique pathway of Peptidoglycan biosynthesis:-

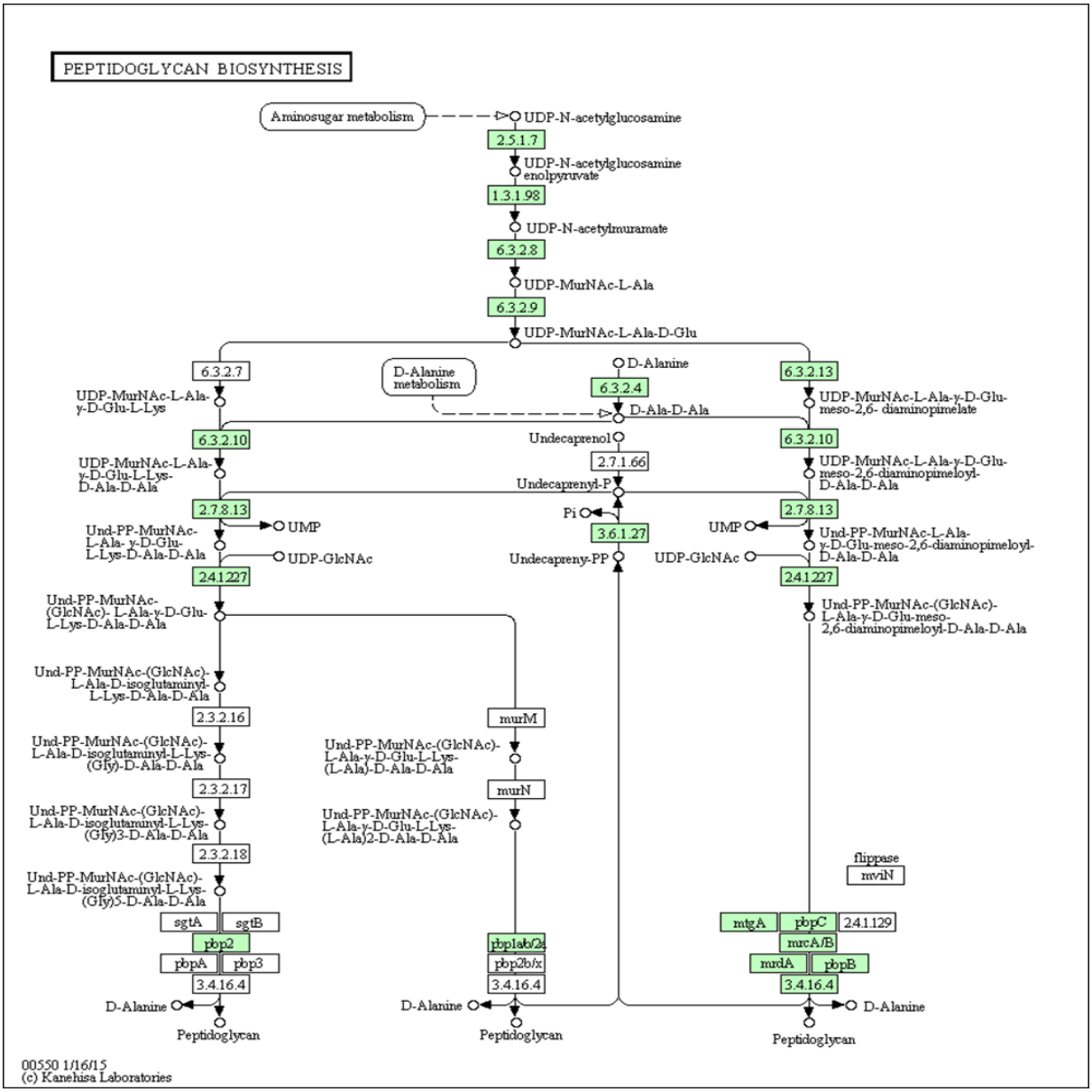


Fig:-6 (The unique pathway of Peptidoglycan biosynthesis)

List of gene from Peptidoglycan biosynthesis:

1. BF638R_3518	Probable putative peptidoglycan biosynthesis related protein;[KO:K00790] [EC:2.5.1.7]
2. BF638R_2870	Possible UDP-N-acetylmuramate dehydrogenase [KO:K00075] [EC:1.3.1.98]
3. BF638R_0305	murC; probable putative UDP-N-acetylmuramate-alanine ligase MurC[KO:K01924] [EC:6.3.2.8]
4. BF638R_0308	murD; probable putative UDP-N-acetylmuramoylalanine-D--glutamate ligase MurD[KO:K01925] [EC:6.3.2.9]
5. BF638R_0310	murE; possible UDP-N-acetylmuramoylalany glutamate--2,6-diaminopimel ligase MurE[KO:K01928] [EC:6.3.2.13]
6. BF638R_0500	ddlA; possible putative D-Alanine-D-Alanine-Ligase DdlA[KO:K01921] [EC:6.3.2.4]
7. BF638R_0457	murF; probable putative UDP-N-acetylmuramoyl- --D-alanyl-d-alanine ligase [KO:K01929] [EC:6.3.2.10]
8. BF638R_0054	possible putative undecaprenol kinase; [KO:K06153] [EC:3.6.1.27]
9. BF638R_0309	mraY; possible putative undecaprenyl-phosphate-phosphatase [KO:K01000] [EC:2.7.8.13]
10. BF638R_0306	murG; possible monofunctional peptidoglycan transglycosylase - undecaprenol N-acetylglucosamate transferase MurG[KO:K02563] [EC:2.4.1.227]
11. BF638R_1930	Possible putative monofunctional peptidoglycan transglycosylase; [KO: K03814] [EC: 2.4.1.-]
12. BF638R_1418	Possible putative penicillin-binding protein [KO:K05367] [EC: 2.4.1.-]
13. BF638R_2336	Possible Putative penicillin-binding peptidoglycan biosynthesis-related exported protein; [EC: 2.4.1. - 3.4.-.-] [KO:K05366] [EC:3.4.-.- 2.4.1.-]

14. BF638R_3952	mrdA; putative penicillin-binding protein 2 [KO:K05515]
15. BF638R_0311	ftsI; putative cell division specific transpeptidase/penicillin-bindingprotein[KO:K03587]
16. BF638R_0024	dacB; Possible putative exported D-alanyl-D-alanine carboxypeptidase penicillin-binding protein [KO:K07259] [EC:3.4.21.- 3.4.16.4]

6.4. SELECTION OF ESSENTIAL GENES:

The essential genes are short out through the DEG database. The parameters that is being utilized to short out the genes is demonstrated as follows, score greater than 100(i.e. 500), expected value is $1 \times e0.5$. This is taken because of specificity of the gene toward the metabolism process.

List of Essential Genes are:

After submitting all the amino acid sequence of the genes of selected four pathways are retrieved from the DEG database or through mail. All the essential genes are shown below.

Gene Entry	Gene Name
1. BF638R_3577	ilvB
2. BF638R_2360	sucD
3. BF638R_2361	sucC
4. BF638R_2594	Not available
5. BF638R_3298	leuB
6. BF638R_3300	leuD
7. BF638R_3301	leuC
8. BF638R_3576	Not available

9. BF638R_0293	gpmI
10. BF638R_0490	ackA
11. BF638R_0525	Not available
12. BF638R_0537	mdh
13. BF638R_0606	apgM
14. BF638R_1202	eno
15. BF638R_1853	Not available
16. BF638R_2143	serC
17. BF638R_2316	glyA
18. BF638R_2333	Not available
19. BF638R_2401	Not available
20. BF638R_3116	fbaB
21. BF638R_3117	gpma
22. BF638R_3162	fba
23. BF638R_3828	mdh
24. BF638R_1442	Not available
25. BF638R_1443	Not available
26. BF638R_1444	Not available
27. BF638R_3952	mrdA
28. BF638R_1446	Not available
29. BF638R_1445	Not available
30. BF638R_0024	DacB

6.5 Identification of non homologous genes:-

The shortlisted 30 essential genes of B.fragilis 638R subjected to BLASTp against human genome.The threshold value was situated to 0.005 and identity ought to be less than 0.35 (35%).

After submitting the AA seq in the box provide, setting the parameters as shown above, then tap on BLAST. A page will open following couple of second. This page will demonstrate the information and identity under the description segment. Detail of indication data and the most significant record configuration of the gene can be accomplished by denoting the gene in the given box then clicking on the download or GenPept, graphical perspective for Graphics. We can decrease the number of column by clicking on the setting symbol on top of right corner.

All the BLAST result is listed below with their biological process and EC number:-

Sl No.	Accession no.and gene name	Location in cell and cellular components	Can be used as drug or not	Biological process	Enzyme commission number
1.	BF638R_3577 :ilvB	Cytoplasm	No	acetolactate synthase activity	2.2.1.6
2.	BF638R_2360 :sucD	Periplasm, Outer Membrane, Extracellular	No	ATP citrate synthase activity	6.2.1.5
3.	BF638R_2361 :sucC	Cytoplasm, plasma Membrane, Periplasm	No	tricarboxylic acid cycle	6.2.1.5
4.	BF638R_2594	Cytoplasm, Plasma Membrane, Periplasm	yes	Thiamine pyrophosphate	2.2.1.6
5.	BF638R_3298	Cytoplasm	yes	Amino-acid biosynthesis; L-leucine biosynthesis;	1.1.1.85
6.	BF638R_3300 :leuD	Cytoplasm, plasma Membrane ,Periplasm	No	Amino-acid biosynthesis, Leucine biosynthesis	4.2.1.33 4.2.1.35
7.	BF638R_3301 :leuC	plasma Membrane ,Periplasm	yes	Amino-acid biosynthesis	4.2.1.33 4.2.1.35
8.	BF638R_3576	Cytoplasm, plasma Membrane ,Periplasm	yes	branched-chain amino acid biosynthetic process	2.2.1.6

9.	BF638R_0293 :gpmI	Cytoplasm, plasma Membrane ,Periplasm	No	Carbohydrate metabolism	5.4.2.12
10.	BF638R_0490 :ackA	Cytoplasm, plasma Membrane ,Periplasm	No	Amino-acid biosynthesis, , glycine biosynthetic process from serine,	2.7.2.1
11.	BF638R_0525	Cytoplasm, plasma Membrane ,Periplasm	yes	Not know	6.2.1.1
12.	BF638R_0537 :mdh	Plasma Membrane, Periplasmic ,Extracellular	yes	cellular carbohydrate metabolic process process,tricarboxylic acid cycle	1.1.1.37
13	BF638R_0606 :apgM	Cytoplasm, Plasma Membrane , Periplasm,OuterMembrane, Extracellular	yes	phosphoglycerate mutase activity	5.4.2.12
14	BF638R_1202 :eno	Cytoplasm	No	cellular carbohydrate metabolic process process,tricarboxylic acid cycle	4.2.1.11
15.	BF638R_1853	Cytoplasm,plasma Membrane , Periplasm, Outer Membrane , Extracellular	No	Carbohydrate metabolism	3.1.3.11
16	BF638R_2143 :serC	Cytoplasm, plasma Membrane ,Periplasm	No	Amino-acid biosynthesis, Pyridoxine biosynthesis, Serine biosynthesis	2.6.1.52
17.	BF638R_2316 :glyA	Cytoplasm, plasma Membrane ,Periplasm	No	Amino-acid biosynthesis, , glycine biosynthetic process from serine,	2.1.2.1
18.	BF638R_2333	Cytoplasm, plasma Membrane ,Periplasm, Outer Membrane , Extracellular	yes	Not known	5.4.2.12
19.	BF638R_2401	Cytoplasm, plasma Membrane ,Periplasm	Close to 35%	L-serine biosynthetic process	3.1.3.3

			No		
20	BF638R_3116 :fbaB	Cytoplasm, plasma Membrane ,Periplasm Outer Membrane, Extracellular	yes	Not known	4.1.2.13
21.	BF638R_3117 :gpmA	Outer Membrane, Extracellular	Close to 35% No	glycolytic process, Glycolysis	5.4.2.11
22.	BF638R_3162 :fba	Cytoplasm	No	fructose 1,6-bisphosphate metabolic process, glycolytic process	4.1.2.13
23.	BF638R_3828 :mdh	Cytoplasm, plasma Membrane ,Periplasm Outer Membrane, Extracellular	yes	Carbohydrate metabolism	1.1.1.37
25.	BF638R_1442	Cell wall , Extracellular	yes	Not known	2.7.1.168
26.	BF638R_1443	Cytoplasm, plasma Membrane ,Periplasm	yes	carbohydrate metabolic process	5.1.3.20
27.	BF638R_1444	Cytoplasm, plasma	yes	Carbohydrate metabolism	3.1.3.82 3.1.3.83
28.	BF638R_3146	Cytoplasm, plasma Membrane ,Periplasm	Close to 35% yes	lipopolysaccharide biosynthetic process	3.1.3.45
29.	BF638R_1445	Cytoplasm, plasma Membrane , Periplasm	Close to 35% No	carbohydrate metabolic process	5.3.1.28
30.	BF638R_0024 :dacB	Cytoplasm, plasma Membrane ,Periplasm, Outer Membrane , Extracellular	yes	carboxypeptidase activity	3.4.16.4 3.4.21.-

6.6 Result of homology modeling:-

Three-dimensional structures will help in the visualization of the binding sites and may prompt the design of potential drug. The 3D structure of DNA BF638R_1443 protein of the B.fragilis was modeled with Deep View; Swiss model; Modeler 9.14 was used for fine building the model and global energy minimization. Modeler results are described in details bellow

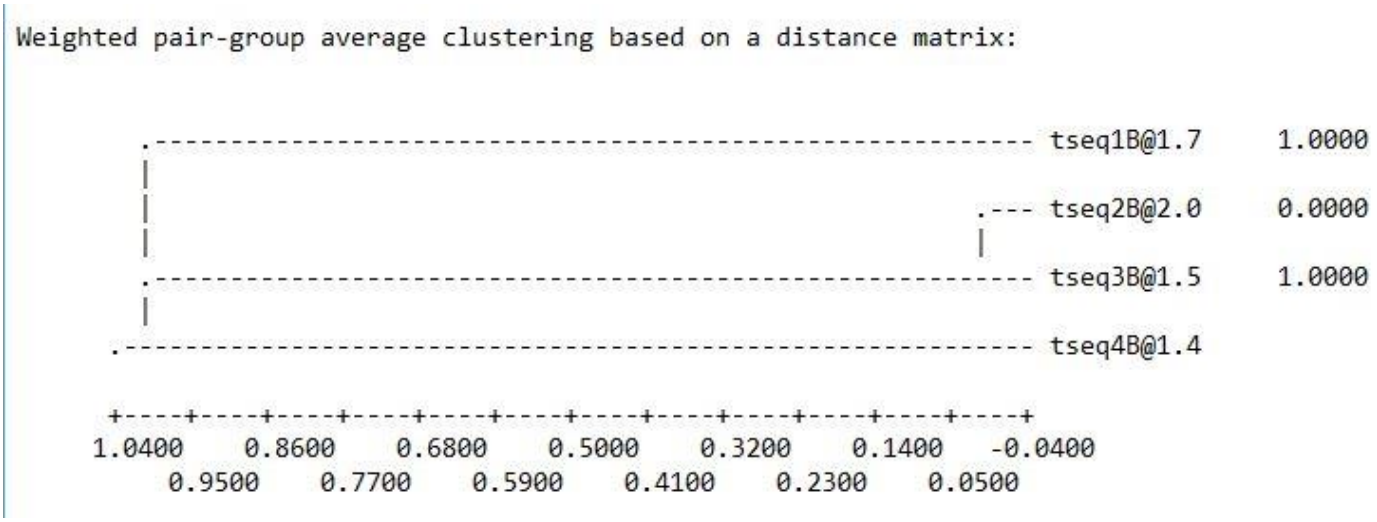


Fig 7 :Pairwise distance matrix of clustering tree

The first four best result from NCBI BLASTP results are named as tseq1, 2, 3&4. There are five templet pdb files are generated and the best model is selected on the basic of DOPE score. Dope score are used to predict the most stable protein templet. If greater the DOPE score ,less is the stability and less is the rank.

Rank of five protein templet are listed below on the basic of DOPE score:

File name(pdb)	Identity (%)	DOPE score	Rank
B99990001	19.811001	-31487.11719	4
B99990002	19.811001	-32676.75195	1
B99990003	19.811001	-32173.77930	2
B99990004	26.127502	-32162.12500	3
B99990005	17.510044	-31393.78711	5

List of energy parameters and values of protein from Modeler:

Parameters	score
% sequence identity	19.811001
Sequence length	318
Compactness	0.325869
Native energy (pair)	-15.293428
Native energy (surface)	-2.364290Jul
Native energy (combined)	-0.810227Jul
Z score (pair)	-3.768817
Z score (surface)	-5.067099
Z score (combined)	-5.785070
GA341 score	0.965544
Total DOPE score	-32676.75195

The three-dimensional surface structure of B99990002 is visualized using RasWin software. We can calculate the number of atom present in each side chain.

6.6.1Structure of the gene: “BF638R_1443” :-

Surface structure of B99990002 protein templet

. The structure of the target protein is basically similar with template if both template and target seq are Similar. Surface structure using Pymol software (3-D Surface structure of B99990002 protein templet)

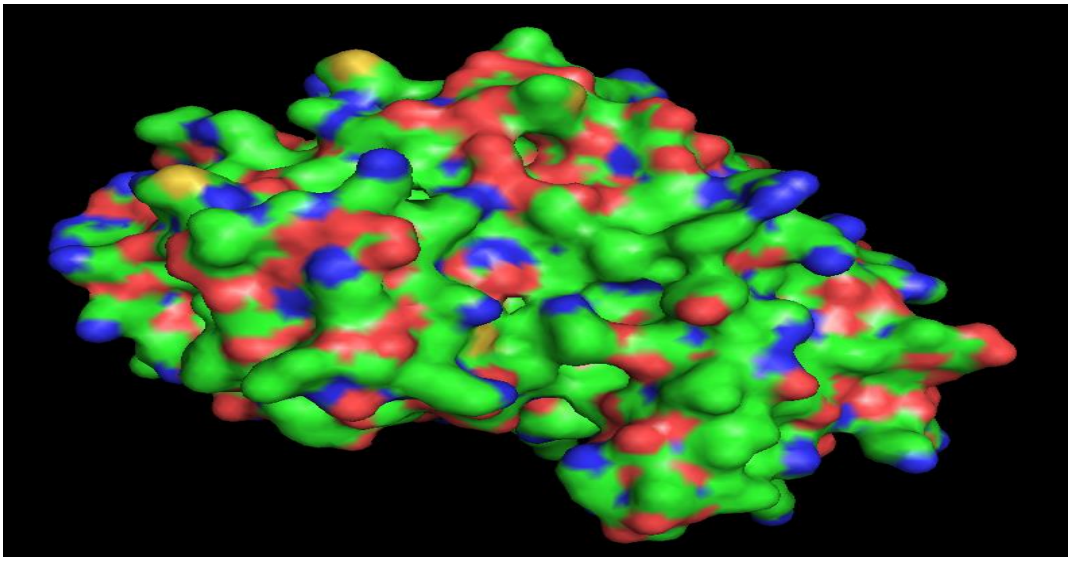


Fig 8 (3-D Surface structure of B99990002 protein templet)

Result of Ramachandran plot

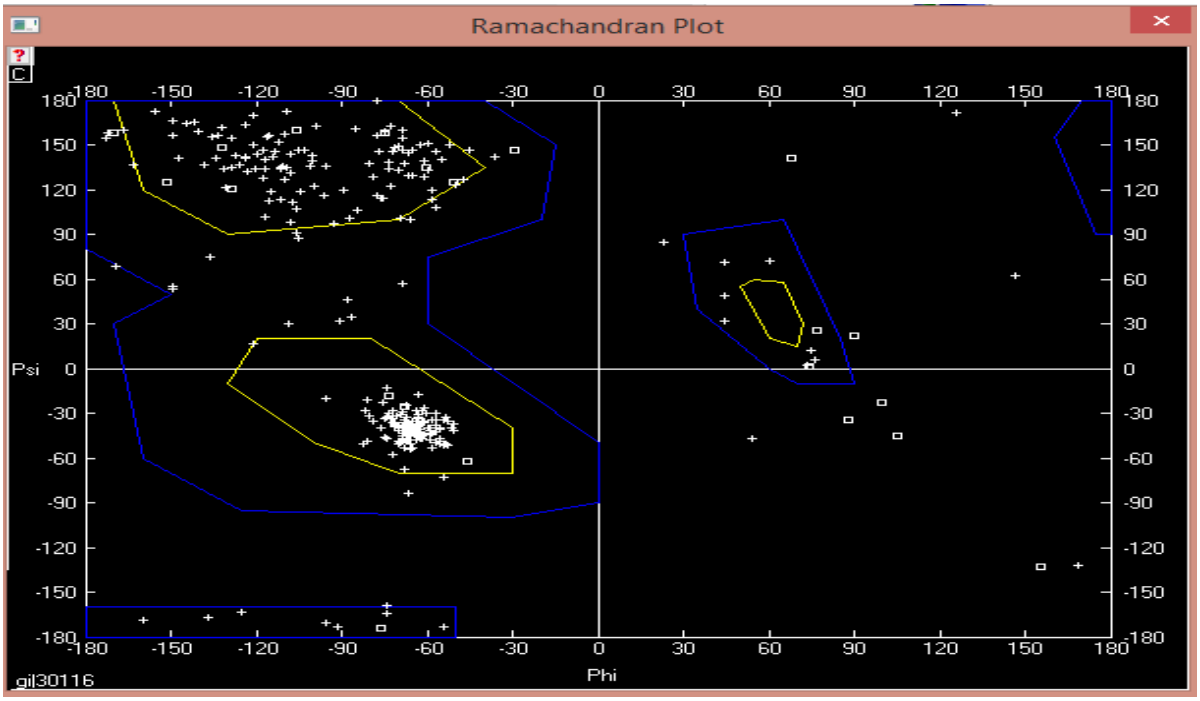


Fig. 9 Ramachandran plot of Holliday junction DNA helicase

6.7 DISCUSSION:-

Among all the human non-homologous essential genes **BF638R_1443**, having EC no: 5.1.3.20 is showing best Blast P result. The NCBI Gene ID of **BF638R_1443** is **11702465**, putative Nucleoside diphosphate sugar epimerase. This gene is present in the Cytoplasm and involves in the biological process like Carbohydrate metabolism process, **Lipopolysaccharide biosynthesis.** The catalytic activity of this protein is: ADP-glyceromanno-heptode 6-epimerase activity. The gene **BF638R_1443** is having 318 amino acid sequence and NCBI-GI is **WP_014298519.1**. The AA (amino acid) sequence of this gene is (Active site Prediction):-

10	20	30	40	50
MAYIVTGGAG	FVGSNMVKKL	NDKGINDVII	IDTYSDDKMK	NLIGLKFIDF
60	70	80	90	100
VDYQDGIKTV	VDYLKAIKNP	QAVFHIGANA	DVLVYDVKKM	MNENYEFSKM
110	120	130	140	150
YCEFANHRNI	PFIYASSSAV	YGNGGKQEVG	AGNEEPHNTY	AWSKWLFDQY
160	170	180	190	200
VMANSSKFLN	KVMGFRFFNV	FGWGEFHKGK	NANIVYRFYR	FIKEKNFIDL
210	220	230	240	250
FNEEIVRDHI	WVEDVAEVMY	QAMIRKELET	GIYNLGGMHP	ISHRQVADIV
260	270	280	290	300
INTMIEEGII	PQDITDRYIT	LIDMPEELRS	KFQFYTHADN	QLTFISEIAK
310				
GNDLKMVEYV KRLIRENK				

The result is taken from Uniport.org

This can be taken as drug target protein, because:-

1. The 3D structure of the protein is known.
2. It has given the best BLASTP result.
3. This gene is liable for the regulation of cell shape and cell wall organization. So, if gene is focused on the wall will disrupt and cell lysis will occur.

4. Bacteria will not be able to go to inactive stage, because this gene is responsible for the cell wall organization.

6.7.1 Final structure of gene BF638R_1443
(Modeled protein ribbon structure using Pymol software):

Fig. 10 ("Modeled protein ribbon structure using Pymol software")

7. Conclusion:-

All the genes of the four important metabolism pathways which are not present in the human are taken as important pathways. The genes responsible for these unique pathways are analyzed for their essentiality and listed in the metabolism procedure of the bacteria through the Data base of essential genes tool. All the shorted genes from the DEG are being shorted for the non-homologous for the human being through the NCBI BLASTp, , as BLASTp takes the amino corrosive grouping which is in charge of the protein exhibit in the bacteria. After performing above steps we have been able to find a target named as putative Nucleoside diphosphate sugar epimerase (**BF638R_1443**), which are present in the Cytoplasm and involves in the biological process like Carbohydrate metabolism process, Lipopolysaccharide biosynthesis**.** This genes capacity can be halted then the microorganisms can be murdered and Multiple Drugs Resistant TB can be cured**.** The availability of computer-aided software and full genome sequences likemodeler, Autodock help to identify probable antimicrobial drug targets. We present here a point by point in-silico analysis of crucial genes, Molecular Modeling of the target protein.

The future study of Bacteroides fragilis 638R project is to find a certain molecule which can be able to drug target this gene and the MDR TB (Multiple Drugs Resistant) can be cured.

8. Reference:-

1. Science reporter, April/2014
2. Ferreira et al-2009 E O Ferreira, C J Smith, L.A. Lobo, R.J.M. Peixoto C.J. Smith, E.R.A Rocha, R.M.P. Domingues, R.B. Zingali. R.M. Peixoto
3. Lesk, M (2005), Bioinformatics introduction, 2nd edition. Oxford University
4. Casey, R. M. (2005). "Basic local alignment search tool Sequences Aid in Genomics and Proteomics ". Business Intelligence Network.
5. Huynen, M., Dandekar, T. and Bork, P. (1998). Differential genome analysis applied to the species-specific features of Helicobacter pylori. FEBS Lett. 426,1-5.
6. Kanehisa M, Goto S (2000). "KEGG: Kyoto Encyclopedia of Genes and Genomes". Nucleic Acids Res 28 (1): 27–30
7. International Union of Tuberculosis and Lung Disease/World Health Organisation. Statement: tuberculosis preventive therapy in HIV-infected individuals. Tuber Lung Disease 1994;75:96–8.
8. JACOBUS, NV. "ACTIVITY OF TROSPECTOMYCIN AGAINST BACTEROIDES-FRAGILIS AND OTHER BACTEROIDES SPECIES." Antimicrobial agents and chemotherapy 32.4 (1988):584-586.
9. Sakharkar, K. R., Sakharkar, M. K. and Chow, V. T. K., (2004). A novel genomics
10. methods for the selection of drug targets in pathogens, with special reference to Pseudomones aerugenosa. In-Silico Biol 4, 0028.
11. Allsop, A. E. (1998). New antibiotic discovery, novel screens, novel targets and impactof microbial genomics. Curr. Opin. Microbiol. 1, 530-534.
12. Judson, N. and Mekalanos, J. J. (2000b). Transposon-based approaches to identifyessential bacterial genes. Trends Microbiol. 8, 521-526.
13. Moir, D.T., Shaw ,K.J., Hare, R.S., Vovis, G.F.,(1999) Genomics and antimicrobialdrug discovery. Antimicrob Agents Chemother. 43:439-446.
14. Huynen, M., Y. & Bork, P. (1997). Differential genome display. Trends Genet. , 389-390.

15. Huynen, M., Dandekar, T. and Bork, Zhang ,R.and Zhang, (2004). Database of Essential Genes.NA Research. 32, 358-386

16. Zhang ,R.and Zhang, (2004). Database of Essential Genes.NA Research. 32, 271-272.

17. Rost, B. (1999).zone of protein sequence Matching. Protein Eng. 12, 85-94.

18. Ruppert, J., Zhang ,R.and Zhang, (2004). Database of Essential Genes.NA Research. 32, 271-272.

19. Connolly, M. L., (1983). Analytical molecular surface identify. Journal of Appl.Crystallogr. 16,548. 1. A. E. Curr. (1998). New antibiotic discovery, novel screens, novel targets and impact of microbial genomics. Opin Microbiol. 1, 530-534.

20. Moxon ER: R., S.R. Peri, and A.R. Allam, *Computational and function prediction of hypothetical protein.* Computat Bioinfo and In Silico Model, 2012. **1**(5): p. 58-62.

21. Morens Fauci AS: East, H., et al., *About 2nd causes of mellitus.* Journal of the Mississippi State Medical Association, 2012. **53**(11): p. 380-383.

22. SCJ Lazaro E: Ampicillin-resistant non-beta-lactamase-producing: recent emergence of clonal isolates with increased resistance to cefotaxime and cefixime. Agents Antimicrob Chemother 2007, 51:2564-257

23. East, H., et al., *About 2nd causes of mellitus.* Journal of the Mississippi State Medical Association, 2012. **53**(11): p. 380-383.

24. Adisakwattana, S. and B. Chanathong, *Alpha-glucosidase inhibitory activity and lipid-lowering mechanisms of Moringa oleifera leaf extract.* Eur Rev Med Pharmacol Sci, 2011. **15**(7): p. 803-8.

25. Rose, P.W., et al., *The RCSB Protein Data Bank: redesigned web site and web services.* Nucleic acids research, 2011. **39**(suppl 1): p. D392-D401.

26. Rambabu, R., S.R. Peri, and A.R. Allam, *Computational and function prediction of hypothetical protein.* Computat Bioinfo and In Silico Model, 2012. **1**(5): p. 58-62.

27. Cui, X., A. Yeliseev, and R. Liu, , *binding site and G protein activation of the mu opioid receptor.* European journal of pharmacology, 2013. **702**(1): p. 309-315.

28. Consortium, U., *Update on activities at the Universal Ligand interaction Protein Resource (UniProt) in 2013.* Nucleic acids research, 2013. **41**(D1): p. D43-D47.

29. Colovos, C., Yeates, T.O.(1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2,1511-1519.

30. Bowman, G.D., ODonnell, M. and Kuriyan, J. (2004) Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature. 429,724-730.

31. Guex, N., Diemand, A. and Peitsch, M.C. (1999). Protein modelling for all. TiBS.24,364-367.

32. Rost, B. (1999). Twilight zone of protein sequence alignments. Protein Eng. 12, 85-94.

33. Isar, J, andISAR. "Suceinic acid production from Bacteroides fragilis: Process optimization and scale up in a bioreactor." Anaerobe 12.5-6 (2006):231-237.