International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 03 | June-2015 www.irjet.net p-ISSN: 2395-0072

SUBTRACTIVE GENOMICS – A Promising way To Combat Pathogens

(A Review)

Pratheek J Madabhavi[1*], V G Shanmuga priya[2], Rakesh N R[3], Preeti S Honagudi[4], Surekha Jiddagi[5]

[1, 5] *Department of Biotechnology, KLE Dr.M.S.Shesgiri College of Engineering and Technology, Belgaum-8, Karnataka, India*

[2,3,4] *R & D Center, Department of Biotechnology, KLE Dr.M.S.Shesgiri College of Engineering and Technology, Belgaum-8, Karnataka, India*

Abstract - *Subtractive genomics is the modern approach which is employed to identify new drug targets in pathogenic organism. This approach involves subtraction of proteins or genes between the host and parasite and provides information of a set of proteins that are likely to be essential to the parasite but absent in the host. A review was carried out on subtractive genomics approach on five pathogenic organisms which are categorized under three different cases. Salmonella typhi, Listeria monocytogenes which are multiple drug resistant organisms and the organisms Lieshmania Donovani and Clostridium botulinum which have no effective drugs till date, and Streptococcus Suis, the organism for which no virulence factors was found till date were considered. By subtractive genomics approach the number of proteins to be used as drug targets was reduced to minimum compared to total proteome. Studies shows that in all cases outer membrane proteins are considered for putative drug targets.*

Key Words: Subtractive genomics, Drug targets

1. Introduction

Currently the drugs used in treatment of diseases caused by pathogenic organisms show small to large side effects in patients and there is an wide alarming rise in the evolution of drug resistance strains. Hence there is a need for identifying new and effective drug drugs to combat the diseases. A modern approach called "Subtractive genomics" is currently widely engaged to identify novel and specific drug targets in pathogenic organism, as a step towards identifying novel and potential drugs. In this study, a review was carried out on subtractive genomics approach on five pathogenic organisms which are categorized under three different cases. *Salmonella typhi, Listeria monocytogenes* which are multiple drug resistant organisms and the organisms *Lieshmania donovani* and *Clostridium botulinum* which have no effective drugs till date, and *Streptococcus suis,* the organism for which no virulence factors was found till date were considered.

2. Subtractive genomics approach:

Subtractive genomics is the process in which the subtraction of sequences between host and the pathogen proteome which helps in providing information for a set of proteins which are essential to pathogen but are not present in the host. Subtractive genomics plays an important role in identification of potential drug targets. These targets are those proteins which are considered essential for the survival of the organismm. [2]

Subtractive genomics studies applied in 3 types of cases:

i. Multi-Drug resistant pathogenic organism:

Salmonella enterica serovar typhi is a human-specific gram-negative pathogen causing enteric typhoid fever, a severe infection of the Reticuloendothelial system. Infection of *S. typhi* leads to the development of typhoid, or enteric fever. This disease is characterized by the sudden onset of a sustained and systemic fever, severe headache, nausea, and loss of appetite. [10] Worldwide, typhoid fever affects roughly millions of people annually, causing deaths. The early administration of antibiotic treatment has proven to be highly effective in eliminating infections, but indiscriminate use of antibiotics has led to the emergence of multidrug-resistant strains of *S. enterica serovar Typhi.*

Listeriosis, the infectious disease caused by *Listeria monocytogenes* causes serious localized and generalized infections in humans especially among pregnant women, the elderly or individuals with a weakened immune system. In serious cases, it can lead to brain infection and even death. Several symptoms are ranged from flu-like illness to severe complications including meningitis,

septicemia, spontaneous abortion or listeriosis of the newborn (FAO/WHO, 2002). The prolonged and uncontrolled use of antibiotics in treatment against many pathogens has caused the multiple drug resistance. [3]

ii. Pathogenic organisms with No effective drugs available:

The *Leishmania donovani is* an intracellular parasite causing kala azar disease, which is almost always fatal if left untreated. As there is no effective medicine available so far, leishmaniasis infection is a worldwide public health challenge. [7]

Botulism (Latin, botulus, "sausage") also known as botulinus intoxication is a rare but serious paralytic illness caused by botulinum toxin, which is produced by the bacterium *Clostridium botulinum.* Severe botulism leads to reduced movement of the muscles of respiration, and this may be experienced as dyspnea (difficulty breathing), but when severe can lead to respiratory failure which may lead to coma and eventually death if untreated. Clostridium botulinum would normally be harmless to humans, but it can infect by a virus, the viral DNA gets integrated into the bacterial genome, causes the host to produce toxins. Neurotoxin production is the unifying feature of the species C. botulinum. Seven types of toxins have been identified and allocated a letter (A-G). Organisms genetically identified as other Clostridium species have caused human botulism, Clostridium butyricum producing type E toxin and Clostridium baratii producing type F toxin. The ability of C. botulinum to naturally transfer neurotoxin genes to other clostridia is concerning, especially in the food industry where preservation systems are designed to destroy or inhibit only *Clostridium botulinum* but no other Clostridium species. [5]

iii. Pathogenic organism with no virulence factor identified:

Streptococcus Suis are a Gram-positive, facultative anaerobe coccus, possessing cell wall antigenic determinants. *S. suis* is responsible for a wide variety of porcine disease syndromes, such as meningitis, septicae.mia, arthritis and endocarditis.It has more than 35 serotype which are responsible in the cause of serious infections in humans. [4]Since no critical virulence factor is well known in the pathogen they applied in silico subtractive proteomics approach to identify potent drug targets in S. suis so that new drugs can be tailored to resist this zoonotic emerging disease.

3. Studies carried out on pathogenic organisms:

Rajendra H M et al (2010) has retrieved the complete proteome sequences of *L.donovani* and *H. sapiens.* Protein sequence of *L. donovani* was searched for sequence homology with human proteome using BLAST program (bit score cut off <100 and minimum expectation value (*E*-value) cut off E-10 were taken to identify homology exhibiting significant with their human counterpart.).Proteins sequences less than 100 amino acids in length were excluded from the analysis which was unlikely to represent essential to parasite hence such sequences were excluded from analysis. Human homologs proteins were then searched against DEG For short listing essential proteins, bit score cut off >100 and *E*-value <E-10 were considered. The function and sub cellular localization of each non homologous protein is identified by using online sub cellular localization prediction tools, CELLO Membrane localized proteins were identified and listed as putative candidate drug targets.[11]

Bhawna Rathi et al (2009) has retrieved the complete proteome of *Salmonella typhi* from SwissProt and protein sequences of *Homo sapiens* were downloaded from NCBI .The Database of Essential genes was accessed and *S. typhi* proteins were purged at 60% using CD-HIT to identify the paralogs or duplicates proteins within the proteome of *S.typhi*. The paralogs are excluded and the remaining sets of protein were subjected to BlastP against *Homo sapiens* protein sequences with the expectation value (E-value) cutoff of 10-4. BLASTP analysis was performed for the non homologous protein sequences of *S.typhi* against DEG with E-value cutoff score of 10-100. A minimum bit score cut-off of 100 was used to screen out genes that appeared to represent essential genes. The protein sequences obtained are non homologous essential proteins of *S.typhi*. Metabolic pathway analysis of the essential proteins of S. typhi was done by KAAS server at KEGG for the identification of potential targets. KAAS (KEGG Automatic Annotation Server) provides functional annotation of genes by BLAST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG Pathways. Sub-cellular localization analysis of the essential protein sequences has been done by Proteome Analyst Specialized Sub cellular Localization Server v2.5 (PA-SUB) to identify the surface membrane proteins which could be probable vaccine candidates. [1]

G. Koteswara Reddy et al(2011) has retrieved and downloaded The complete genome and protein sequences of Clostridium botulinum A strain from the National Center for the Biotechnology Information (NCBI) server (ftp://ftp.ncbi.nlm.nih.gov/genomes/) .Coding sequences having less than 100 amino acids were screened out because Coding sequences having less than 100 amino acids were less likely to represent essential genes from protein table (www.ncbi.nlm.nih.gov/sites/entrez).These coding sequences were subjected to BLASTX against the DEG database (http://tubic.tju.edu.cn/deg/).Expectation value (E-value) cut-off of 0.00001 was used to screen out coding sequences that are likely to be essential. Remaining coding sequences were subjected to BLASTX against

human genome provided by the NCBI server (http://www.ncbi.nlm.nih.gov/) with default parameters to find out essential and non human homologs. The homologs to human genome were excluded; the essential, non-human homologs were listed out. Among the essential, non-human homolog coding sequences, their functional elements i.e. enzymes were listed out because they are potential drug targets. The protein products corresponding to the final selected genes were further analyzed with the database of protein sub cellular localization in bacteria (http://db.psort.org/) to compile the final list of proteins which were presumably located on the surface to design vaccine and drug targets (http://www.imtech.res.in/raghava/pslpred/submit.html).[6]

Pramod Shinde1(2013) et al has retrieved the complete genome and protein sequences of S. suis from the National Center for the Biotechnology Information (NCBI) and UniProtKB and Proteins sequences with sequence length less than 100 amino acids were excluded from the study because they unlikely to represent essential for pathogenicity. Each protein sequence of S. suis was searched for sequence homology with human proteome using Blastp tool available at NCBI, bit score cut off<100 and minimum expectation value (E -value) cut off E -10 was taken to identify homology exhibiting significant differences with their human host counterpart. Non-human homologs proteins of pathogen, which are possibly unique to S. suis, were then subjected to identify its homolog, standard BLASTp program was used. Functional family prediction of the non homologous essential proteins was done by using the Predict Protein and HNB network. Analysis of the metabolic pathways of the essential proteins of S. suis is done by KEGG Automatic Annotation Server was accessed to for the identification of potential drug targets .Non homologous essential surface membrane proteins of bacteria illustrate their potential of becoming the possible vaccine targets. It is identified by using CELLO, PSLpred
and PSORTb tools. [9]

Md. Musharaf Hossain et.al (2013) has retrieved the complete proteome of *L. monocytogenes* F2365 from National Center for Biotechnological Information (http://www.ncbi.nlm.nih.gov/). The database of essential genes (DEG) was accessed at http://tubic.tju.edu.cn/deg/ (Zhang & Lin, 2009) for the selection of essential genes/proteomes that are vital for the normal activity of the pathogen. The assumption described by Dutta et al. (2006) was followed in this analysis and the DEG search parameter was set to an expectation value (E-value) cut off of 10-10 and bit score not less than 100 (≥100) were selected. Prior to set the DEG search parameter, proteins of less than 100 amino acids and those share >60% sequence identity were excluded from the analysis. Selected sequences form DEG search were BlastP searched against human proteome in the National Center for Biotechnological Information (NCBI) database to identify human non-homologous sequences with threshold E-value of 10-3. Computational prediction of sub cellular localization is necessary for genome analysis and annotation in bacterial pathogens since the prediction of proteins on the cell surface is of particular interest due to the potential of such proteins to be primary drugs or vaccine targets. Sub cellular localization analysis of the essential proteins has been done by Psortb.Metabolic pathway analysis of the essential proteins of *Listeria monocytogenes* F2365 was done by KEGG Automatic Annotation Server (KAAS). Comparative analysis of the metabolic pathways of the host and pathogen was performed by using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (Kanehisa & Goto, 2000) to trace out essential proteins involved in pathogen specific metabolic pathways for the identification of potential drug targets. [8]

5. CONCLUSION

A review was done on five pathogenic organisms wherein subtractive genomics approach was carried out to identify organism specific drug targets. A review was done on five pathogenic organisms which are involved in three different cases. Mostly in all the studies outer membrane proteins were considered as putative drug targets. The result analysis shows that subtractive genomics approach can be successfully applied to identify drug targets in a pathogenic organisms which is specific for that organisms and hence drugs against it cannot show

Cross reactions with human proteins.wet lab analysis of the derived essential organism specific targets obtained from the above said studies can enhance the drug development process

4. Comparative result analysis:

Table 1: Results obtained from five different pathogenic organisms

Steps carried out:	*Salmonella typhi*	*Clostridium botulinum*	*Leishmania donovani*	*Streptococcus suis*	*Listeria monocytogenes*
Total Number of proteins	4718	3404	446	482	2821
Genes whose products are greater than 100 amino acids	-	2938	431	-	2509
Non-paralogs	4559	-	-	249	
Non-human homologous proteins (E-value 10-4)	3570	180	29	37	275
Essential protein in DEG (E-value 10-100)	300	1458 (carried prior)	-	-	745 (carried prior)
Essential proteins involved in metabolic pathways	149	-	-	16	-
Membrane associated non-human homologs of essential genes	11	22	16	7	46

6. REFERENCES:

1. Bhawna Rathi, Aditya N. Sarangi, Nidhi Trivedi, "Genome subtraction for novel target definition in *Salmonella typhi",* Bioinformation (2009) ISSN 0973-2063.

2. Dutta, A., Singh, S. K., Ghosh, P., Mukherjee, R., Mitter, S. and Bandyopadhyay, D., In silico identification of potential therapeutic targets in the human pathogen Helicobacter pylori, In Silico Biol., 6: 0005, (2006).

3. FAO/WHO Microbiological Risk Assessment Series. 2002. Risk Assessment of *Listeria monocytogenes* in ready–to–eat foods: Interpretative Summary http://www.who.int/foodsafety/micro/jemra/assessment/*listeria*/

4. Higgins, R. & Gottschalk, M., Streptococcal diseases. In Diseases of Swine, 1998; Chap. 38. Iowa State University Press.

5. J.F. Tomb, O. White, A.R. Kerlavage, R.A. Clayton, G.G. Sutton and R.D.Fleischmann "Comparative Genomics and Understanding of Microbial Biology" Nature (1997) 388,539-47.

6. Koteswara Reddy.G, K. Nagamalleswara Rao and Prof. P. Rajendra Prasad, "identification of drug and vaccine targets in clostridium botulinum-a by the approach in-silico subtractive genomics", International Journal of Pharmaceutical Studies and Research (2011) E-ISSN 2229-4619.

7. Maltezou, H. C., Drug Resistance in Visceral Leishmaniasis, Biomed and Biotech, 2010: 11-16,

8. Md. Musharaf Hossain ,A.T.M.J. Mosnaz ,Abdul Mohin Sajib ,Pradip Kumar Roy ,Shahriar Kabir Shakil ,S.M. Sarid Ullah ,Shamsul H. Prodhan , "Identification of putative drug targets of *Listeria monocytogenes* F2365 by subtractive genomics approach", *J. BioSci. Biotech.* 2013, 2(1): 63-71. ISSN: 1314-6246

9. Pramod Shinde, Komal Patil, Nikita Malavkar,Aditee Dalvi, Madhura Janve ,"In Silico Subtractive Proteomics Approach for Identification and Characterization of Drug Targets in Streptococcus suis", Journal of Computational Biology(2013), ISSN 2319- 3433.

10 P Everest *et al., Trends Microbial.* 9: 316 (2001) [PMID: 11435104]

11 Rajendra haribhau mandage and amol shriram wadnerkar," Subtractive genomics approach to identify potential therapeutic drug targets in *leishmania donovani",* International Journal of Pharma and Bio Sciences (2010) ISSN 0975-6299.

