# POPULATION LEVEL MODELING OF GENETIC REGULATION OF BIOFILMS

Sreenath Rajagopal*,Prakruthi A, Nainesh K, Varada Anirudan, Santanu Datta, Ramanujan Kadambi S
*Organisation:* Bugworks Research India Pvt Ltd, India
*Email link: sreenath@bugworksresearch.com

## ABSTRACT

Microbial biofilms are a large heterogeneous collection of cells working as a close-knit community located on a suitable substrate and held together within a matrix of polymers and other molecules that have been secreted by this heterogeneous collection. The cells in this collection have varied genetic expressions accounting for division into two fundamentally different states – Planktonic and Biofilm.

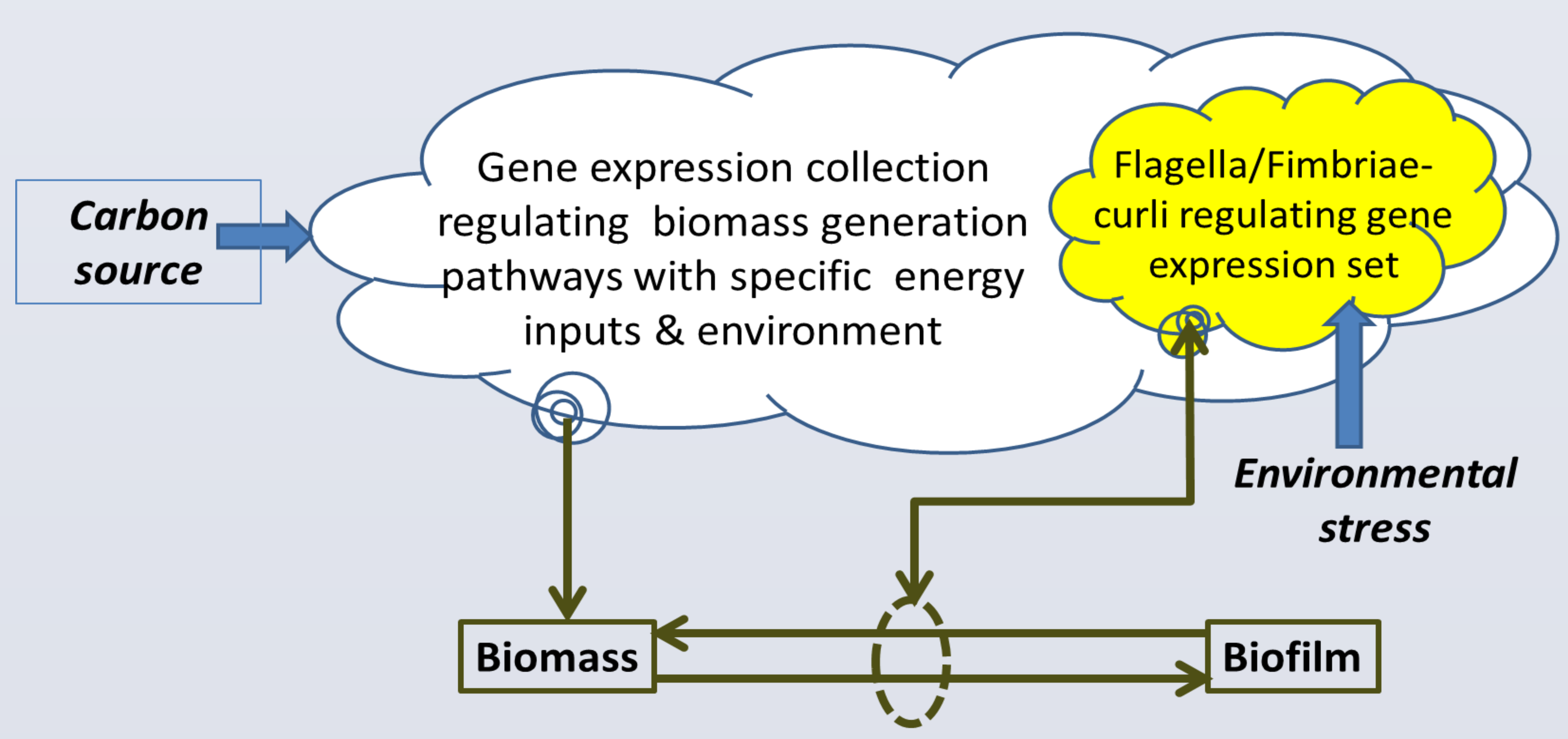
Matrix Of Multiple Solutions (MOMS) is a novel quantitative approach to analyze the multiple genetic expression levels. It is built over the known Flux Based Analysis (FBA) method supported by an Ordinary Differential Equation (ODE) regulatory network. The biofilm associated gene pool of about 100 non-essential genes controls the genetic expression of the Flagellar and Curli/Fimbriae modes, which are mutually exclusive and initiate, sustain and stabilize biofilms. The model is built using the known genetic annotation of the *E.coli* model–ijO1366 and the additional 110 genes connected with biofilm.

The MOMS set shows the distribution of the solutions that encompass a large set of biofilm regulatory gene expression profiles. Specific matrix vector elements of this MOMS set correspond to possible genetic outcomes towards formation of biofilm and modulation of the Flagellar and Curli/Fimbriae pathways. The energy and redox factors in these possible regulatory states provide a predictive method and a "what-if" analysis from simulation results to understand and look at possible drugs targets to prevent and disrupt biofilms.

## MOMS

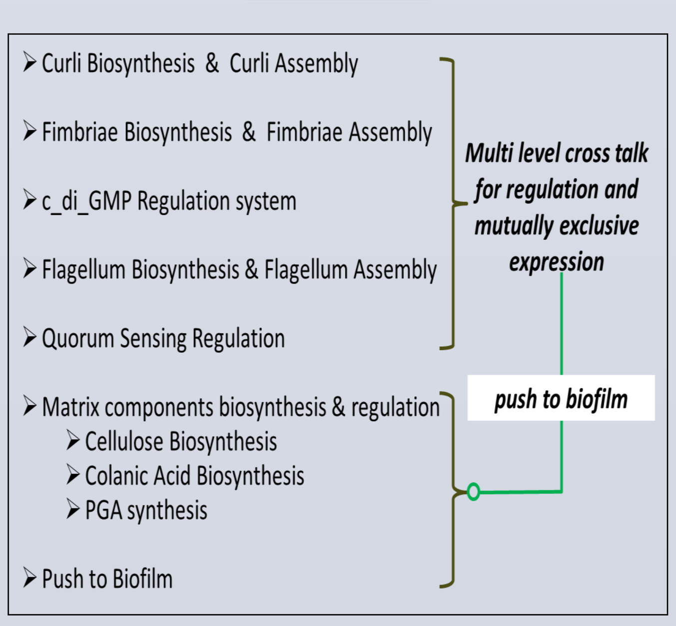
- Matrix Of Multiple Solutions (MOMS): A Proprietary algorithm developed by Bugworks Research Inc. works on underdetermined system of equations and fetches probable solutions for a given condition.
- Underdetermined system - In mathematics, a system of linear equations is considered underdetermined if there are fewer equations than unknowns. An underdetermined system of linear equations has an infinite number of solutions.
- For example consider the given system of linear equations :-
  - X +Y + Z = 10
  - X + 2Y + 3Z = 15
- In the above example it can be seen that there are 2 equations (X +Y + Z = 10; X + 2Y + 3Z = 15) and 3 unknowns (X, Y and Z). When you solve for the same we can find that there are infinite values of X, Y and Z that would satisfy the equations like X = 5.833333, Y = 3.333333, Z = 0.8333333; X = 5.612867, Y = 3.774265, Z = 0.6128673; X = 5.423184, Y = 4.153631, Z = 0.4231843; X = 5.221705, Y = 4.556590, Z = 0.2217052 etc.
- In a biological metabolic network, there are more reactions (Unknowns) than metabolites (Equations). Hence there are infinite number of solutions where each solution corresponds to a way in which a cell can behave for a given condition.
- Similarly in a regulatory network, a given protein acts as repressor and also activator in the cell (eg. CsrA) giving rise to situation of more unknowns than the equations.

## MOMS Model For Biofilm

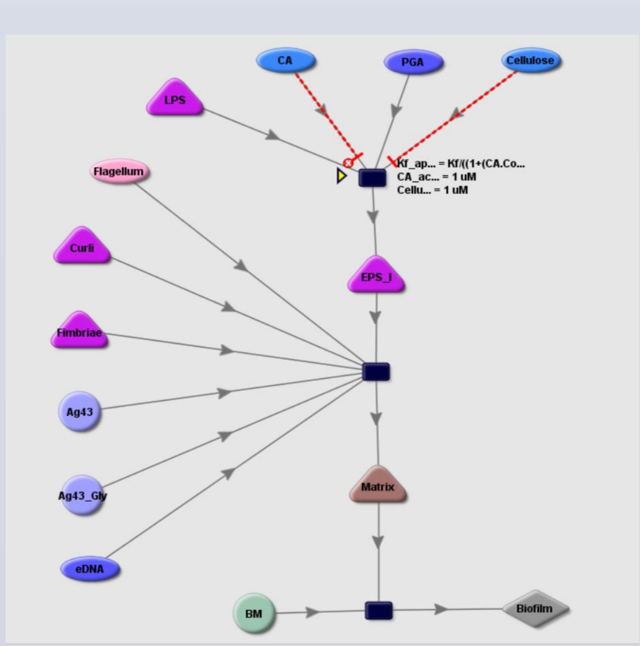



- Microbial biofilms are a large heterogeneous collection of cells working as a close-knit community located on a suitable substrate and held together within a matrix of polymers and other molecules that have been secreted by this heterogeneous collection. Hence understanding the way in which the microorganism acts as a population shall provide a key in deciphering the biofilm behavior.
- A systems biology approach to study biofilm regulation during shift from planktonic phase to biofilm. Changes are governed by a selective subset of pre-dominantly non essential genes responding to stress/altered environmental factors.
- The mutually exclusive parts of this subset trigger a number of specific protein expressions that allow the reconstitution of the biomass from its planktonic phase to a biofilm phase.
- Each gene expression in the respective systemic pathways is combined with a set of activator and repressor genes enabling overall regulation of the Flagellar Vs Curli-Fimbriae modes.
- The gene expressions are divided into the functional blocks (Table-1) terminating towards the biosynthesis of the matrix components that would eventually initiate and constitute the matrix.

## Table - 1



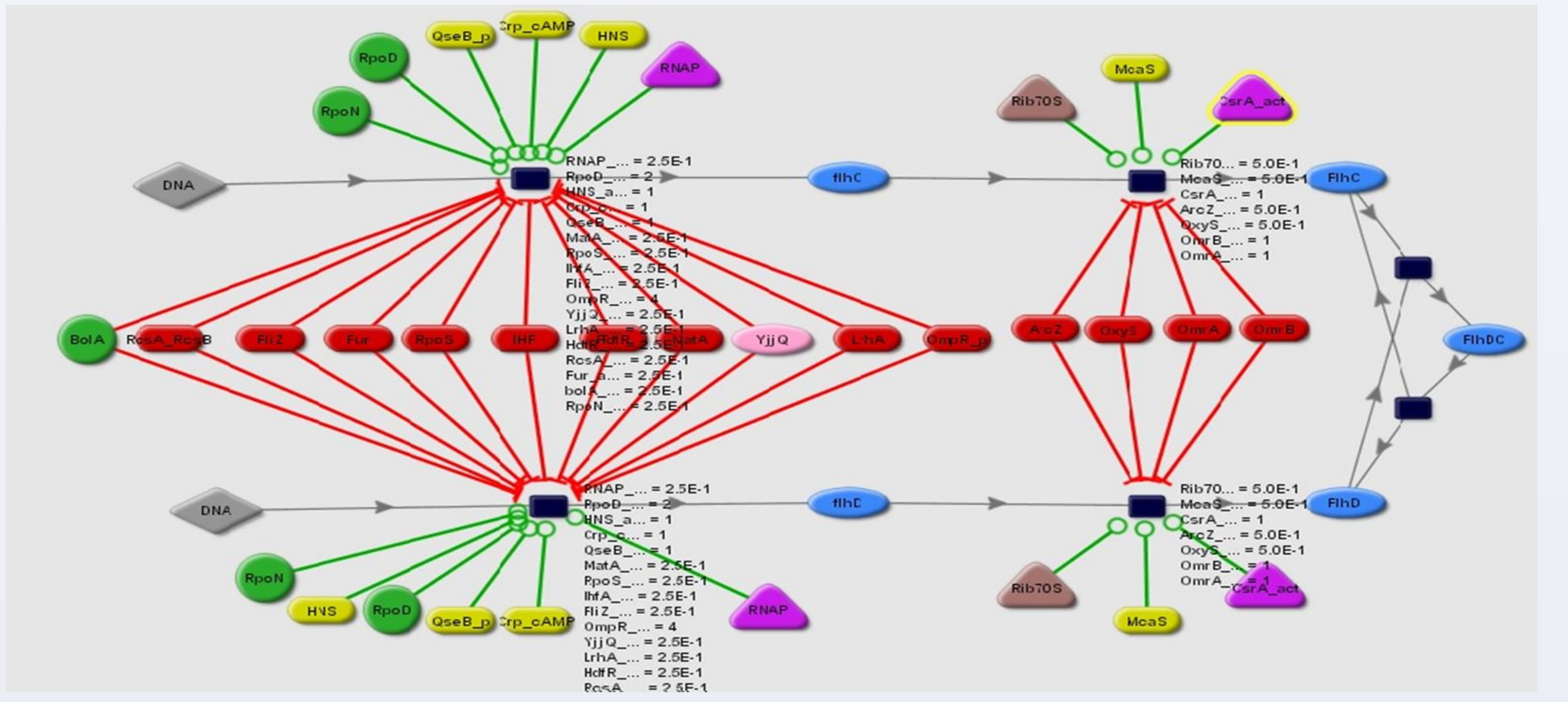

## FIG-1 : End Point Biofilm Matrix



## MOMS Solver

- The solver maps to a computational set of solutions each of which estimates a possible regulatory state of a cell in a population.
- Each possible solution is a cell phenotype mapped to specific values of biofilm matrix formative components and this drives to a quantitative endpoint of matrix components for a given level of carbon source like glucose and thus to varied levels of biofilm.
- The combination of these possible states can be viewed as a plot of scatter points of each of the flux elements that characterize a specific pathway of the systemic network that is regulated.
- These plots provide a probabilistic view of the states that could eventually be the transitive result towards biofilm propensity of a given environment / stress condition.
- Although computationally very intensive, the MOMS approach provides a possible way to get to a probabilistic estimation of the specific gene expressions that form key links to the formation of biofilm.
- The MOMS scatter plots can be considered as estimation of the probability of the occurrence of the respective regulatory pathway expression when paired with other pathway expressions.

## Fig -2 : Typical Biofilm Regulation Block



## Results

- Developed MOMS model with a selective subset of non-essential genes along with the metabolic network, involved in biofilm formation in response to the varying environmental conditions.
- Generated a collection of possible MOMS regulatory solutions.
- The following illustrations show the heterogeneity within a population for a given condition.

**Fig - 3 : ΔcsrA: Effect of Carbon Storage Regulatory system *in silico***

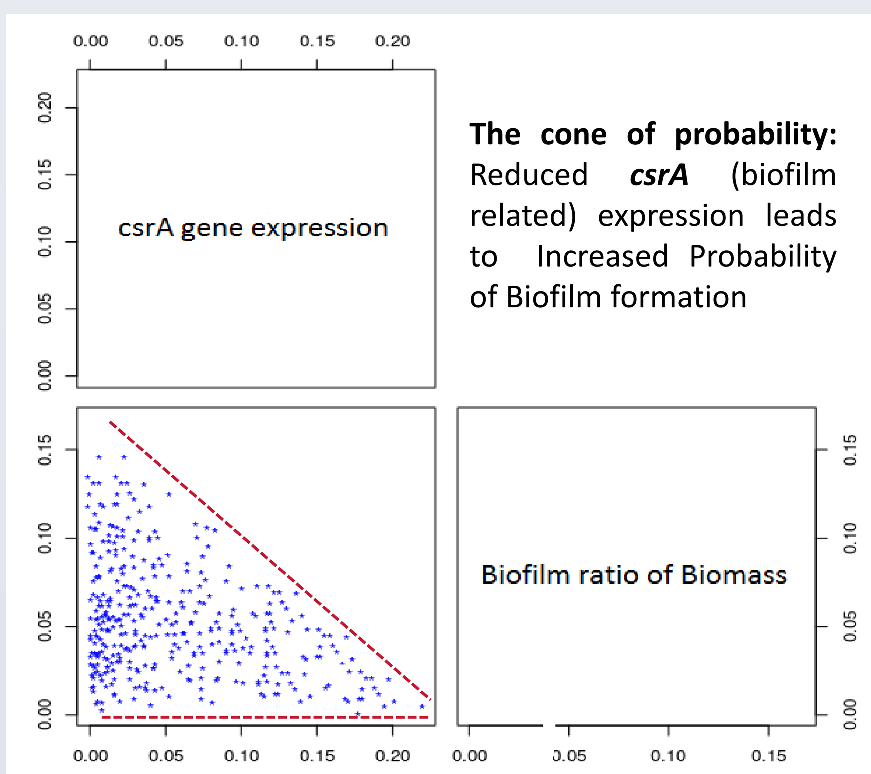
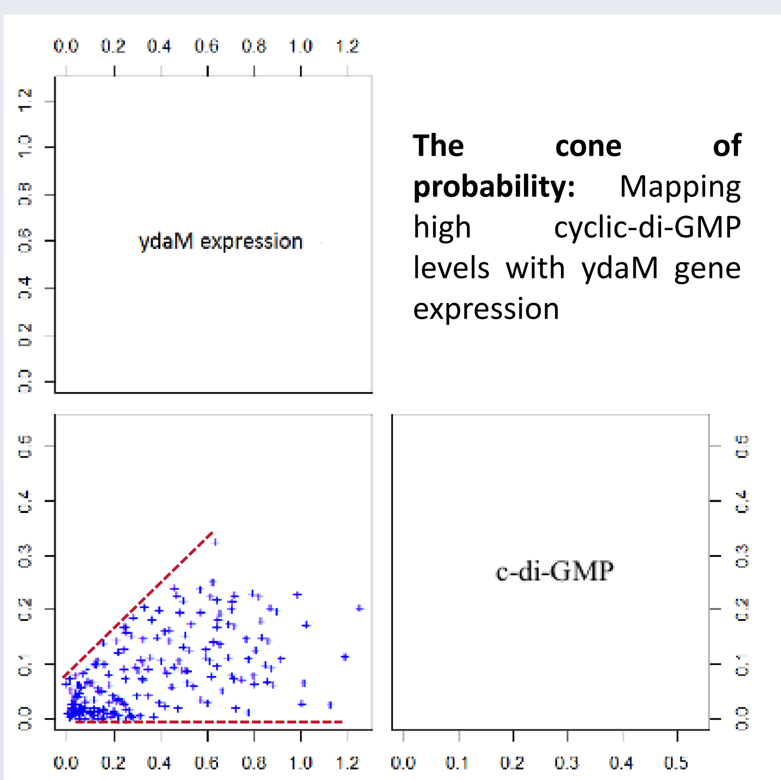



**Fig – 4 : Effect of ydaM expression on Cyclic di-3',5'-guanylate.**




### Validation of model with specific examples

- An inverse correlation between csrA levels and the biofilm formation was established through the MOMS approach (Fig-3).
- *In vitro* experiment was carried out to test the same and it was found that the csrA knockout indeed produced more biofilm than the wild type. (Fig-5), validating the MOMS approach.
- Relation between Di-guanylate Cyclase and ydaM gene was developed (fig-4) – MOMS indicated that overexpression of ydaM leads to increased c-di-GMP levels.
- *E. coli* partial ΔcsrA (deletion of last 30 nucleotides) strain forms biofilm in comparison to the wild type *E. coli* strain. Results similar to that observed in the MOMS prediction (Fig – 5)
- csrA D51-61::kan created by transforming a PCR fragment of Kanamycin cartridge with csrA upstream and downstream homologous regions into the *E. coli* BW25113 carrying lRed plasmid. The recombinant was confirmed by PCR (Fig-6).

**Fig -6 : PCR confirmation for *csrA* knock-out**

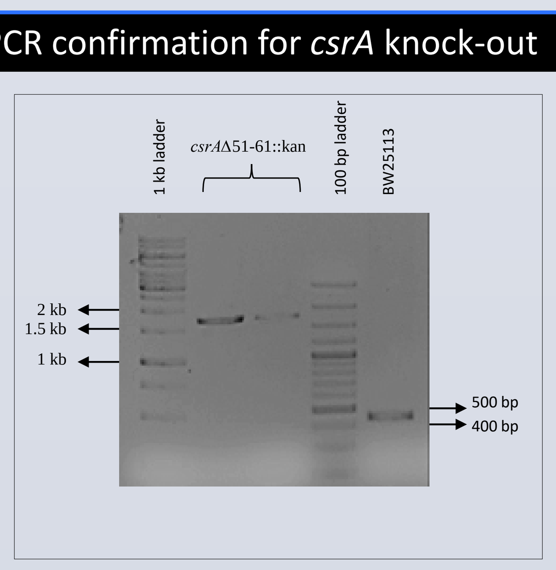
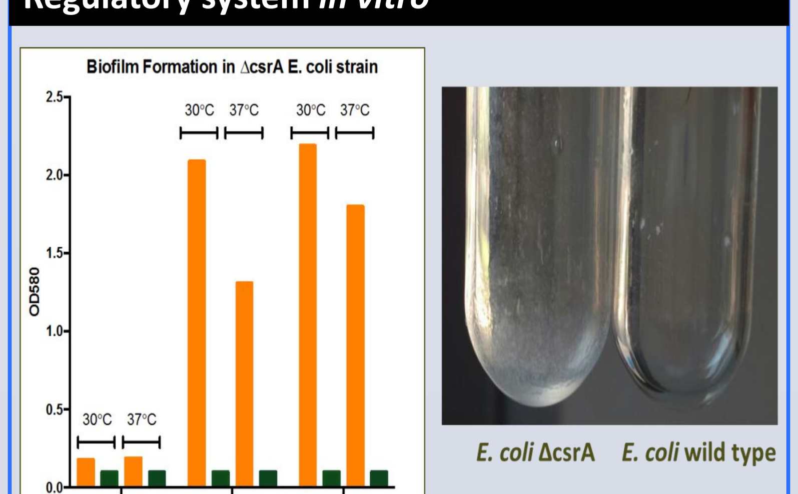



**Fig -5 : ΔcsrA: Effect of Carbon Storage Regulatory system *in vitro***




## Conclusions

- The MOMS approach to solving biological network quantitative behavior in the presence of complex feedforward and feedback loops provides a method to estimate probability of a certain specific end point behavior of the network in the absence of exact gene expression data.
- With large computational power and the ability to handle big data, MOMS approach provides a good handle to predict behavior of a microbial population.
- The distribution of the scatter in a given area when plotted against the end point (formation of biofilm) provides a view of the possible relationship between the gene expression in consideration to the propensity of forming biofilm since the key biofilm matrix expressions converge to the formation as shown in Fig-1.
- This provides a window of opportunities to find possible gene targets for discovery of novel drugs that can prevent/disrupt biofilm.

## References

- Romeo T, Gong M, Liu MY, Brun-Zinkernagel A. Identification and Molecular Characterization of csrA, a Pleiotropic Gene from Escherichia coli That Affects Glycogen Biosynthesis, Gluconeogenesis, Cell Size, and Surface Properties. *Journal of Bacteriology*. 1993;175(15):4744–4755.
- Serra DO, Hengge R. Stress responses go three dimensional - the spatial order of physiological differentiation in bacterial macrocolony biofilms. *Environmental Microbiology*. 2014;16(6):1455–1471. doi:10.1111/1462-2920.12483.
- Hengge R. Novel tricks played by the second messenger c-di-gMP in bacterial biofilm formation. *The EMBO Journal*. 2013;32(3):322–323. doi:10.1038/emboj.2012.351.
- Jackson DW, Suzuki K, Oakford L, Simecka JW, Hart ME, Romeo T. Biofilm formation and dispersal under the influence of the global regulator CsrA of Escherichia coli. *Journal of Bacteriology*. 2002;184(1):290–301. doi:10.1128/jb.184.1.290-301.2002.
- Matrix Of Multiple Solutions (MOMS): A Proprietary algorithm developed by Bugworks Research Inc.



Workshop on Biofilm Biology to Drug Development - 2-3 September 2016, SASTRA University, Thanjavur