

# **Biofilm-i: A platform for predicting biofilm inhibitors using Quantitative structure–activity relationship (QSAR) based regression models to curb antibiotic resistance**

**Akanksha Rajput<sup>1</sup>, Kailash T. Bhamare<sup>1,2</sup>, Anamika Thakur<sup>1,2</sup> and Manoj Kumar<sup>1,2\*</sup>**

<sup>1</sup>Virology Unit and Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research (CSIR), Sector 39-A, Chandigarh-160036, India

<sup>2</sup>Academy of Scientific and Innovative Research (AcSIR), Ghaziabad-201002, India

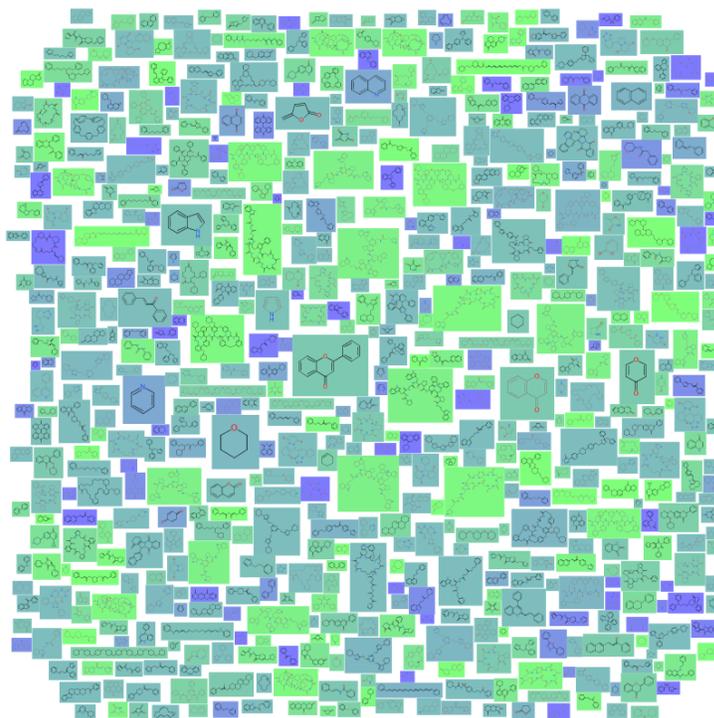
## **\* Correspondence:**

Corresponding Author  
[manojk@imtech.res.in](mailto:manojk@imtech.res.in)

## **Supplementary information**

**Figure S1.** Scaffold cloud view of 884 experimentally validated biofilm inhibitors where biofilm inhibition efficiency shown in colors (blue color depicts 0% and green color displays 100%)

**Figure S2.** Input output of generalised predictor available in biofilm-i web portal



**Figure S1.** Scaffold cloud view of 884 experimentally validated biofilm inhibitors where biofilm inhibition efficiency shown in colors (blue color depicts 0% and green color displays 100%)

**BIOFILM-i**  
A platform for predicting biofilm inhibitors to curb antibiotic resistance

HOME PREDICTOR ALGORITHM TOOLS ANALYSES MORE

**GENERALISED** (3 based regression models to predict)

**GENERALISED Predictor**  
To predict biofilm inhibition efficacy of chemical(s)

**Staphylococcus Aureus**  
(Specific Specific Predictor)  
To predict biofilm inhibition efficacy of chemical(s) against S. aureus (C9orf64 & C9orf64E bacterial)

**Escherichia Coli**  
(Specific Specific Predictor)  
To predict biofilm inhibition efficacy of chemical(s) against E. coli (C9orf64 & C9orf64E bacterial)

**Candida Albicans**  
(Specific Specific Predictor)  
To predict biofilm inhibition efficacy of chemical(s) against C. albicans (C9orf64 & C9orf64E bacterial)

**Anti-Biofilm Peptides**  
To predict biofilm inhibition efficacy of peptide(s)

**Pseudomonas Aeruginosa**  
(Specific Specific Predictor)  
To predict biofilm inhibition efficacy of chemical(s) against P. aeruginosa (C9orf64 & C9orf64E bacterial)

**Predict the biofilm inhibition efficacy (GENERALIZED) OF THE QUERY CHEMICAL(S) IN BATCH MODE**

Input chemical in SDF format

**PREDICTED RESULTS OF GENERALIZED PREDICTOR OF QUERY CHEMICAL(S)**

SMILES of query chemical (s)

Structures of query chemical (s)

Predicted inhibition efficiencies of query chemical (s)

Search similar chemical (s) in aBiofilm resource

Calculation of important descriptors

S.No.	Query ID	Molecular Formula	Current Charge	H Bond acceptors	H Bond donors	UPLINK rotatable bonds	LINKED rotatable bonds	PSA	Rotatable bonds	LogP	Molecular weight
2	600803	C21H24NO6	0	8	5	11	5	0	7	3.240	360.488
3	3084072	C20H15ON2O3	0	3	3	5	3	0	1	5.537	302.764

**Figure S2.** Input output of generalised predictor available in biofilm-i web portal

## Supplementary Table

**Supplementary Table S1.** Performance of all the eight predictors (both training/testing and independent validation) using regression based approach developed using Random Forest along with the final descriptors employed individually

<b>Model used</b>	<b>Data Sets</b>	<b>Feature s</b>	<b>Pr-cor</b>	<b>RMS E</b>	<b>MAE</b>
Chemicals (Overall)	Training/Testing data set (T <sup>800</sup> )	265	0.52	24.22	18.46
	Independent Validation data set (V <sup>84</sup> )		0.45	27.17	21.12
Gram-positive bacteria	Training/Testing data set (T <sup>350</sup> )	177	0.68	22.03	16.11
	Independent Validation data set (V <sup>34</sup> )		0.67	22.49	16.65
Gram-negative bacteria	Training/Testing data set (T <sup>450</sup> )	387	0.57	23.95	18.87
	Independent Validation data set (V <sup>48</sup> )		0.53	22.03	16.26
Fungus/Yeast	Training/Testing data set (T <sup>140</sup> )	111	0.65	19.10	12.55
	Independent Validation data set (V <sup>18</sup> )		0.53	17.33	11.57
<i>Pseudomonas aeruginosa</i>	Training/Testing data set (T <sup>270</sup> )	81	0.65	21.79	16.51
	Independent Validation data set (V <sup>31</sup> )		0.62	22.46	16.27
<i>Staphylococcus aureus</i>	Training/Testing data set (T <sup>210</sup> )	90	0.80	19.36	14.02
	Independent Validation data set (V <sup>29</sup> )		0.78	16.74	12.15
<i>Candida albicans</i>	Training/Testing data set (T <sup>140</sup> )	76	0.63	19.15	12.83
	Independent Validation data set (V <sup>12</sup> )		0.79	15.06	8.88

<i>Escherichia coli</i>	Training/Testing data set (T <sup>93</sup> )	52	0.63	22.56	17.59
	Independent Validation data set (V <sup>10</sup> )		0.86	17.40	13.59