TCBB-2023-12-0805

Machine learning-assisted high-throughput screening for Anti-MRSA compounds

Fadi Shehadeh, *Member, IEEE*, LewisOscar Felix, Markos Kalligeros, Adnan Shehadeh, Beth Burgwyn Fuchs, Frederick M. Ausubel, Paul P. Sotiriadis, *Fellow, IEEE* and Eleftherios Mylonakis

***Abstract*— Background: Antimicrobial resistance is a major public health threat, and new agents are needed. Computational approaches have been proposed to reduce the cost and time needed for compound screening.**

Aims: A machine learning (ML) model was developed for the *in silico* screening of low molecular weight molecules.

Methods: We used the results of a high-throughput *Caenorhabditis elegans* methicillin-resistant *Staphylococcus aureus* (MRSA) liquid infection assay to develop ML models for compound prioritization and quality control.

Results: The compound prioritization model achieved an AUC of 0.795 with a sensitivity of 81% and a specificity of 70%. When applied to a validation set of 22,768 compounds, the model identified 81% of the active compounds identified by high-throughput screening (HTS) among only 30.6% of the total 22,768 compounds, resulting in a 2.67-fold increase in hit rate. When we retrained the model on all the compounds of the HTS dataset, it further identified 45 discordant molecules classified as non-hits by the HTS, with 42/45 (93%) having known antimicrobial activity.

Conclusion: Our ML approach can be used to increase HTS efficiency by reducing the number of compounds that need to be physically screened and identifying potential missed hits, making HTS more accessible and reducing barriers to entry.

***Index Terms*— antimicrobial drug resistance, high-throughput screening, *in silico* screening, machine learning**

Corresponding authors: Fadi Shehadeh, Eleftherios Mylonakis

Fadi Shehadeh is with the Department of Medicine, Houston Methodist Research Institute, Houston, TX 77030, USA and the Department of Electrical and Computer Engineering, National Technical University of Athens, Athens 15780, Greece (e-mail: fshehadeh@houstonmethodist.org).

LewisOscar Felix is with the Department of Medicine, Houston Methodist Research Institute, Houston, TX 77030, USA (e-mail: lfelixrajlucas@houstonmethodist.org).

Markos Kalligeros is with the Division of Internal Medicine, Warren Alpert Medical School of Brown University, Rhode Island Hospital, Providence, RI 02903, USA (e-mail: markos_kalligeros@brown.edu).

Adnan Shehadeh is with the Department of Wine, Vine and Beverage Sciences, University of West Attica, Athens 12243, Greece (e-mail: adnanshehad@uniwa.gr).

Beth Burgwyn Fuchs is with the Division of Infectious Disease, Warren Alpert Medical School of Brown University, Rhode Island Hospital, Providence, RI 02903, USA (e-mail: helen_fuchs@brown.edu).

Frederick M. Ausubel is with the Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA and the Department of Genetics, Harvard Medical School, Boston, MA 02115, USA (e-mail: ausubel@molbio.mgh.harvard.edu).

Paul P. Sotiriadis is with the Department of Electrical and Computer Engineering, National Technical University of Athens, Athens 15780, Greece (e-mail: pps@ieee.org).

Eleftherios Mylonakis is with the Department of Medicine, Houston Methodist Research Institute, Houston, TX 77030, USA and Weill Cornell Medicine, New York, NY 10065, USA (e-mail: emylonakis@houstonmethodist.org).

Color versions of one or more of the figures in this article and supplemental materials are available online at http://ieeexplore.ieee.org

I. INTRODUCTION

ANTIMICROBIAL resistance represents a significant public health concern and is projected to result in up to 10 million deaths annually by 2050, according to estimates from the World Health Organization (WHO) [1-3]. The ongoing emergence of new resistance mechanisms and multidrug-resistant bacteria is compounded by the stagnation in the development of new antibiotics due to scientific challenges and market inefficiencies [4, 5].

Antimicrobial drug discovery typically involves *in vitro* screening of compound libraries for antimicrobial activity, followed by further *in vitro/in vivo* testing for toxicity and refinement through structure-activity relationship (SAR) analysis [6]. Virtual libraries of compounds have been designed using robust chemical transformations [7, 8] and consist of billions of compounds that have never been synthesized or screened previously. While advancements in automation and high-throughput screening (HTS) have made it feasible to screen libraries containing over a million compounds [9], the scale of the chemical search space provided by these libraries exceeds the capabilities of HTS [10]. As a result, this expansion of the chemical search space, along with the high cost of HTS, has led to the use of computational approaches and machine learning (ML) for high-throughput antibiotic discovery [11].

HTS in whole-animal infection models is a powerful tool for identifying compounds with antibacterial activity and low host toxicity [12] but is costly and challenging to automate [13]. In this study, we employed the results of a combined *in vivo* and *in vitro* high-throughput *Caenorhabditis elegans* methicillin-resistant *Staphylococcus aureus* (MRSA) liquid infection assay [14] to develop and evaluate an ML model for an *in silico* approach to screen drug compounds. In this host-pathogen infection model, anti-*S. aureus* compounds can prolong the survival of worms exposed to MRSA in a 384-well liquid assay format. As a metric readout, the screen used

Sytox staining to differentiate between live and dead worms. Sytox is a cell-impermeable die that stains dead worms but not live worms. The compounds in wells that caused a reduction in Sytox staining were considered hits. Efficacious compounds could prolong nematode survival by inhibiting *S. aureus* directly as antibacterial compounds, inhibit bacterial virulence, or modulate nematode host defenses. The process was labor intensive in terms of establishing a screening pipeline and follow-up evaluation of the hit compounds.

Our ML model has the potential to decrease the time and cost associated with whole-animal HTS by reducing the chemical search space and the number of compounds requiring *in vivo* HTS. Furthermore, we demonstrate that this approach could be used for quality control and identification of compound hits missed during HTS.

II. Materials and Methods

High throughput liquid assay. We used data from a high-throughput *C. elegans*-MRSA liquid infection assay as previously described [15]. The assay used robotic pinning to dispense 100 nL of investigational compounds into a 384-well plate format along with methicillin-resistant *S. aureus* (MRSA) strain MW2 and synchronized young adult *glp-4*::*sek-1 C. elegans*. After a 5-day incubation period at 25°C, worms were washed free of bacteria and stained with Sytox to distinguish live from dead worms using automated microscopy and Cellprofiler image analysis. Investigational compound libraries were made available by the Institute of Chemistry and Cell Biology (ICCB) at Harvard Medical School. Worm survival was measured for each well by calculating the ratio of Sytox worm area to bright field worm area. Hits were identified by calculating a Z-score:

$$Z = \frac{x - \mu}{\sigma}$$

where x is the survivability score of each well, μ is the mean of the survivability scores from all the wells, and σ is the standard deviation of the survivability scores from all wells. All compounds were screened in duplicate, and a hit was defined as a compound with $Z \geq 2$ in both screens, as previously described [15]. A threshold of $Z \geq 2$ ensures that hits are compounds with activity at least two standard deviations above the mean, indicating a significantly stronger activity compared to the other compounds on the plate. The assay and the compound screening are described in detail in [14, 15].

Featurization and model development. For each molecule, we calculated physiochemical compound properties such as the topological surface area [16], the exact molecular weight, the number of valence electrons the molecule has, the number of heteroatoms, and the partition coefficient (logP) [17]. Physiochemical drug properties have been associated with drug toxicity [18] and are widely used as input features in ML models for virtual drug screening [19].

We represented the molecular substructures of the compounds as vectors using the Mol2vec unsupervised ML model [20]. Mol2vec uses the Morgan algorithm [21] to represent chemical substructures and applies the Word2vec algorithm [22] on the corpus of compounds by considering compound substructures derived from the Morgan algorithm as "words" and compounds as "sentences." We trained the unsupervised Mol2vec model using more than 20 million chemical compounds from the ZINC12 database [23], with replaced uncommon identifiers using a Skip-gram architecture, a window size of 10, generating 100-dimensional embeddings of Morgan substructures. We used this pre-trained Mol2vec model for the featurization of our high-throughput screening (HTS) hit data, creating a high-dimensional embedding of the screened compounds where related functional groups and molecules are close in the generated vector space.

To explore the differences in the molecular properties of hit and non-hit compounds in the dataset, we conducted a series of boxplot analyses using descriptors associated with Lipinski's Rule of Five descriptors (Ro5) and specifically the molecular weight, partition coefficient (logP), number of hydrogen bond acceptors, and number of hydrogen bond donors. The Ro5 descriptors assess the drug-likeness of compounds based on their molecular properties [24].

To predict if a compound is likely to have antimicrobial or antivirulence activity against *S. aureus* or not, we trained a Balanced Random Forest classifier. A Random Forest classifier is an ensemble classifier that fits decision trees on various random sub-samples of the dataset and aggregates the predictions of the decision trees by majority voting [25]. A Balanced Random Forest classifier is an ensemble classifier for imbalanced data [26]. For each iteration, a bootstrap sample from the minority class and the same number of cases from the majority class are drawn and provided to the tree. The predictions of the ensemble are then aggregated to provide the final prediction [26].

The HTS hit data is an extremely imbalanced dataset as a small minority of the compounds screened were identified as hits. To balance our data and get a more representative split for training and validation, we split the dataset in a stratified fashion by creating splits with the same percentage of hits as in the complete set.

Compound prioritization model. For the compound prioritization model we trained a Balanced Random Forest classifier [26]. To calculate the model hyperparameters, we used a stratified 10-fold cross-validation on the training dataset and estimated the optimal sampling ratio (number of samples in the minority class/number of samples in the majority class after resampling) and the number of trees per forest. We used the recall (sensitivity) score to evaluate the performance of the cross-validated model.

We split our dataset into a training and a validation set using a 70%-30% ratio. We used 70% of the HTS data to train the ML model and 30% to estimate the performance of the model and evaluate its generalizability on new data that were not used during training. We evaluated the performance of the compound prioritization ML model by calculating the sensitivity and specificity of the model for the validation dataset. Sensitivity and specificity were defined as:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%$$

$$Specificity = \frac{TN}{FP + TN} \times 100\%$$

where TP (true positives) is the number of compounds correctly identified by the model as hits, based on the HTS results; TN (true negatives) is the number of compounds correctly identified by the model as non-hits, based on the HTS results; FP (false positives) is the number of compounds incorrectly identified by the model as hits, which are non-hits according to the HTS; and FN (false negatives) is the number of compounds incorrectly identified by the model as non-hits, which are hits according to the HTS.

We selected sensitivity and specificity as the most appropriate metrics to evaluate the performance of the compound prioritization ML model due to their relevance in balancing the identification of true positives (hit compounds) and true negatives (non-hit compounds). In our use case, specificity is critical for reducing the number of non-hit compounds that proceed to costly and time-consuming in vitro/in vivo validation. Meanwhile, sensitivity ensures that most of the true hit compounds are retained in the subset for further testing, thereby minimizing the risk of missing potential therapeutic candidates.

To assess the robustness of our model, we performed y-randomization by randomly shuffling the response variable in the training dataset while keeping the feature variables unchanged [27]. This process was repeated 100 times, and the model performance was evaluated for each iteration to determine if the performance of our model was significantly better than that of the randomized models.

To define the region within the chemical space where the model is anticipated to deliver accurate predictions, we performed applicability domain (AD) analysis as described in [28, 29]. We used t-distributed stochastic neighbor embedding (t-SNE) to visually represent and compare the feature space of the training and validation dataset. Compounds that fall within the boundary of the training dataset are considered to fall within the applicability domain of the model.

Quality control model. To identify compounds that may have been misclassified as non-hits by the HTS, we trained a separate Balanced Random Forest classifier after performing minority class augmentation by generating stereoisomers of the hit compounds. This resulted in a much larger number of compounds being defined as hits for the purpose of training, creating a more balanced dataset. We fit the classifier using all the compounds of the HTS dataset, and we used the resulting model for the *in silico* screening of the same dataset.

We did not calculate classification performance metrics for the quality control ML model, as it was trained on the full dataset and was solely used for the identification of possible hits that were misclassified by the HTS. To evaluate the performance of this model and verify if the discordant hits had antimicrobial activity, we performed a literature review of the identified discordant hits.

The training and evaluation of the ML models was performed using sklearn v0.24.2 [30] and processing of the molecules using rdkit v2021.03.5 [31]. We used T-distributed Stochastic Neighbor Embedding (t-SNE) [32] to visualize our results in 2-dimensions.

III. Results and Discussion

High-throughput screening dataset

The ML model in this study was trained on a *C. elegans*-MRSA HTS dataset of 82,286 compounds and extracts, which were tested in duplicate for their ability to protect *C. elegans* from MRSA-mediated killing [14]. The diversity of the screen included U.S. Food and Drug Administration (FDA)-approved drugs, commercially available libraries, and aqueous fractions from plants and microbes. After removing compounds and extracts that could not be represented by the simplified molecular-input line-entry system (SMILES) notation [33], duplicate compounds, and compounds that were identified as potential hits in only one of the two replicates, the HTS data set consisted of 75,891 unique compounds with 335 (0.44%) being designated as hits ($Z \geq 2$). The boxplots presented in **Supplementary Figure 1** illustrate the distribution and variation of the Ro5 descriptors within the hit and non-hit compound groups. To estimate the performance of the ML model and evaluate its generalizability on new data that were not used during training, we divided the dataset into training and validation sets using stratified sampling and created splits with the same percentage of hits as in the complete set. The training dataset included 53,123 compounds with 234 (0.44%) hits, and the validation dataset included 22,768 compounds with 101 (0.44%) hits. The validation dataset provided an unbiased estimate of the performance and generalizability of the model as it was not used for fitting the model.

Selection of compound prioritization machine learning model

The hit rate in whole-animal *C. elegans* HTS is typically low, with less than 0.5% of compounds identified as hits [13, 34], leading to high screening costs and long lead times for screening campaigns [35]. The imbalanced nature of the datasets, *i.e.,* the fact that there is a very small number of hit compounds compared to the total number of compounds screened, presents a challenge for the classification performance of ML algorithms [36]. When facing imbalanced scenarios, standard classifiers such as logistic regression, support vector machines (SVM), and decision trees often provide suboptimal classification results, with good coverage of the majority classes while distorting the minority [37]. To overcome these challenges, we used a Balanced Random Forest classifier, which samples the training data in each iteration by drawing a sample from the minority class and the same number of cases from the majority class [26], thus balancing the training data of each iteration. 10-fold cross-validation yielded an optimal sampling ratio of 100% (number of samples in the minority class over the number of samples in the majority class after resampling) and 360 trees per forest. The maximum depth hyperparameter was set to "None".

We compared the performance of the Balanced Random Forest classifier with that of an SVM and a Random Forest classifier, and the detailed metrics are shown in **Supplementary Table 1**.

To verify that the model selected is anticipated to deliver

accurate predictions for the compounds of the validation set, we performed AD analysis. AD analysis ensures that the model is used within the same chemical space it was trained on. As seen in **Supplementary Figure 2**, the chemical spaces of the training and validation datasets overlap significantly. This overlap indicates that the compounds in the validation set fall within the same chemical space as those used to train the model, demonstrating that the model is suitable for predicting the activity of the validation compounds.

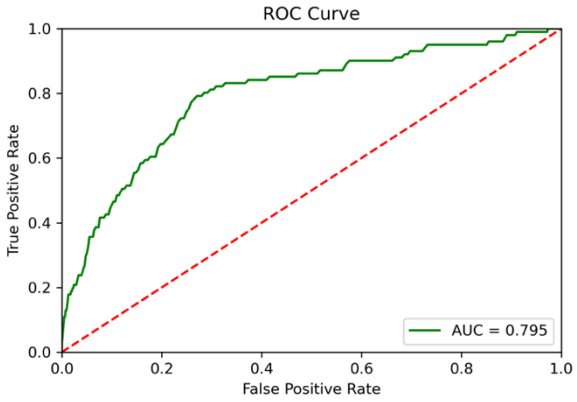

Fig. 1. ROC curve of the hit classification model in the validation set.

Evaluation of the compound prioritization machine learning model

The compound prioritization model achieved an area under the receiver operating characteristic (ROC) curve (AUC) of 0.795 (**Fig 1**), with a sensitivity of 81% and a specificity of 70%. In total, 6,966 compounds were classified as potential hits by our ML model in the validation set, and among these compounds were 82 of the 101 compounds identified as hits using the traditional HTS *C. elegans*-MRSA model. Meanwhile, our model classified concordantly 15,783 out of 22,667 non-hit compounds in the validation dataset. The sensitivity of the model was 81% (82/101) and it indicates the effectiveness of our approach in identifying hit compounds and ensuring that most potential therapeutic candidates were captured. On the other hand, the specificity reflects the ability of the model to correctly exclude 70% (15,783/22,667) of the non-hit compounds, thereby minimizing unnecessary screening efforts. Additionally, we calculated the model performance metrics for each compound library in the validation set. The ML model had a median per-plate sensitivity of 89% (IQR 72%-100%) and a median per-plate specificity of 62% (IQR 51%-71%). Detailed metrics are shown in **Supplementary Table 2**.

Finally, the y-randomization test showed that the performance of the randomized models was significantly lower than that of the original model. The mean sensitivity and specificity for the randomized models was 49.5% and 49.2%, respectively, compared to 81% and 70% for the original model. This significant decrease in performance metrics for the randomized models indicates that the observed high performance of our original model is not due to random chance but is instead reflective of its ability to accurately capture the underlying relationships in the data.

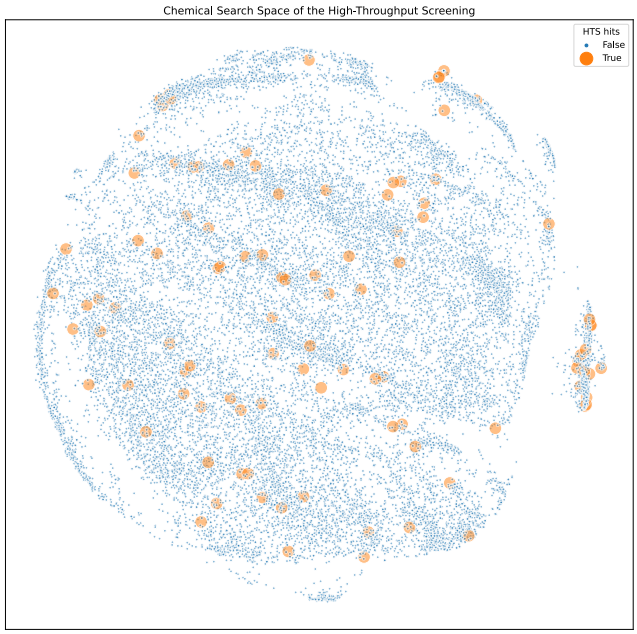

Fig. 2A. The complete chemical space of the validation dataset.

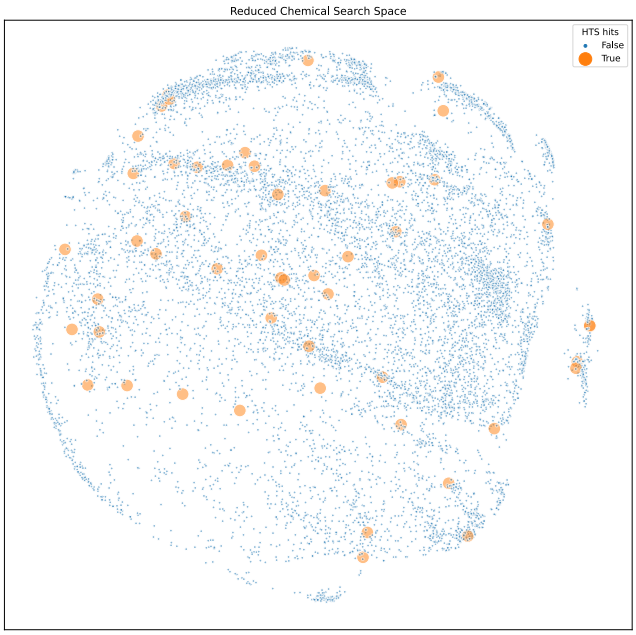

Fig. 2B. The reduced chemical search space after removing the compounds that were identified as non-hits by our machine learning model.

Iterative protocol for compound prioritization

The results from the previous section suggest a practical iterative protocol using our ML model to reduce the number of compounds in the *C. elegans*-MRSA HTS platform [38, 39]. In step one, a fraction of a set of compounds to be tested is subjected to screening using the C. elegans HTS procedure. In step two, the results from this partial screening are used to train the ML model. In step three, the remaining compounds are analyzed *in silico* to

identify a subset of potential hits. In step four, the potential hit compounds are tested in the *C. elegans* model. In this scenario, the 15,802/22,768 compounds that were classified as non-hits by the model would be excluded from further *in vitro/in vivo* screening with the expectation of identifying approximately 81% (82/101) of the hit molecules, thus reducing the number of compounds to be screened from 22,768 to 6,966, a very significant decrease with respect to time, effort, and cost.

Figure 2A depicts the complete chemical space of the validation dataset, while **Figure 2B** depicts the reduced chemical search space after removing the compounds that were identified as non-hits by our ML model. By screening only these 6,966 prioritized compounds (30.6% of the compounds), the high-throughput *C. elegans*-*S. aureus* screening assay would have identified 82/101 hits (81% of the active compounds). In a similar ML approach, Dreiman *et al*. [35] used cell-free and cell-based HTS datasets available through PubChem and showed that by training a random forests model with 35% of the data, they could predict 55%-80% of the active compounds. Likewise, Paricharak *et al*. [39] performed a similarity search using circular and HTS fingerprints in a large proprietary HTS dataset of 1.3M compounds and retrieved diverse compounds belonging to the top 0.5% of their HTS campaign by retrospectively screening ~1% of their compound collection. However, their performance was inferior in cell-based assays compared to cell-free assays. Our study extends these approaches to whole-animal HTS, and our results expand these previous reports, using a much smaller number of hit compounds from a *C. elegans*-*S. aureus* liquid infection assay.

The use-case scenario of the prioritization model described above would have decreased the number of compounds screened by 69.3%, resulting in a hit rate of 1.17% (82/6,966), which is 2.67 times higher than the original hit rate. Our ML approach for compound selection has the potential to make resource-intensive HTS techniques, such as whole-animal HTS, more facile and accessible and reduce the barrier to entry for academic and resource-constrained groups while also extending the chemical space that can be explored.

Identification of missed hit compounds using machine learning

HTS assays are subject to random and systemic variability due to differences in biological activity, random errors, differences between reagent lots, compound concentration, uneven temperature, and failed compound or reagent transfer [40-42]. Various statistical methods are commonly used to identify hits in HTS and reduce the impact of these errors, but no single method is optimal for every assay [42]. To identify compounds that may have been misclassified by the *C. elegans*-MRSA HTS described above, we retrained the ML model on all the compounds of the HTS dataset. This model identified 17 384-well plates with possible hits that were not identified by the HTS. We evaluated the HTS quality for these plates by calculating the Z-factor, which is a dimensionless statistic used for assay quality assessment and can indicate whether there is adequate separation between positive and negative controls [43]. These plates had a mean Z-factor of -0.98 (SD: 1.39), indicating that there was no separation band between the positive and negative controls, and thus the results of the HTS in these plates could not be trusted. The Z-factor of each plate and the number of compounds with published biological activity in the literature are provided in **Table 1**. Our quality control model allows for the real-time evaluation of results so the part of the library that did not meet technical standards can be identified and rescreened.

TABLE 1
Z-factor and number of compounds with published biological activity in the literature for the plates identified by the quality control model.

Library	Plate	HTS Z' (A)	HTS Z' (B)	Discordant hits (N)	Verified as active in literature (N)
Chembridge3	1585	-2.6241408	0.31067307	1	0
ChemDiv5	1711	0.41737861	0.0180734	1	0
NINDS2	1920	0.42185919	0.37378134	3	3
NINDS3	1921	-0.1353595	-0.1407361	2	2
NINDS4	1922	0.05531238	-1.3998514	8	8
NINDS5	1923	0.14255031	0.0521821	2	2
Biomol lCCB3	1990	-0.8503376	0.00191202	1	1
Biomol4	2090	0.14117253	-0.0675108	6	6
Prestwick2	2095	-3.0531884	-3.1146967	2	2
Prestwick2	2096	-1.8546921	-3.6324069	3	3
Prestwick2	2097	-0.8341563	-1.7144365	4	4
Prestwick2	2098	-3.8852087	-3.4002068	2	2
LOPAC1	3260	-0.3247594	-0.9053192	2	2
LOPAC1	3261	-0.4595694	0.587314	1	1
LOPAC1	3263	-0.1902764	0.48519967	1	1
NCC1-2012	3393	-1.6607249	-3.6364711	4	3
NCC1-2012	3394	-0.9455457	-1.48888	2	2

Analysis of discordant results

In the set of 75,891 unique compounds screened, the ML model classified 45 molecules, which were classified as non-hits by the whole-animal HTS, as hits. We reviewed the published data related to these compounds, and for 42/45 of the discordant molecules, we found published results supporting their antimicrobial activity. A detailed list of these compounds and the related literature is summarized in **Table 2**. Of these 42 compounds, 37 are known antibiotics of various classes. Specifically, we identified 12 cephalosporins (2 first generation, 1 second generation, and 9 third generation), 7 macrolides, 6 fluoroquinolones, 4 rifamycins, 2 antitumor compounds with antibiotic activity, 2 topical antibiotics, 1 aminoglycoside, 1 aminocoumarin antibiotic, 1 ionophore antibiotic, and 1 penicillin. All these antibiotic compounds have published evidence of anti-staphylococcal activity, while 11 of them have documented anti-MRSA activity (**Table 2)**.

Of the 5 non-antibiotic compounds, 2 are known disinfectants,

1 is an antineoplastic agent, 1 is a protease inhibitor, and 1 is an FDA-approved antihypertensive agent. Upon further literature review, we found that 4 out these 5 compounds possess anti-staphylococcal and anti-MRSA activity [44-46]. Specifically, the disinfectants chlorhexidine and methylbenzethonium chloride, although toxic for systemic use, are both effective against *S. aureus* [44, 45]. Also, the antihypertensive agent (candesartan cilexetil, PubChem CID: 2540) has been tested and was shown to have antimicrobial and specifically anti-*S. aureus* activity [46]. Finally, the antineoplastic/antimetabolite agent (floxuridine, PubChem CID: 5790) has inhibitory and anti-virulence activity against *S. aureus* [47], and a recent report by Sharma *et al*. also revealed strong synergy between cefoxitin and floxuridine against MRSA [48].

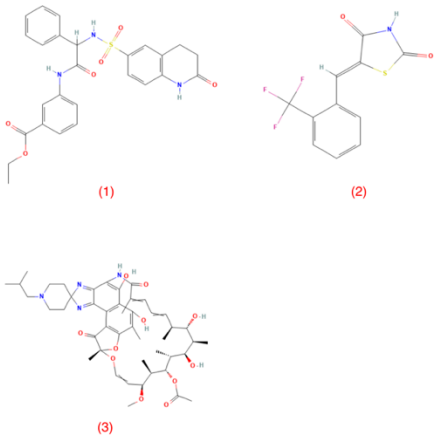

Fig. 3. 2D Structures of compounds 1-3.
(1) PubChem CID: 20864422, URL: https://pubchem.ncbi.nlm.nih.gov/compound/20864422#section=2D-Structure,
(2) PubChem CID: 1327345, URL: https://pubchem.ncbi.nlm.nih.gov/compound/1327345#section=2D-Structure ,
(3) PubChem CID: 51619, URL: https://pubchem.ncbi.nlm.nih.gov/compound/51619#section=2D-Structure

TABLE 2
Compounds identified by the quality control model with published biological activity in the literature.

PubChem CID	Name	Compound type	Anti-SA & anti-MRSA Activity
135550179	Rifampin	Rifamycin	Liu *et al.* [49]
23675312	Cephapirin sodium	1st generation cephalosporin	Ster *et al.* [50]
23675322	Cefazolin sodium	1st generation cephalosporin	Pant *et al.* [51]
23672568	Cefamandole sodium	2nd generation cephalosporin	Coppens *et al.* [52, 53]
6398970	Cefdinir	3rd generation cephalosporin	Giordano *et al.* [54]
6321411	Cefixime	3rd generation cephalosporin	Bergeron *et al.* [55]
2713	Chlorhexidine	Disinfectant	Hayden *et al.* [45]
101526	Moxifloxacin hydrochloride	Fluoroquinolone	Lemaire *et al.* [56]
12971800	Ceftriaxone sodium	3rd generation cephalosporin	Lowe *et al.* [57]
23672566	Cefotaxim sodium salt	3rd generation cephalosporin	Aldridge K.E. [58]
5379	Gatifloxacin	Fluoroquinolone	Blondeau *et al.* [59]
135749824	Rifapentine	Rifamycin	Karau *et al.* [60]
4539	Norfloxacin	Fluoroquinolone	Gade *et al.* [61]
72111	Rifaximin (Xifaxan)	Rifamycin	Jiang *et al.* [62]
5479527	Cefotaxim sodium	3rd generation cephalosporin	Aldridge K.E. [58]
23670319	Cefuroxime sodium salt	3rd generation cephalosporin	Rasmussen *et al.* [63]
23581806	CID 23581806	3rd generation cephalosporin	PubChem Bioassay Record for AID 720641 [64]
60196280	Ceftriaxone	3rd generation cephalosporin	Lowe *et al.* [57]
11957499	Calcimycin	Ionophore antibiotic	Westhead J.E. [65]
23695850	Fusidin	Topical antibiotic	Ayliffe *et al.* [66]
23670321	AKOS015994666	3rd generation cephalosporin	Rasmussen *et al.* [63]
5790	Floxuridine	Antineoplastic	Yeo *et al.* [47], Sharma *et al.* [48]
5702003	LSM-1590	Antitumor antibiotic	Jacobs *et al.* [67]
5746	Mitomycin	Antitumor antibiotic	Jacobs *et al.* [67]
4583	Ofloxacin	Fluoroquinolone	Fu *et al.* [67]
149096	Levofloxacin	Fluoroquinolone	Fu *et al.* [67]
23690471	Sodium fusidate	Topical antibiotic	Ayliffe *et al.* [66]
56208	Sarafloxacin	Fluoroquinolone	Asadipour *et al.* [68]
5480431	(Z)-roxithromycin	Macrolide	Rayner *et al.* [69]
443953	Erythromycin ethylsuccinate	Macrolide	Rayner *et al.* [69]
5702238	Methylbenzethonium chloride	Disinfectant	Bearden *et al.* [44]
6915744	Roxithromycin	Macrolide	Rayner *et al.* [69]
84029	Clarithromycin	Macrolide	Rayner *et al.* [69]
12560	Erythromycin	Macrolide	Rayner *et al.* [69]
447043	Azithromycin	Macrolide	Rayner *et al.* [69]
135549317	Prestwick3 001109	Rifamycin	Karau *et al.* [60]
6473883	Dirithromycin	Macrolide	Rayner *et al.* [69]
54723343	Novobiocin sodium salt	Aminocoumarin	Walsh *et al.* [70]
2540	Candesartan cilexetil	Antihypertensive	Xu *et al.* [46]
118796988	CHEBI:95195 (Roxithromycin)	Macrolide	Rayner *et al.* [69]
25102585	flucloxacillin sodium	Penicillin	Leder *et al.* [53]
439647	Tosylphenylalanyl chloromethyl ketone	Protease inhibitor	

Antibacterial activity of compounds

We also evaluated the activity of the 3 remaining compounds identified by the ML model for which there are no published reports of antimicrobial activity (**Figure 3**). This was carried out using the *S. aureus* strain MW2 (SA-MW2) *in vitro* by performing broth microdilution to calculate the minimum inhibitory concentration (MIC) [71]. To assess *in vivo* antibacterial efficacy, the compounds were evaluated using the wax moth caterpillar *Galleria mellonella* infection model [72, 73].

Compound 1 ($C_{26}H_{25}N_3O_6S$, IUPAC: ethyl 3-[[2-[(2-oxo-3,4-dihydro-1*H*-quinolin-6-yl)sulfonylamino]-2-

phenylacetyl]amino]benzoate, PubChem CID: 20864422) consists of three chemically active parts; ethyl benzoate (ester), 2-amino-2-phenylacetamide, and 2-oxo-1,2,3,4-tetrahydroquinoline-6-sulfonamide. Compound 1 did not exhibit direct anti-MRSA activity (MIC > 256 µg/mL against MRSA isolate MW2). On the other hand, Compound 1 significantly prolonged larval survival in the *G. mellonella* infection assay (**Figure 4A**; LD_{50} = 24 h at 3 mg/kg, Kaplan-Meier p = 0.0428), suggesting that it modulates host defenses, disrupts *S. aureus* virulence, or is a prodrug that needs to be metabolized by a host metabolic enzyme to generate an active antimicrobial compound. If Compound 1 functions as an antimicrobial prodrug, it seems most likely that the 2-oxo-1,2,3,4-tetrahydroquinoline-6-sulfonamide moiety is the antimicrobial component since many sulfonamides are known to inhibit the growth of bacteria by disrupting the synthesis of folic acid [74], and sulfonamide derivatives have been extensively studied for their anti-*S. aureus* activity [75, 76].

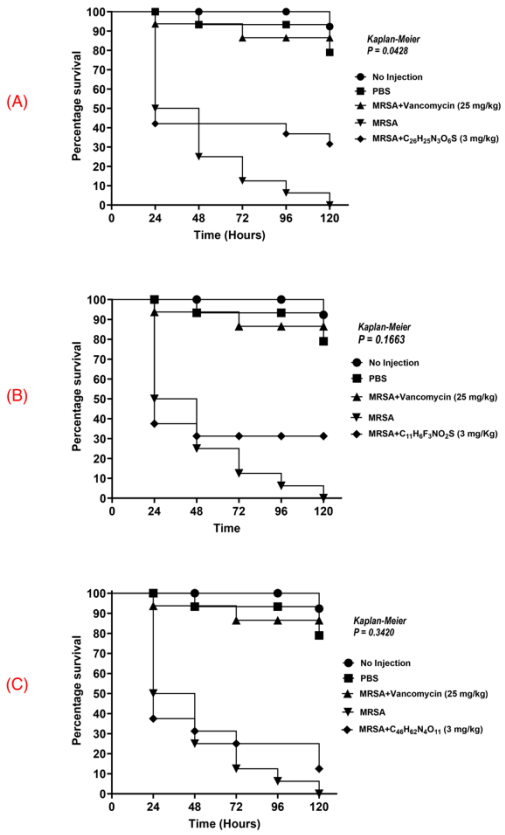

Fig. 4. *Galleria Mellonella* survival assay. Survival rate of compound 1 (A), compound 2 (B) and compound 3 (C).

Compound 2 ($C_{11}H_6F_3NO_2S$, IUPAC: (5Z)-5-[[2-(trifluoromethyl)phenyl]methylidene]-1,3-thiazolidine-2,4-dione, PubChem CID: 1327345) demonstrated weak inhibitory activity against *S. aureus* strain MW2 with an MIC of 32 µg/mL but had no efficacy in prolonging larval survival in the *G. mellonella-S. aureus* infection assay (**Figure 4B**; LD_{50} = 48 h at 3mg/kg, Kaplan-Meier p = 0.166). Compound 2 consists of two chemically active parts: trifluoromethyl benzene and thiazolidine-2,4-dione (TZD). Previous studies have revealed that compounds containing the 2,4-thiazolidinedione ring exhibit potent antimicrobial activity, particularly against drug-resistant bacteria such as *S. aureus*. [77-80].

Compound 3 ($C_{46}H_{62}N_4O_{11}$ IUPAC: [(7S,11S,12R,13S,14R,15R,16R,17S,18S)-2,15,17,32-tetrahydroxy-11-methoxy-3,7,12,14,16,18,22-heptamethyl-1'-(2-methylpropyl)-6,23-dioxospiro[8,33-dioxa-24,27,29-triazapentacyclo[23.6.1.14,7.05,31.026,30]tritriaconta-1(31),2,4,9,19,21,25(32),26,29-nonaene-28,4'-piperidine]-13-yl] acetate, PubChem CID: 51619) inhibited the growth of *S. aureus in vitro* with an MIC of 16 µg/mL but did not show any evidence of efficacy in the *G. mellonella* survival assay, possibly due to compound cytotoxicity (**Figure 4C**; LD_{50} = 48 h at 3mg/kg, Kaplan-Meier p = 0.342). Compound 3 is a structural isomer of rifabutin (PubChem CID: 135398743), which is an FDA-approved rifamycin. Similar to other rifamycins, rifabutin exerts its antimicrobial action by inhibiting bacterial RNA polymerase and is primarily used as an anti-tuberculous agent against *M. avium* as well as an adjunctive therapy for resistant *Acinetobacter* [81]. Rifabutin also has comparable *in vitro* activity to that of rifampin against staphylococcal biofilms [82]. A recent study by Kumar *et al*. using a rat model also demonstrated its potential as an alternative to rifampin for the management of MRSA periprosthetic joint infection when used in conjunction with vancomycin [60]. These findings have sparked new interest in rifabutin as an anti-MRSA compound, but clinical trial data on its efficacy against MRSA are lacking [83].

To summarize this section, the three compounds identified as potential antimicrobial compounds by the ML model, but for which there are no published reports of antimicrobial activity, contain components that are similar to compounds known to exhibit antimicrobial activity (Compounds 1 and 2) or is an isomer of a known antimicrobial compound (Compound 3).

IV. Conclusion

The results of our study show that ML models are effective at identifying compounds with antimicrobial activity, which may have been overlooked by traditional HTS methods, especially the *C. elegans* HTS platform. The model achieved an AUC of 0.795 with a sensitivity of 81% and a specificity of 70%. When pseudo-prospectively applied to a validation set of 22,768 compounds, the model was able to identify 81% of the active compounds by screening only 30.7% of the total compounds, resulting in a 2.67-fold increase in hit rate compared to the original hit rate.

Our ML approach effectively overcomes the challenges presented by imbalanced HTS hit data. Of the 45 discordant molecules classified as non-hits by the HTS but as hits by the ML model, 42 (93%) have known antimicrobial activity, while the 3 unknown compounds revealed either *in vitro* or *in vivo* activity against MRSA. Importantly, the use of ML models also appears to identify potentially toxic compounds such as chlorhexidine and methylbenzethonium chloride that possess anti-*S. aureus* activity, something that could not have been

possible with the use of the *G. mellonella* survival assay due to compound toxicity. This ability to identify chemical classes and compounds that exhibit antimicrobial activity but may be toxic can be used to guide the synthesis and/or testing of less toxic isomers for further evaluation. Furthermore, ML models may also be able to overcome other limitations inherent in HTS methods, such as solubility issues, thus providing an additional level of insight into the discovery of new antimicrobial compounds.

In summary, our ML approach can be used to complement HTS methods, including whole-animal HTS, by making HTS more widely accessible and efficient by significantly decreasing the number of compounds that need to be screened. This approach expands previous reports and highlights the potential of ML in the field of compound prioritization and drug discovery.

REFERENCES

[1] World Health Organization. "Global action plan on antimicrobial resistance," 4 Jan 2023; https://www.who.int/publications/i/item/9789241509763.

[2] Centers for Disease Control and Prevention, *Antibiotic resistance threats in the United States, 2019*: US Department of Health and Human Services, Centres for Disease Control and Prevention, 2019.

[3] Interagency Coordination Group on Antimicrobial Resistance. "No Time to Wait: Securing the Future from Drug-Resistant Infections. Report to the Secretary-General of the United Nations," 4 Jan 2023; https://www.who.int/docs/default-source/documents/no-time-to-wait-securing-the-future-from-drug-resistant-infections-en.pdf.

[4] C. Lepore, L. Silver, U. Theuretzbacher, J. Thomas, and D. Visi, "The small-molecule antibiotics pipeline: 2014-2018," *Nat Rev Drug Discov,* vol. 18, no. 10, pp. 739, Sep, 2019.

[5] National Academies of Sciences, Engineering, Medicine, *Combating Antimicrobial Resistance and Protecting the Miracle of Modern Medicine*, Washington, DC: The National Academies Press, 2022.

[6] A. Desalermos, X. Tan, R. Rajamuthiah, M. Arvanitis, Y. Wang, D. Li, T. K. Kourkoumpetis, B. B. Fuchs, and E. Mylonakis, "A multi-host approach for the systematic analysis of virulence factors in Cryptococcus neoformans," *J Infect Dis,* vol. 211, no. 2, pp. 298-305, Jan 15, 2015.

[7] Q. Hu, Z. Peng, J. Kostrowicki, and A. Kuki, "LEAP into the Pfizer Global Virtual Library (PGVL) space: creation of readily synthesizable design ideas automatically," *Methods Mol Biol,* vol. 685, pp. 253-76, 2011.

[8] T. Sterling, and J. J. Irwin, "ZINC 15--Ligand Discovery for Everyone," *J Chem Inf Model,* vol. 55, no. 11, pp. 2324-37, Nov 23, 2015.

[9] R. H. Clare, C. Bardelle, P. Harper, W. D. Hong, U. Borjesson, K. L. Johnston, M. Collier, L. Myhill, A. Cassidy, D. Plant, H. Plant, R. Clark, D. A. N. Cook, A. Steven, J. Archer, P. McGillan, S. Charoensutthivarakul, J. Bibby, R. Sharma, G. L. Nixon, B. E. Slatko, L. Cantin, B. Wu, J. Turner, L. Ford, K. Rich, M. Wigglesworth, N. G. Berry, P. M. O'Neill, M. J. Taylor, and S. A. Ward, "Industrial scale high-throughput screening delivers multiple fast acting macrofilaricides," *Nat Commun,* vol. 10, no. 1, pp. 11, Jan 2, 2019.

[10] Y. O. Adeshina, E. J. Deeds, and J. Karanicolas, "Machine learning classification can reduce false positives in structure-based virtual screening," *Proc Natl Acad Sci U S A,* vol. 117, no. 31, pp. 18477-18488, Aug 4, 2020.

[11] M. C. R. Melo, J. Maasch, and C. de la Fuente-Nunez, "Accelerating antibiotic discovery through artificial intelligence," *Commun Biol,* vol. 4, no. 1, pp. 1050, Sep 9, 2021.

[12] N. Tharmalingam, R. Rajmuthiah, W. Kim, B. B. Fuchs, E. Jeyamani, M. J. Kelso, and E. Mylonakis, "Antibacterial Properties of Four Novel Hit Compounds from a Methicillin-Resistant Staphylococcus aureus-Caenorhabditis elegans High-Throughput Screen," *Microb Drug Resist,* vol. 24, no. 5, pp. 666-674, Jun, 2018.

[13] L. P. O'Reilly, C. J. Luke, D. H. Perlmutter, G. A. Silverman, and S. C. Pak, "C. elegans in high-throughput drug discovery," *Adv Drug Deliv Rev,* vol. 69-70, pp. 247-53, Apr, 2014.

[14] W. Kim, W. Zhu, G. L. Hendricks, D. Van Tyne, A. D. Steele, C. E. Keohane, N. Fricke, A. L. Conery, S. Shen, W. Pan, K. Lee, R. Rajamuthiah, B. B. Fuchs, P. M. Vlahovska, W. M. Wuest, M. S. Gilmore, H. Gao, F. M. Ausubel, and E. Mylonakis, "A new class of synthetic retinoid antibiotics effective against bacterial persisters," *Nature,* vol. 556, no. 7699, pp. 103-107, Apr 5, 2018.

[15] R. Rajamuthiah, B. B. Fuchs, E. Jayamani, Y. Kim, J. Larkins-Ford, A. Conery, F. M. Ausubel, and E. Mylonakis, "Whole animal automated platform for drug discovery against multi-drug resistant Staphylococcus aureus," *PLoS One,* vol. 9, no. 2, pp. e89189, 2014.

[16] P. Ertl, B. Rohde, and P. Selzer, "Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties," *J Med Chem,* vol. 43, no. 20, pp. 3714-7, Oct 5, 2000.

[17] S. A. Wildman, and G. M. Crippen, "Prediction of Physicochemical Parameters by Atomic Contributions," *J Chem Inf Comput Sci,* vol. 39, no. 5, pp. 868-873, 1999.

[18] J. D. Hughes, J. Blagg, D. A. Price, S. Bailey, G. A. Decrescenzo, R. V. Devraj, E. Ellsworth, Y. M. Fobian, M. E. Gibbs, R. W. Gilles, N. Greene, E. Huang, T. Krieger-Burke, J. Loesel, T. Wager, L. Whiteley, and Y. Zhang, "Physiochemical drug properties associated with in vivo toxicological outcomes," *Bioorg Med Chem Lett,* vol. 18, no. 17, pp. 4872-5, Sep 1, 2008.

[19] S. D'Souza, K. V. Prema, and S. Balaji, "Machine learning models for drug-target interactions: current knowledge and future directions," *Drug Discov Today,* vol. 25, no. 4, pp. 748-756, Apr, 2020.

[20] S. Jaeger, S. Fulle, and S. Turk, "Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition," *J Chem Inf Model,* vol. 58, no. 1, pp. 27-35, 2018.

[21] H. L. Morgan, "The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service," *J Chem Doc,* vol. 5, no. 2, pp. 107-113, 2002.

[22] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[23] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, and R. G. Coleman, "ZINC: a free tool to discover chemistry for biology," *J Chem Inf Model,* vol. 52, no. 7, pp. 1757-68, Jul 23, 2012.

[24] N. Schaduangrat, N. Anuwongcharoen, P. Charoenkwan, and W. Shoombuatong, "DeepAR: a novel deep learning-based hybrid framework for the interpretable prediction of androgen receptor antagonists," *J Cheminform,* vol. 15, no. 1, pp. 50, May 6, 2023.

[25] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, no. 1, pp. 5-32, 2001/10/01, 2001.

[26] C. Chen, A. Liaw, and L. Breiman, "Using random forest to learn imbalanced data," *University of California, Berkeley,* vol. 110, no. 1-12, pp. 24, 2004.

[27] C. Rucker, G. Rucker, and M. Meringer, "y-Randomization and its variants in QSPR/QSAR," *J Chem Inf Model,* vol. 47, no. 6, pp. 2345-57, Nov-Dec, 2007.

[28] N. Schaduangrat, N. Homdee, and W. Shoombuatong, "StackER: a novel SMILES-based stacked approach for the accelerated and efficient discovery of ERalpha and ERbeta antagonists," *Sci Rep,* vol. 13, no. 1, pp. 22994, Dec 27, 2023.

[29] N. Schaduangrat, N. Anuwongcharoen, M. A. Moni, P. Lio, P. Charoenkwan, and W. Shoombuatong, "StackPR is a new computational approach for large-scale identification of progesterone receptor antagonists using the stacking strategy," *Sci Rep,* vol. 12, no. 1, pp. 16435, Sep 30, 2022.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *J Mach Learn Res,* vol. 12, pp. 2825-2830, Oct, 2011.

[31] G. Landrum, "RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling," Academic Press Cambridge, 2013.

[32] L. van der Maaten, and G. Hinton, "Visualizing Data using t-SNE," *J Mach Learn Res,* vol. 9, pp. 2579-2605, Nov, 2008.

[33] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," *J Chem Inf Model,* vol. 28, no. 1, pp. 31-36, 1988.

[34] S. Mondal, E. Hegarty, C. Martin, S. K. Gokce, N. Ghorashian, and A. Ben-Yakar, "Large-scale microfluidics providing high-resolution and high-throughput screening of Caenorhabditis elegans poly-glutamine aggregation model," *Nat Commun,* vol. 7, pp. 13023, Oct 11, 2016.

[35] G. H. S. Dreiman, M. Bictash, P. V. Fish, L. Griffin, and F. Svensson, "Changing the HTS Paradigm: AI-Driven Iterative Screening for Hit Finding," *SLAS Discov,* vol. 26, no. 2, pp. 257-262, Feb, 2021.

[36] S. Korkmaz, "Deep Learning-Based Imbalanced Data Classification for Drug Discovery," *J Chem Inf Model,* vol. 60, no. 9, pp. 4180-4190, Sep 28, 2020.

[37] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst Appl,* vol. 73, pp. 220-239, 2017.

[38] J. Bajorath, "Integration of virtual and high-throughput screening," *Nat Rev Drug Discov,* vol. 1, no. 11, pp. 882-94, Nov, 2002.

[39] S. Paricharak, I. J. AP, A. Bender, and F. Nigsch, "Analysis of Iterative Screening with Stepwise Compound Selection Based on Novartis In-house HTS Data," *ACS Chem Biol,* vol. 11, no. 5, pp. 1255-64, May 20, 2016.

[40] N. Malo, J. A. Hanley, S. Cerquozzi, J. Pelletier, and R. Nadon, "Statistical practice in high-throughput screening data analysis," *Nat Biotechnol,* vol. 24, no. 2, pp. 167-75, Feb, 2006.

[41] B. Gunter, C. Brideau, B. Pikounis, and A. Liaw, "Statistical and graphical methods for quality control determination of high-throughput screening data," *J Biomol Screen,* vol. 8, no. 6, pp. 624-33, Dec, 2003.

[42] T. Y. Shun, J. S. Lazo, E. R. Sharlow, and P. A. Johnston, "Identifying actives from HTS data sets: practical approaches for the selection of an appropriate HTS data-processing method and quality control review," *J Biomol Screen,* vol. 16, no. 1, pp. 1-14, Jan, 2011.

[43] J. H. Zhang, T. D. Chung, and K. R. Oldenburg, "A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays," *J Biomol Screen,* vol. 4, no. 2, pp. 67-73, 1999.

[44] D. T. Bearden, G. P. Allen, and J. M. Christensen, "Comparative in vitro activities of topical wound care products against community-associated methicillin-resistant Staphylococcus aureus," *J Antimicrob Chemother,* vol. 62, no. 4, pp. 769-72, Oct, 2008.

[45] M. K. Hayden, K. Lolans, K. Haffenreffer, T. R. Avery, K. Kleinman, H. Li, R. E. Kaganov, J. Lankiewicz, J. Moody, E. Septimus, R. A. Weinstein, J. Hickok, J. Jernigan, J. B. Perlin, R. Platt, and S. S. Huang, "Chlorhexidine and Mupirocin Susceptibility of Methicillin-Resistant Staphylococcus aureus Isolates in the REDUCE-MRSA Trial," *J Clin Microbiol,* vol. 54, no. 11, pp. 2735-2742, Nov, 2016.

[46] L. Xu, P. She, L. Chen, S. Li, L. Zhou, Z. Hussain, Y. Liu, and Y. Wu, "Repurposing Candesartan Cilexetil as Antibacterial Agent for MRSA Infection," *Front Microbiol,* vol. 12, pp. 688772, 2021.

[47] W. S. Yeo, R. Arya, K. K. Kim, H. Jeong, K. H. Cho, and T. Bae, "The FDA-approved anti-cancer drugs, streptozotocin and floxuridine, reduce the virulence of Staphylococcus aureus," *Sci Rep,* vol. 8, no. 1, pp. 2521, Feb 6, 2018.

[48] A. D. Sharma, and W. G. Gutheil, "Synergistic Combinations of FDA-Approved Drugs with Ceftobiprole against Methicillin-Resistant Staphylococcus aureus," *Microbiol Spectr*, pp. e0372622, Dec 15, 2022.

[49] C. Liu, A. Bayer, S. E. Cosgrove, R. S. Daum, S. K. Fridkin, R. J. Gorwitz, S. L. Kaplan, A. W. Karchmer, D. P. Levine, B. E. Murray, J. R. M, D. A. Talan, H. F. Chambers, and A. Infectious Diseases Society of, "Clinical practice guidelines by the infectious diseases society of america for the treatment of methicillin-resistant Staphylococcus aureus infections in adults and children," *Clin Infect Dis,* vol. 52, no. 3, pp. e18-55, Feb 1, 2011.

[50] C. Ster, V. Lebeau, J. Leclerc, A. Fugere, K. A. Veh, J. P. Roy, and F. Malouin, "In vitro antibiotic susceptibility and biofilm production of Staphylococcus aureus isolates recovered from bovine intramammary infections that persisted or not following extended therapies with cephapirin, pirlimycin or ceftiofur," *Vet Res,* vol. 48, no. 1, pp. 56, Sep 21, 2017.

[51] N. Pant, S. C. Wallis, J. A. Roberts, and D. P. Eisen, "In vitro effect of synovial fluid from patients undergoing arthroplasty surgery on MRSA biofilm formation," *J Antimicrob Chemother,* vol. 77, no. 4, pp. 1041-1044, Mar 31, 2022.

[52] L. Coppens, B. Hanson, and J. Klastersky, "Therapy of staphylococcal infections with cefamandole or vancomycin alone or with a combination of cefamandole and tobramycin," *Antimicrob Agents Chemother,* vol. 23, no. 1, pp. 36-41, Jan, 1983.

[53] K. Leder, J. D. Turnidge, T. M. Korman, and M. L. Grayson, "The clinical efficacy of continuous-infusion flucloxacillin in serious staphylococcal sepsis," *J Antimicrob Chemother,* vol. 43, no. 1, pp. 113-8, Jan, 1999.

[54] P. A. Giordano, D. Elston, B. K. Akinlade, K. Weber, G. F. Notario, T. A. Busman, M. Cifaldi, and A. M. Nilius, "Cefdinir vs. cephalexin for mild to moderate uncomplicated skin and skin structure infections in adolescents and adults," *Curr Med Res Opin,* vol. 22, no. 12, pp. 2419-28, Dec, 2006.

[55] M. G. Bergeron, and A. Turcotte, "Penetration of cefixime into fibrin clots and in vivo efficacy against Escherichia coli, Klebsiella pneumoniae, and Staphylococcus aureus," *Antimicrob Agents Chemother,* vol. 30, no. 6, pp. 913-6, Dec, 1986.

[56] S. Lemaire, K. Kosowska-Shick, P. C. Appelbaum, Y. Glupczynski, F. Van Bambeke, and P. M. Tulkens, "Activity of moxifloxacin against intracellular community-acquired methicillin-resistant Staphylococcus aureus: comparison with clindamycin, linezolid and co-trimoxazole and attempt at defining an intracellular susceptibility breakpoint," *J Antimicrob Chemother,* vol. 66, no. 3, pp. 596-607, Mar, 2011.

[57] R. A. Lowe, K. E. Barber, J. L. Wagner, A. M. Bell-Harlan, and K. R. Stover, "Ceftriaxone for the Treatment of Methicillin-susceptible Staphylococcus aureus Bacteremia: A Case Series," *J Pharmacol Pharmacother,* vol. 8, no. 3, pp. 140-144, Jul-Sep, 2017.

[58] K. E. Aldridge, "Cefotaxime in the treatment of staphylococcal infections. Comparison of in vitro and in vivo studies," *Diagn Microbiol Infect Dis,* vol. 22, no. 1-2, pp. 195-201, May-Jun, 1995.

[59] J. M. Blondeau, S. Borsos, and C. K. Hesje, "Antimicrobial efficacy of gatifloxacin and moxifloxacin with and without benzalkonium chloride compared with ciprofloxacin and levofloxacin against methicillin-resistant Staphylococcus aureus," *J Chemother,* vol. 19, no. 2, pp. 146-51, Apr, 2007.

[60] M. J. Karau, S. M. Schmidt-Malan, M. Albano, J. N. Mandrekar, C. G. Rivera, D. R. Osmon, C. P. Oravec, D. J. Berry, M. P. Abdel, and R. Patel, "Novel Use of Rifabutin and Rifapentine to Treat Methicillin-Resistant Staphylococcus aureus in a Rat Model of Foreign Body Osteomyelitis," *J Infect Dis,* vol. 222, no. 9, pp. 1498-1504, Oct 1, 2020.

[61] N. D. Gade, and M. S. Qazi, "Fluoroquinolone Therapy in Staphylococcus aureus Infections: Where Do We Stand?," *J Lab Physicians,* vol. 5, no. 2, pp. 109-12, Jul, 2013.

[62] Z. D. Jiang, and H. L. DuPont, "Rifaximin: in vitro and in vivo antibacterial activity--a review," *Chemotherapy,* vol. 51 Suppl 1, pp. 67-72, 2005.

[63] J. B. Rasmussen, J. D. Knudsen, M. Arpi, H. C. Schonheyder, T. Benfield, and C. Ostergaard, "Relative efficacy of cefuroxime versus dicloxacillin as definitive antimicrobial therapy in methicillin-susceptible Staphylococcus aureus bacteraemia: a propensity-score adjusted retrospective cohort study," *J Antimicrob Chemother,* vol. 69, no. 2, pp. 506-14, Feb, 2014.

[64] National Center for Biotechnology Information. "PubChem Bioassay Record for AID 720641, Source: ICCB-Longwood Screening Facility, Harvard Medical School," January 30, 2023.

[65] J. E. Westhead, "Semiautomated turbidimetric bioassay for the ionophore A23187," *Antimicrob Agents Chemother,* vol. 11, no. 5, pp. 916-8, May, 1977.

[66] G. A. J. Ayliffe, M. A. Buckles, M. W. Casewell, B. D. Cookson, R. A. Cox, G. J. Duckworth, G. L. French, M. A. Griffiths-Jones, R. Heathcock, H. Humphreys, C. T. Keane, R. R. Marples, D. C. Shanson, R. Slack, and E. Tebbs, "Revised guidelines for the control of methicillin-resistant Staphylococcus aureus infection in hospitals: Report of a combined working party of the British Society for Antimicrobial Chemotherapy, the Hospital Infection

Society and the Infection Control Nurses Association," *J Hosp Infect,* vol. 39, no. 4, pp. 253-290, 1998/08/01/, 1998.

[67] J. Y. Jacobs, Y. Wax, J. Michel, and T. Sacks, "Synergism between gentamicin and mitomycin C in staphylococcal infections in mice," *Chemotherapy,* vol. 31, no. 5, pp. 389-94, 1985.

[68] A. Asadipour, M. H. Moshafi, L. Khosravani, S. Moghimi, E. Amou, L. Firoozpour, G. Ilbeigi, K. Beiki, E. Soleimani, and A. Foroumadi, "N-substituted piperazinyl sarafloxacin derivatives: synthesis and in vitro antibacterial evaluation," *Daru,* vol. 26, no. 2, pp. 199-207, Dec, 2018.

[69] C. Rayner, and W. J. Munckhof, "Antibiotics currently used in the treatment of infections caused by Staphylococcus aureus," *Intern Med J,* vol. 35 Suppl 2, pp. S3-16, Dec, 2005.

[70] T. J. Walsh, H. C. Standiford, A. C. Reboli, J. F. John, M. E. Mulligan, B. S. Ribner, J. Z. Montgomerie, M. B. Goetz, C. G. Mayhall, D. Rimland, and et al., "Randomized double-blinded trial of rifampin with either novobiocin or trimethoprim-sulfamethoxazole against methicillin-resistant Staphylococcus aureus colonization: prevention of antimicrobial resistance and effect of host factors on outcome," *Antimicrob Agents Chemother,* vol. 37, no. 6, pp. 1334-42, Jun, 1993.

[71] M. Lehtinen, E. Pelttari, and H. Elo, "Antimicrobial activity of formylchromones: detection by a micro-scale method," *Z Naturforsch C J Biosci,* vol. 66, no. 11-12, pp. 562-70, Nov-Dec, 2011.

[72] R. Khader, N. Tharmalingam, B. Mishra, L. Felix, F. M. Ausubel, M. J. Kelso, and E. Mylonakis, "Characterization of Five Novel Anti-MRSA Compounds Identified Using a Whole-Animal Caenorhabditis elegans/Galleria mellonella Sequential-Screening Approach," *Antibiotics (Basel),* vol. 9, no. 8, Jul 27, 2020.

[73] M. Giannouli, A. T. Palatucci, V. Rubino, G. Ruggiero, M. Romano, M. Triassi, V. Ricci, and R. Zarrilli, "Use of larvae of the wax moth Galleria mellonella as an in vivo model to study the virulence of Helicobacter pylori," *BMC Microbiol,* vol. 14, pp. 228, Aug 27, 2014.

[74] K. A. Scott, and J. T. Njardarson, "Analysis of US FDA-Approved Drugs Containing Sulfur Atoms," *Top Curr Chem (Cham),* vol. 376, no. 1, pp. 5, Jan 22, 2018.

[75] Y. Genc, R. Ozkanca, and Y. Bekdemir, "Antimicrobial activity of some sulfonamide derivatives on clinical isolates of Staphylococus aureus," *Ann Clin Microbiol Antimicrob,* vol. 7, pp. 17, Aug 20, 2008.

[76] C. R. Mizdal, S. T. Stefanello, V. da Costa Flores, V. A. Agertt, P. C. Bonez, G. G. Rossi, T. C. da Silva, F. A. Antunes Soares, L. de Lourenco Marques, and M. M. A. de Campos, "The antibacterial and anti-biofilm activity of gold-complexed sulfonamides against methicillin-resistant Staphylococcus aureus," *Microb Pathog,* vol. 123, pp. 440-448, Oct, 2018.

[77] X. F. Liu, C. J. Zheng, L. P. Sun, X. K. Liu, and H. R. Piao, "Synthesis of new chalcone derivatives bearing 2,4-thiazolidinedione and benzoic acid moieties as potential anti-bacterial agents," *Eur J Med Chem,* vol. 46, no. 8, pp. 3469-73, Aug, 2011.

[78] K. Sena, R. F. V. Mendes, E. X. Botelho, R. O. Araujo-Melo, C. J. A. Silva, H. N. P. Costa Junior, B. Amorim-Carmo, I. Z. Damasceno, M. F. Fernandes-Pedrosa, J. S. Aguiar, T. G. Silva, G. M. S. Lima, J. F. C. Albuquerque, and R. M. Ximenes, "Antibacterial and antibiofilm activities of thiazolidine-2,4-dione and 4-thioxo-thiazolidin-2-one derivatives against multidrug-resistant Staphylococcus aureus clinical isolates," *J Appl Microbiol,* vol. 133, no. 6, pp. 3558-3572, Dec, 2022.

[79] H. A. Aziz, A. M. M. El-Saghier, M. Badr, G. E. A. Abuo-Rahma, and M. E. Shoman, "Thiazolidine-2,4-dione-linked ciprofloxacin derivatives with broad-spectrum antibacterial, MRSA and topoisomerase inhibitory activities," *Mol Divers,* vol. 26, no. 3, pp. 1743-1759, Jun, 2022.

[80] O. Bozdag-Dundar, O. Ozgen, A. Mentese, N. Altanlar, O. Atli, E. Kendi, and R. Ertan, "Synthesis and antimicrobial activity of some new thiazolyl thiazolidine-2,4-dione derivatives," *Bioorg Med Chem,* vol. 15, no. 18, pp. 6012-7, Sep 15, 2007.

[81] M. C. Phillips, N. Wald-Dickler, K. Loomis, B. M. Luna, and B. Spellberg, "Pharmacology, Dosing, and Side Effects of Rifabutin as a Possible Therapy for Antibiotic-Resistant Acinetobacter Infections," *Open Forum Infect Dis,* vol. 7, no. 11, pp. ofaa460, Nov, 2020.

[82] J. B. Doub, E. L. Heil, A. Ntem-Mensah, R. Neeley, and P. R. Ching, "Rifabutin Use in Staphylococcus Biofilm Infections: A Case Series," *Antibiotics (Basel),* vol. 9, no. 6, Jun 13, 2020.

[83] H. F. Chambers, "Rifabutin to the Rescue?," *J Infect Dis,* vol. 222, no. 9, pp. 1422-1424, Oct 1, 2020.