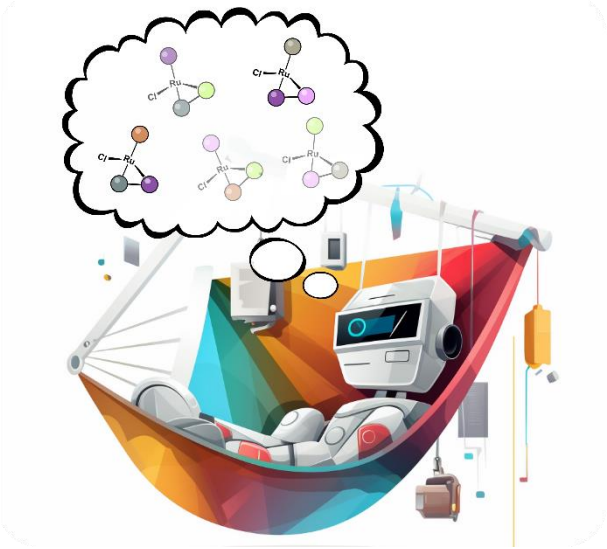
Using Machine Learning to Predict the Antibacterial Activity of Ruthenium Complexes

Markus Orsi[#a], Boon Shing Loh[#b], Dr. Cheng Weng,[#b] Prof. Wee Han Ang[b,c], Dr. Angelo Frei*[a]*

[a]Department of Chemistry, Biochemistry & Pharmaceutical Sciences, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland

[b]Department of Chemistry, National University of Singapore, 4 Science Drive 2, Singapore 117544, Singapore

[c]NUS Graduate School – Integrated Science and Engineering Programme (ISEP), National University of Singapore, 21 Lower Kent Ridge Rd, Singapore 119077, Singapore



Abstract

Rising antimicrobial resistance (AMR) and lack of innovation in the antibiotic pipeline necessitate novel approaches to discovering new drugs. Metal complexes have proven to be promising antimicrobial compounds, but the number of studied compounds is still low compared to the millions of organic molecules investigated so far. Lately, machine learning (ML) has emerged as a valuable tool for guiding the design of small organic molecules, potentially even in low-data scenarios. For the first time, we extend the application of ML to the discovery of metal-based medicines. Utilising 288 modularly synthesized ruthenium arene Schiff-base complexes and their antibacterial properties, a series of ML models were trained. The models perform well and are used to predict the activity of 54 new compounds. These displayed a 5.7x higher hit-rate (53.7%) against methicillin-resistant *Staphylococcus aureus* (MRSA) compared to the original library (9.4%), demonstrating that ML can be applied to improve the success-rates in the search of new metalloantibiotics. This work paves the way for more ambitious applications of ML in the field of metal-based drug discovery.

Introduction

Drug-resistant bacterial infections are already causing 1.3 million deaths per year world-wide.[1] At the same time antibiotic use has increased during the COVID-19 pandemic[2,3] and is still on the rise in agriculture.[4] Yet against this backdrop, most pharmaceutical companies have shut down their antibiotic development programs. This is reflected in a meagre clinical pipeline with only a small number of compounds with novel modes of action in clinical trials, even though there has been a slight but promising increase in recent years.[5] As the conventional sources for antibiotics, i.e. natural products and small organic molecules are not enough to contain the worsening antimicrobial resistance (AMR) problem, novel approaches are urgently needed.

Transition metal complexes have made a lasting impact in medicine with the platinum-based anticancer drugs constituting some of the most effective chemotherapeutic cancer regimens in the clinic today.[6,7] Over the last decade, more and more studies have highlighted the promising antimicrobial properties of metal complexes.[8–11] The recent progress in this field has been highlighted in several review articles.[12,13] We have recently shown that metal complexes have superior hit-rates against critical bacteria[14] and fungi[15] *in vitro* compared to purely organic molecules. We could show that metal complexes did not possess higher rates of cytotoxicity or haemolysis against human cells compared to their organic counterparts.[14] Ruthenium complexes in particular seem to have promising antibacterial properties, with one compound family currently undergoing preclinical evaluations.[9,16,17] However, at this stage the vast transition metal chemical space remains largely uncharted, particularly in comparison with the vast screening campaigns that have been performed on organic molecules.

Machine-learning (ML) approaches have been applied to different facets of the drug discovery process, including the prediction of bioactive compounds. In 2020, Stokes *et al.* reported the application of deep learning to the prediction of antibacterial compounds. The authors trained a deep neural network on *Escherichia coli* growth inhibition data of 2,335 unique compounds. The library comprised a large portion of US Food and Drug Administration (FDA) approved compounds as well as natural products. With the trained model, the authors identified halicin from the Drug Repurposing Hub as an antibacterial compound and verified its activity both *in vitro* and *in vivo.*[18] A similar approach was utilized to discover the antibiotic abaucin which displayed narrow-spectrum activity against *Acinetobacter baumanii* both *in vitro* and *in vivo*, highlighting that ML models can be utilized to predict both broad-spectrum as well as narrow-spectrum antibiotics.[19] Recently, Capecchi *et al.* have shown that ML can also be utilised to predict non-hemolytic antimicrobial peptides.[20] Despite these encouraging examples, their numbers are still sparse and they rely on available databases such as already FDA approved drugs or collections such as the Drug Repurposing Hub, limiting their potential to discover entirely new compounds.

The application of ML approaches to transition metal complexes has been even slower as the number of curated datasets is hitherto very limited. The Kulik group has made notable developments towards the application of ML towards the efficient search of

vast chemical spaces for optimised transition metal complexes.[21–24] Balcells *et al.* recently described a new representation for deep graph learning on transition metal complexes.[25] In the field of homogeneous catalyst development, the group of Corminboeuf reported the application of genetic optimisation as a tool to accelerate catalyst discovery.[26]

An additional difficulty when dealing with metal complexes is that standard string representations such as SMILES tend to not be generally applicable. However, many cheminformatics tools rely on these representations to translate molecules into a computer-readable format that can be utilised for training of ML models.

To tackle these challenges, we have opted to utilize systematic data generated inhouse, maximising its robustness. Secondly, we have focused this study on a single metal scaffold. Herein we report the application of ML to the prediction of antibacterial ruthenium complexes. We present a new, yet simple fingerprint that describes these metal complexes sufficiently for ML models to be trained on antimicrobial activity data. We then use these models to predict novel active ruthenium complexes from a virtual library of over 70 million possible compounds. We validate the ML approach by acquiring and synthesizing a selection of predicted compounds showing that we could significantly increase our hit-rate through the ML-guided building block selection to discover entirely novel antibacterial ruthenium complexes.

Results and Discussion

In ongoing work we have prepared a combinatorial library of novel ruthenium-arene Schiff-based complexes and evaluated their antimicrobial properties.[27] This was based on an combinatorial synthesis approach described by the Ang group in earlier work.[28–31]

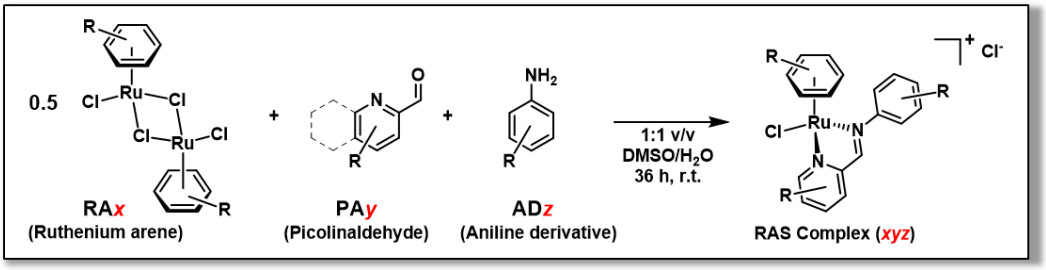


Figure 1. General reaction scheme for the synthesis of combinatorial ruthenium arene Schiff-base complexes.

By combining 6 picolinaldehydes with 12 aniline derivatives and 4 ruthenium-arene precursors, 288 novel compounds could be prepared (Figure 1). All 288 compound crudes were directly screened through the CO-ADD initiative[32] for their ability to inhibit microbial growth against a panel of ESKAPE pathogens (*Escherichia coli*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, MRSA) and two fungi (*Cryptococcus neoformans* (yeast) and *Candida albicans*). No significant inhibition was found against the Gram-negative bacteria and only three compounds showed some inhibition of *C. neoformans*. However, 27 compounds (9.4%) showed significant growth inhibition against Gram-positive MRSA at 20 µM. For 16 (5.6%) of these a minimum inhibitory concentration (MIC) <20 µM could be determined (Figure

2). Further investigation of lead compounds is currently underway and will be reported separately.

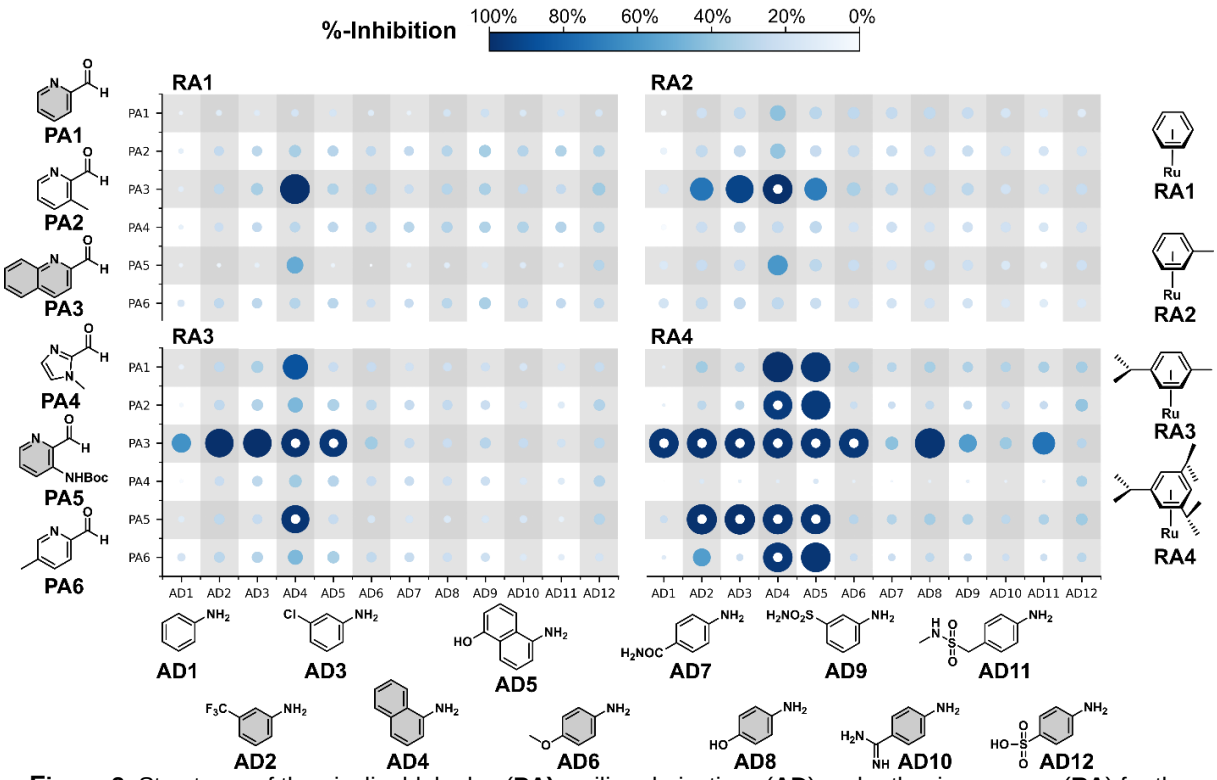


Figure 2. Structures of the picolinaldehydes (**PA**), aniline derivatives (**AD**) and ruthenium arenes (**RA**) for the combinatorial library and the MRSA growth inhibition-% data at 20 µM given for the 288 tested compounds (white dots indicate a confirmed MIC < 20 µM, for specific MICs see Table S1).

It is evident that even though we have prepared 288 complexes with high ease and low costs, this is barely scratching the surface of the possible compounds that could be made. To construct a more robust library towards antimicrobial efforts without blindly making more compounds, a more guided approach could be applied to the data obtained and direct future synthetic explorations.

While ML approaches have gained widespread attention and application in organic drug discovery, they have not been widely applied to metal complexes yet. One reason is that most conventional methods to generate descriptors or feature vectors for molecules rely on molecular representations, such as SMILES, which cannot be easily extrapolated onto metal complexes with multiple coordinating ligands. While some solutions have been proposed, they have not been widely applied yet.[33,34] In our case we took advantage of the fact that all 288 tested compounds and any compound of this class we wished to predict shared several similarities. All compounds contain ruthenium(II), have a chlorido ligand and in an approximation adopt the same pseudo-'piano stool' geometry. We therefore hypothesized that it would be sufficient to represent each ruthenium complex as a linear combination of the molecular fingerprints of its ligand components (Figure 3). Since each component (Ru-arene, aniline-derivative and picolinaldehyde) is a conventional organic molecule, we were able to use available python libraries to generate fingerprint vectors. In our case, we opted for the RDKit[35] implementation of the widely applied extended-connectivity fingerprints (ECFP4[36]). With regards to the fingerprint length, we opted for 512-bits

as a trade-off between encoding efficiency and feature-size, since we are operating in a low-data regime.

Initially we considered a XOR-type combination of the three fingerprints, resulting in a single vector where all the bits that were activated in the fingerprints of the three components remained activated in the final vector. However, we decided against this representation as some information would be lost in the final vector if the same bit is active in two or more of the building blocks. We settled on the second approach, which entailed combining fingerprints through bit-wise summation. This method allowed us to retain more information about the three distinct components in the final vector. Essentially, the fingerprint indicates the presence of any circular substructure and how many components contain the substructure. We consider this to be a more appropriate representation for the combinatorial library, as it allows the models to capture whether certain moieties need to be present in one or more components.

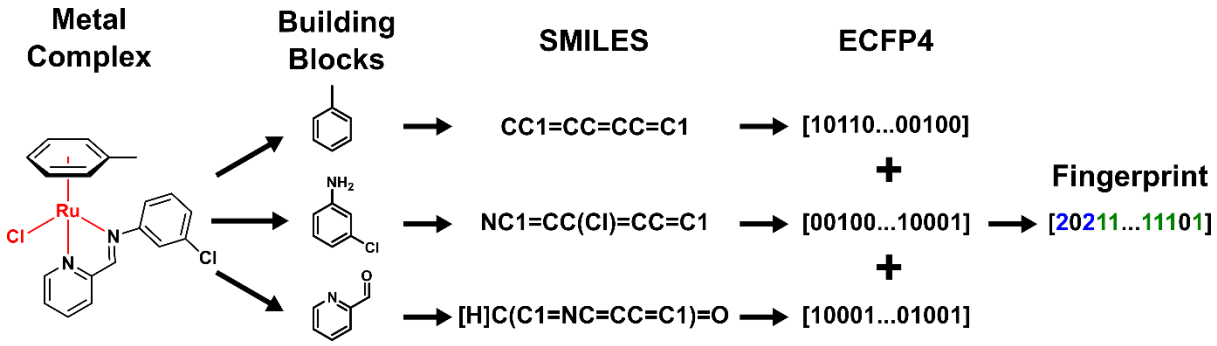


Figure 3. Schematic illustration of how fingerprints for the ruthenium-arene Schiff-base complexes were generated.

As we only had a limited dataset of 288 compounds, we converted the percentage inhibition data into binary classification data. Taking advantage of the more detailed dose-response data allowed us to exclude some false positives from the actives, leaving 16 compounds with confirmed activity. All compounds were either labelled inactive (‘0’) or active (‘1’). With the fingerprints and labels in hand, we evaluated six classical machine learning models in a 10-fold cross validation utilizing the scikit-learn package. The models used were: Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGB) and k-Nearest Neighbours (kNN). As the dataset is highly unbalanced, a model guessing all compounds to be inactive would still give accurate predictions 90% of the time. To make sure the models are not learning the underlying data distribution, we rerun the 10-fold cross validations with scrambled labels for all models. Additionally, we paid close attention to the performance of the models with respect to the generation of false-negatives and false-positives (confusion matrices for all models are provided in the SI).

In general, apart from Naïve Bayes, all models demonstrated comparable performance in the classification task, indicating that the utilisation of a linear combination of ligand fingerprints is a suitable proxy for the encoding of the entire metal complex in our specific use-case. The best training results were obtained for the SVM and MLP algorithms, with a mean AUC of 0.98 (Figure 4, and Figure S1-S6). Both models also showed very low numbers of false-positives and false-negatives. At

the same time, the models performed no better than random when scrambled labels were used, indicating that they did not just replicate the underlying data distribution.

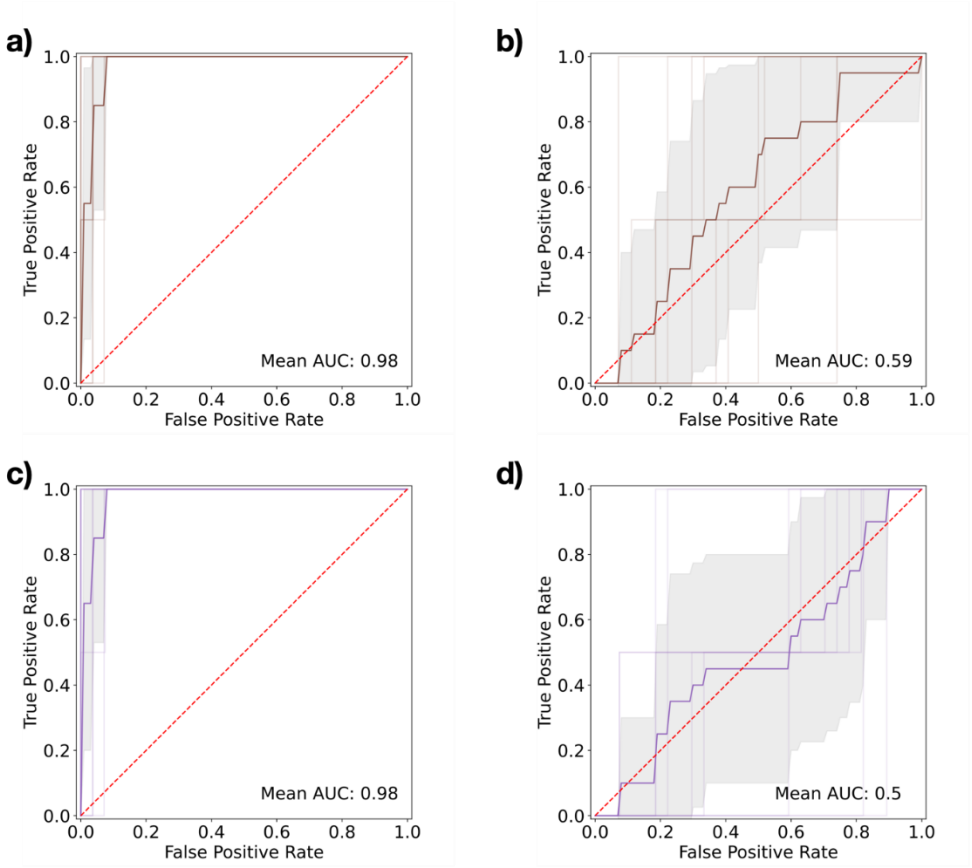


Figure 4. Performance evaluation of selected models. a) AUC of SVM model for all 10 cross-validations. The mean AUC of all cross-validations is displayed in brown; the standard deviation interval is displayed in grey. b) AUCs for the SVM model with scrambled labels. c) AUC of MLP model for all 10 cross-validations. The mean AUC of all cross-validations is displayed in purple; the standard deviation interval is displayed in grey. d) AUCs for the MLP model with scrambled labels.

To evaluate the accuracy and usefulness of these models, we created a virtual library of possible complexes to synthesize next. To assemble this library, we conducted targeted substructure searches on the curated chemistry database Reaxys. The goal was to generate a library of building blocks that can, in principle, be commercially acquired. To this end all results were filtered for 'commercial availability' during the searches. As the ruthenium arene precursor was the most difficult to synthesize and not much variety was available, we selected only 14 arenes for the virtual library. In total, 864 commercially available picolinaldehydes were included. In the case of aniline-derivatives, the results were further filtered for a molecular weight smaller than 200 Da, leaving 6356 amines. The full lists of the selected building blocks are available in our GitHub repository.[37] Altogether, these building blocks generated a virtual library containing 76,906,368 possible ruthenium complexes. We now utilized our trained SVM and MLP models to evaluate the virtual library, saving only compounds which were predicted to be active by both models (Figure S7). This left 2,299,553 possible ruthenium compounds that, according to our trained ML models, were likely to be active against MRSA.

Breakdown of the predicted actives into their respective building blocks showed that some building blocks were clearly favoured for activity over others (The ranked lists of building block frequencies are provided in the GitHub repository[37]). To investigate whether the ML might just correlate simple properties of the molecules such as logP, topological polar surface area (TPSA) or molecular weight with activity, we calculated these properties for all ranked building blocks. No correlation between rank/frequency and the properties logP, TPSA and molecular weight was found (Figure S8-S13). The feature importance was also analysed from the support vectors of SVM. Oxygen and chloro-substituents were shown to contribute positively to activity while sulfonyl and amidine groups such as the ones on **AD9-12** overall contributed negatively to activity in the SVM model (Figure S14-S15).

To verify the ML model predictions, we aimed to synthesize a small set of the compounds predicted to be active. As the **RA4** ruthenium-arene building block seemed to be highly favored over the others, we kept it constant and selected different picolinaldehydes (**PA**) and aniline-derivative (**AD**) building blocks. Starting from the most frequent building blocks, commercial suppliers were searched that provided a given building block at a reasonable price-point (at least 250 mg at <200 USD) and within a realistic time-period. Finally, 6 new picolinaldehydes (**nPA1**-**nPA6**) and 9 new aniline derivatives (**nAD1**-**nAD9**) were obtained. The new building blocks differ significantly from the original library in structure with a mean Tanimoto similarity of 0.22 ± 0.07 between the two **AD** groups and 0.27± 0.1 between the **PAs** (Figure S16-S17)

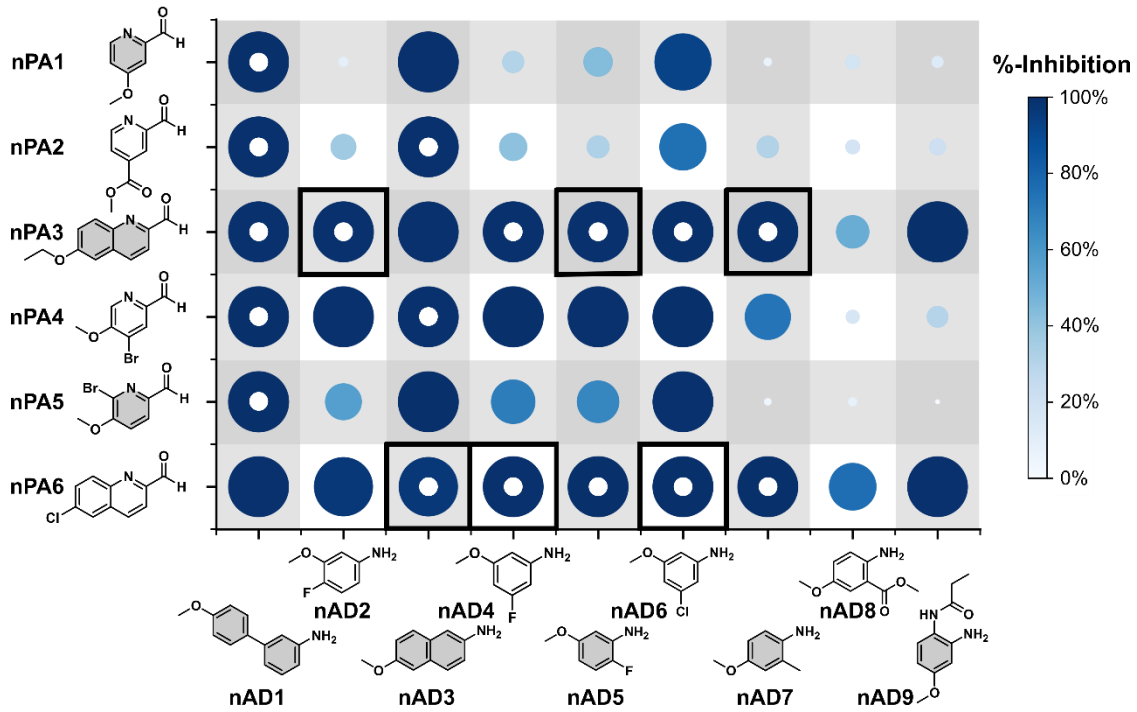


Figure 5. Structures of new picolinaldehydes (**nPA**) and aniline derivatives (**AD**) utilised together with **RA4** to prepare the ML-predicted combinatorial library with MRSA growth inhibition-% for the new library at 20 µM (white dots indicate confirmed complete growth inhibition at 5 µM.) The six compounds resynthesized for verification are highlighted with black squares.

With these building blocks in hand, we assembled a new small combinatorial library of 54 novel ruthenium(II) arene Schiff-base complexes. The reaction progress was

monitored by LC-MS, confirming that the target complexes had formed (Table S1). The average assembly yield was 66 ± 21%, indicating that in most crudes the putative complex was the major species. The crude compounds where then tested for their antimicrobial activity against both a methicillin-susceptible *S. aureus* (MSSA) as well as a MRSA strain. At 20 µM, 29/54 (53.7%) showed significant growth inhibition against MRSA, representing a 5.7x higher hit-rate compared to the initial 'blind' testing (27/288 or 9.4% active, Figure 5 and Table S2-S3). When screened at 5 µM, 17/54 (31.5%) still showed complete growth inhibition. When one considers that considers that only 16 out of the 27 active compounds in the original library showed an MIC ≤10 µM (Table S4), the improvement in the predicted library is even better. The activity-levels against MSSA were the same (29/54, 53.7% showing growth inhibition at 20 µM, Table S5). Even if one only compared the activity rates of compounds containing arene **RA4** between the old and new library, a significant 3.2x improvement was evident, increasing from 12/72 (16.7%) to the aforementioned 29/54 (53.7%). None of the new building blocks showed any bacterial growth inhibition by themselves at 20 µM (Table S6) supporting the conclusion that any observed activity stemmed from the putative ruthenium complexes.

As a final verification of these results, we selected 6 RAS complexes within the new library for targeted synthesis so that the isolated complexes could be directly evaluated for their antibacterial efficacies. To challenge the validity of the ML-predicted compounds, we chose RAS combinations that showed activity at 5 µM against MRSA but displayed the lowest assembly yields (Figure 5, black squares). As a negative control we choose **RA4-PA6-AD10** from the original library which showed no bacterial growth inhibition. These compounds were re-synthesized on batch scale, purified by flash silica column chromatography and characterised by ^{1}H, ^{13}C-NMR, HPLC and HR-MS (cf. Supporting Information), verifying their purity (>92%) and identity. With the purified compounds in hand, their antibacterial activity was determined against both MRSA and MSSA strains (Figure S18-19, and Table 1). All 6 ruthenium compounds that showed activities in the crude screening exhibited high efficacies against both bacterial strains with MIC values as low as 0.24 µM, which are comparable, and in some cases better than the standard-of-care vancomycin. The negative control compound did not show any antibacterial activity at the highest tested concentration. These results not only provide further evidence that our ML-models are indeed able to predict bioactive metal complexes but also support the crude-screening approach given that even in cases where the assembly yield is low, the antibacterial activity of the putative compound can be detected reliably.

Table 1. Minimum inhibitory concentration (MIC) of selected purified RAS complexes against MRSA 1768 and MSSA RN4220 in µM.

Compound[1]	MRSA 1768	MSSA RN4220
RA4-nPA3-nAD2	2.41	2.41
RA4-nPA3-nAD5	0.51	0.51
RA4-nPA3-nAD7	1.08	1.08
RA4-nPA6-nAD3	0.49	0.24
RA4-nPA6-nAD4	2.13	2.13
RA4-nPA6-nAD6	0.98	1.97
RA4-PA6-AD10 (negative control)	>43.5	>43.5
Ampicillin	25	0.2
Vancomycin	0.78	0.78

[1]Each experiment were performed with 3 biological replicates per concentration and at least 2 independent experiments were carried out. Final concentrations were adjusted based on ICP-OES measurements of stock solutions. MIC value is indicated by vertical dotted line in the dose-response graph (Figure S18-S19).

In conclusion, we have demonstrated for the first time the application of ML for the prediction of bioactive metal complexes. A linear combination of the building blocks' ECFP fingerprints was successfully utilized as a descriptor for a library of ruthenium complexes. The trained ML models enabled us to screen close to 77 million possible compounds and narrowed our search to ~2 million which the two best models predicted to be active. Further filtering by building block frequency and commercial availability led to the acquisition of 15 new building blocks and synthesis of a small new library. We found a significantly improved rate of antibacterial activity in the ML-predicted library compared to the initial one. The activity of the predicted compounds was verified further by resynthesizing and purifying six metal complexes. These isolated compounds showed low MICs comparable with currently employed antibiotics. This is but the exciting first step into the application of ML methods to the discovery of bioactive metal compounds. With this successful proof of concept, the door is open for future work with larger compound libraries and higher resolution models able to predict more precise activity levels but also other properties such as toxicity, haemolysis, solubility, stability etc. Better descriptors that include the properties of the transition metal will be needed to train models that are able to predict molecular properties across the periodic table.

Acknowledgements

A.F. thanks Prof. J-L. Reymond for generously hosting and supporting our research group and this project. A.F. gratefully acknowledges funding from the Swiss National

Science Foundation Ambizione grant PZ00P2_202016 and support by the University of Bern. Antimicrobial screening was partially performed by CO-ADD (Community for Open Antimicrobial Drug Discovery), funded by the Wellcome Trust (UK) and The University of Queensland (Australia). M.O. is grateful for support by an European Research Council Grant (885076). W.H.A. acknowledges support from Ministry of Education and National University of Singapore (A-0004134-00-00 and A-0004539-00-00).

Author contributions

A.F conceived the project. W. H. A. developed the combinatorial chemistry approach. M.O. and A. F. prepared, processed and evaluated the data and wrote the code for the ML. L. B. S. and W.C. performed the synthesis and characterisation of the libraries and the antibacterial testing of the predicted library. A.F. composed the manuscript. All authors discussed, commented and approved the final manuscript.

Conflicts of interest

There are no conflicts to declare.

Data Availability

All code and data used for training is available on the Frei Lab GitHub repository[37]

Materials and Methods

^{1}H, $^{13}C\{^{1}H\}$, $^{19}F\{^{1}H\}$ NMR spectra were obtained using either Bruker Avance III HD 400 or Bruker AVNEO 500 spectrometer, and the chemical shifts (δ) are reported in parts per million (ppm) with reference to residual solvent peaks.

The HPLC was Shimadzu Prominence System equipped with a DGU-20A_3 degasser, two LC-20AD liquid chromatography pumps, a SPD-20A UV/vis detector, and an Agilent, ZORBAX Eclipse Plus C18 column (4.6 × 150 mm, 5 µM) with flow rate of 1.0 mL/min with 254 and 280 nm detection wavelength. The gradient elution conditions were 20-95% of solvent B (MeCN with 0.1% trifluoroacetic acid) and solvent A (H_2O with 0.1% trifluoroacetic acid) over a 30 min elution period.

Electrospray-ionization mass spectrometry (ESI-MS) spectra were obtained using a Thermo Finnigan MAT ESI-MS System.

Determination of Ru concentration of stock solution was made using Perkin Elmer Avio 500 Inductively Coupled Plasma-Optical Emission Spectrometer (ICP-OES) operated by CMMAC, NUS, using internal indium standards at 0.5 ppm analyzed at 325 nm wavelength.

Dataset preparation

The set for training and evaluation of the machine learning models was generated starting from the components (anilines, picolinaldehydes, Ru-arenes) of the original combinatorial library. For all components, the ECFP4 fingerprints were calculated as 512-bit sized vectors using the RDKit (2022.3.4) implementation. All possible combinations of the obtained fingerprints were then generated by bit-wise addition for a total of 288 molecular fingerprints. Finally, the 288 molecular fingerprints were labelled based on their experimental antimicrobial activity (0 inactive, 1 active).

The screening library was generated in a similar fashion by combining the ECFP4 fingerprints of 14 Ru-arenes, 864 picolinaldehydes and 6356 anilines, for a total of 76,906,368 combinations.

Model evaluation

Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGB) and k-Nearest Neighbours (kNN) models were implemented using the python scikit-learn (0.22.1) package. All models were evaluated in a 10-fold cross validation with a train/test split of 0.75/0.25 and performances compared with scrambled labels to check for meaningful fitting. The mean area under the receiving operating characteristic curve (ROC-AUC) along all cross validations was used to evaluate and select the best models for prediction on the large screening library.

Library Synthesis

All experiments were carried out without exclusion of moisture and air. All chemicals and solvents were obtained from commercial sources without further treatment. All chemicals and solvents at analytical grade or high-performance liquid chromatography (HPLC) grade were purchased from commercially available sources. $RuCl_3 \cdot xH_2O$ precursor was purchased from Precious Metal Online. $[(\eta^6\text{-}1,3,5\text{-Triisopropylbenzene})RuCl_2]_2$ was synthesized according to previously reported protocols.[28]

Separate stock solutions containing $[(\eta^6\text{-}1,3,5\text{-Triisopropylbenzene})RuCl_2]_2$ dimer (10 mM), picolinaldehyde **PA** (40 mM), and aniline derivatives **AD** (40 mM) were prepared in DMSO. The reactions were then carried out on a 96-well flat-bottom plate (Greiner) with sequential addition of H_2O (100 µL), **PA** (25 µL), **AD** (25 µL) and **RA** (50 µL), added to each well in one portion, yielding RAS complexes (5 mM) in DMSO/H_2O (1:1 v/v, 200 µL). The plates were sealed and incubated with shaking at room temperature for 36 h.

To ascertain assembly efficiency, RAS complexes (50 µL) were freeze dried and reconstituted in 0.9 % (w/v) NaCl solution and characterized using LC-MS Thermo Scientific Vanquish HPLC-PDA tandem Thermo Scientific LTQ XL instrument and Thermo Xcaliber Qual Browser software. The acquisition method was obtained with ZORBAX Eclipse Plus C18 column (4.6 × 150 mm, 5 µM) with flow rate of 1.0 mL/min using solvent B (MeCN with 0.1% formic acid) and solvent A (H_2O with 0.1% formic

acid). The gradient elution conditions were 20 → 80 % solvent B over 20 min followed by a consistent 80% solvent B for 10 min.

Single Concentration Antibacterial Testing

All antimicrobial susceptibility testing followed clinical laboratory standards institute (CLSI) guideline.[38] Methicillin-susceptible *Staphylococcus aureus* (MSSA) RN4220, methicillin-resistant *Staphylococcus aureus* (MRSA) BAA-1768 were first inoculated in Lysogency broth and incubated at 37 °C for 4-6 h. Afterwards, bacterial cells were inoculated into 96-well flat-bottom plates (Greiner) with a density of 5×10^4 CFU/100 µL per well, followed by 5 µL addition of stock solution of assembled RAS complexes into 100 µL per well Mueller-Hinton II broth to obtain the desired corresponding concentration. 5 µL of 1:1 v/v DMSO/H_2O were added for blank controls and Mueller-Hinton II broth was added for background subtraction. The plate was sealed and incubated in a shaker at 37 °C for 16 h. The viability was determined by absorbance at OD_{600} using a microplate reader.

Determination for Minimum Inhibitory Concentration (MIC)

All antimicrobial susceptibility testing followed clinical laboratory standards institute (CLSI) guideline.[38] Methicillin-susceptible *Staphylococcus aureus* (MSSA) RN4220, methicillin-resistant *Staphylococcus aureus* (MRSA) BAA-1768 were first inoculated in Lysogency broth and incubated at 37 °C for 4-6 h. Afterwards, bacterial cells were inoculated into 96-well flat-bottom plates (Greiner) with a density of 5×10^4 CFU/100 µL per well. The RAS complexes were prepared in DMSO stock and serially diluted across the wells to obtain a range of concentrations in 100 µL per well Mueller-Hinton II broth. Ru concentrations of RAS complexes stock solutions were determined by ICP-OES. Blanks using Mueller-Hinton II broth was added for background subtraction. Antibiotics ampicillin and vancomycin were included for comparison purposes. The plate was sealed and incubated in a shaker at 37 °C for 16 h. The viability was determined by absorbance at OD_{600} using a microplate reader. The MIC was determined by the lowest concentration that induced no bacterial growth (as indicated on the graph with a vertical line).

General procedure for the synthesis and purification of RAS complex. The picolinaldehyde **PA** (1.2 equiv) and aniline derivatives **AD** (1.2 equiv) were added to MeOH (10 mL) and stirred at r.t. over 24 h. Next, the Ru dimer **RA4** (0.5 equiv) in MeOH (10 mL) was added to the reaction mixture for reflux over 24 h. The solvent was removed in vacuo, and the resultant crude product was purified by flash column chromatography (EtOH/$CHCl_3$) to afford the desired RAS complex.

RA4-nPA3-nAD5. Yield: 23.3 mg (60.8 %). ^{1}H NMR (400 MHz, DMSO-d_6): δ 0.92 (9H, d, J=6.80 Hz), 1.17 (9H, d, J=6.84 Hz), 1.48 (3H, t, J=6.96 Hz), 2.35 (3H, sep, J=6.80 Hz), 3.81 (3H, s), 4.33 (2H, q, J=6.96 Hz), 5.72 (3H, s), 7.16 (1H, dt, J=3.20 Hz), 7.59 (1H, dd, J=9.25, 10.61 Hz), 7.71 (1H, d, J=2.76 Hz), 7.77 (1H, q, J=3.20 Hz), 7.81 (1H,

dd, J=2.82, 9.51 Hz), 8.32 (1H, d, J=8.36 Hz), 8.74 (2H, t, J=9.15 Hz), 9.13 (1H, d, J=2.40 Hz) ppm. $^{13}C\{^{1}H\}$-NMR (125 MHz, DMSO-d_6): δ 14.43, 21.40, 22.26, 30.87, 55.94, 64.55, 77.64, 107.03, 111.08, 113.16, 116.33, 116.37, 117.55, 117.72, 125.37, 125.93, 131.03, 131.52, 138.61, 144.58, 152.80, 155.64, 159.58, 173.41 ppm. ^{19}F NMR (376 MHz, DMSO-d_6): δ 136.34 ppm. HRMS (+ mode): m/z calculated for $C_{34}H_{41}ClFN_2O_2Ru$, $[M]^+$ = 665.1885, found: 665.1890. Purity (HPLC) = 95.6%.

RA4-nPA3-nAD2. Yield: 26.2 mg (71.7 %). ^{1}H NMR (400 MHz, DMSO-d_6): δ 0.87 (9H, d, J=6.84 Hz), 1.18 (9H, d, J=6.80 Hz), 1.48 (3H, t, *J*=6.94 Hz), 2.42 (3H, sep, *J*=6.82 Hz), 3.97 (3H, s), 4.31 (2H, q, J=6.94 Hz), 5.57 (3H, s), 7.50 (1H, dd, J=10.95, 8.75 Hz), 7.68 (2H, q, J=3.34 Hz), 7.73 (1H, dd, J=2.82, 9.51 Hz), 7.98 (1H, dd, J=2.42, 7.74 Hz), 8.20 (1H, d, J=8.40 Hz), 8.68 (1H, d, J=8.40 Hz), 8.85 (1H, d, J=9.53 Hz), 9.06 (1H, s) ppm. $^{13}C\{^{1}H\}$-NMR (125 MHz, DMSO-d_6): δ 14.43, 21.31, 22.25, 30.84, 56.30, 64.46, 73.79, 107.03, 109.21, 116.13, 117.40, 124.81, 125.66, 131.08, 131.24, 138.44, 144.28, 147.10, 147.13, 147.19, 150.71, 153.46, 159.25, 168.93 ppm. ^{19}F NMR (376 MHz, DMSO-d_6): δ 132.75 ppm. HRMS (+ mode): m/z calculated for $C_{34}H_{41}ClFN_2O_2Ru$, $[M]^+$ = 665.1885, found: 665.1894. Purity (HPLC) = 95.0%.

RA4-nPA3-nAD7. Yield: 32.7 mg (93.7 %). ^{1}H NMR (400 MHz, DMSO-d_6): δ 1H-NMR (DMSO) 0.92 (9H, d, J=6.76 Hz), 1.11 (9H, d, J=6.88 Hz), 1.48 (3H, t, J=6.96 Hz), 2.28 (3H, sep, J=6.81 Hz), 2.57 (3H, s), 3.85 (3H, s), 4.33 (2H, q, J=6.88 Hz), 5.80 (3H, s), 6.97 (1H, dd, J=2.78, 8.87 Hz), 7.10 (1H, d, J=2.76 Hz), 7.70 (1H, d, J=2.80 Hz), 7.82 (1H, dd, J=2.80, 9.53 Hz), 7.91 (1H, d, J=8.85 Hz), 8.26 (1H, d, J=8.36 Hz), 8.68 (2H, d, J=9.73 Hz), 8.71 (2H, d, J=8.44 Hz), 8.91 (1H, s) ppm. $^{13}C\{^{1}H\}$-NMR (125 MHz, DMSO-d_6): δ 14.45, 18.53, 20.89, 22.96, 30.67, 55.64, 64.49, 80.20, 107.12, 110.47, 112.45, 115.95, 124.75, 125.45, 125.53, 130.51, 130.61, 131.21, 138.52, 143.80, 144.45, 153.00, 158.93, 159.29, 171.75 ppm. HRMS (+ mode): m/z calculated for $C_{35}H_{44}ClN_2O_2Ru$, $[M]^+$ = 661.2136, found: 661.2146. Purity (HPLC) = 92.9%.

RA4-nPA6-nAD6. Yield: 39.5 mg (85.4 %). ^{1}H NMR (400 MHz, DMSO-d_6): δ 0.89 (9H, d, J=6.84 Hz), 1.18 (9H, d, J=6.80 Hz), 2.43 (3H, sep, J=6.81 Hz), 3.92 (3H, s), 5.65 (3H, s), 7.31 (1H, t, J=1.96 Hz), 7.69 (1H, t, J=1.94 Hz), 7.76 (1H, t, J=1.64 Hz), 8.15 (1H, dd, J=2.40, 9.33 Hz), 8.32 (2H, t, J=4.20 Hz), 8.50 (1H, d, J=2.36 Hz), 8.84 (1H, d, J=8.36 Hz), 8.91 (1H, d, J=9.37 Hz), 9.17 (1H, s) ppm. $^{13}C\{^{1}H\}$-NMR (125 MHz, DMSO-d_6): δ 21.35, 22.17, 30.85, 56.25, 75.27, 79.19, 108.16, 115.46, 116.35, 125.66, 127.92, 129.90, 131.24, 133.26, 134.26, 134.93, 139.93, 146.64, 152.22, 156.21, 160.35, 170.52 ppm. HRMS (+ mode): m/z calculated for $C_{32}H_{36}Cl_3N_2ORu$, $[M]^+$ = 673.0928, found: 673.0921. Purity (HPLC) = 95.3%.

RA4-nPA6-nAD4. Yield: 33.8 mg (97.8 %). ^{1}H NMR (400 MHz, DMSO-d_6): δ 0.89 (9H, d, J=6.84 Hz), 1.18 (9H, d, J=6.84 Hz), 2.44 (3H, sep, J=6.84 Hz), 3.91 (3H, s), 5.65 (3H, s), 7.14 (1H, td, J=2.18, 10.69 Hz), 7.55 (1H, td, J=1.93, 9.55 Hz), 7.58 (1H, t, J=1.84 Hz), 8.15 (1H, dd, J=2.42, 9.35 Hz), 8.32 (1H, d, J=8.36 Hz), 8.50 (1H, d, J=2.40 Hz), 8.84 (1H, d, J=8.40 Hz), 8.92 (1H, d, J=9.37 Hz), 9.16 (1H, s) ppm. $^{13}C\{^{1}H\}$-NMR (125 MHz, DMSO-d_6): δ 21.27, 22.25, 30.85, 56.28, 74.98, 103.02, 103.21, 105.68, 116.69, 125.66, 127.92, 129.90, 131.23, 133.26, 134.92, 139.90,

146.62, 152.14, 152.24, 156.20, 160.89, 170.36 ppm. ^{19}F NMR (376 MHz, DMSO-d_6): δ 109.44 ppm. HRMS (+ mode): m/z calculated for $C_{32}H_{36}Cl_2FN_2ORu$, $[M]^+$ = 655.1230, found: 655.1238. Purity (HPLC) = 95.8%.

RA4-nPA6-nAD3. Yield: 28.7 mg (70.9 %). ^{1}H NMR (400 MHz, DMSO-d_6): δ 0.84 (9H, d, J=6.88 Hz), 1.18 (9H, d, J=6.76 Hz), 2.41 (3H, sep, J=6.79 Hz), 3.96 (3H, s), 5.61 (3H, s), 7.36 (1H, dd, J=2.50, 8.95 Hz), 7.52 (1H, d, J=2.36 Hz), 7.97 (1H, d, J=9.09 Hz), 8.07 (1H, d, J=8.97 Hz), 8.12 (1H, dd, J=2.40, 9.33 Hz), 8.20 (1H, dd, J=2.16, 8.89 Hz), 8.34 (1H, d, J=10.13 Hz), 8.48 (1H, d, J=2.40 Hz), 8.60 (1H, d, J=1.80 Hz), 8.82 (1H, d, J=8.40 Hz), 8.98 (1H, d, J=9.29 Hz), 9.26 (1H, s) ppm. ^{13}C{^{1}H}-NMR (125 MHz, DMSO-d_6): δ 21.38, 22.08, 30.87, 55.49, 74.31, 106.42, 117.51, 120.43, 121.59, 121.97, 125.29, 127.36, 127.90, 129.76, 130.36, 131.30, 133.07, 134.61, 134.91, 139.75, 145.99, 146.56, 156.76, 158.82, 167.88 ppm. HRMS (+ mode): m/z calculated for $C_{35}H_{39}Cl_2N_2ORu$, $[M]^+$ = 687.1481, found: 687.1486. Purity (HPLC) = 97.3%.

References

[1] C. J. Murray, K. S. Ikuta, F. Sharara, L. Swetschinski, G. R. Aguilar, A. Gray, C. Han, C. Bisignano, P. Rao, E. Wool, S. C. Johnson, A. J. Browne, M. G. Chipeta, F. Fell, S. Hackett, G. Haines-Woodhouse, B. H. K. Hamadani, E. A. P. Kumaran, B. McManigal, R. Agarwal, S. Akech, S. Albertson, J. Amuasi, J. Andrews, A. Aravkin, E. Ashley, F. Bailey, S. Baker, B. Basnyat, A. Bekker, R. Bender, A. Bethou, J. Bielicki, S. Boonkasidecha, J. Bukosia, C. Carvalheiro, C. Castañeda-Orjuela, V. Chansamouth, S. Chaurasia, S. Chiurchiù, F. Chowdhury, A. J. Cook, B. Cooper, T. R. Cressey, E. Criollo-Mora, M. Cunningham, S. Darboe, N. P. J. Day, M. D. Luca, K. Dokova, A. Dramowski, S. J. Dunachie, T. Eckmanns, D. Eibach, A. Emami, N. Feasey, N. Fisher-Pearson, K. Forrest, D. Garrett, P. Gastmeier, A. Z. Giref, R. C. Greer, V. Gupta, S. Haller, A. Haselbeck, S. I. Hay, M. Holm, S. Hopkins, K. C. Iregbu, J. Jacobs, D. Jarovsky, F. Javanmardi, M. Khorana, N. Kissoon, E. Kobeissi, T. Kostyanev, F. Krapp, R. Krumkamp, A. Kumar, H. H. Kyu, C. Lim, D. Limmathurotsakul, M. J. Loftus, M. Lunn, J. Ma, N. Mturi, T. Munera-Huertas, P. Musicha, M. M. Mussi-Pinhata, T. Nakamura, R. Nanavati, S. Nangia, P. Newton, C. Ngoun, A. Novotney, D. Nwakanma, C. W. Obiero, A. Olivas-Martinez, P. Olliaro, E. Ooko, E. Ortiz-Brizuela, A. Y. Peleg, C. Perrone, N. Plakkal, A. Ponce-de-Leon, M. Raad, T. Ramdin, A. Riddell, T. Roberts, J. V. Robotham, A. Roca, K. E. Rudd, N. Russell, J. Schnall, J. A. G. Scott, M. Shivamallappa, J. Sifuentes-Osornio, N. Steenkeste, A. J. Stewardson, T. Stoeva, N. Tasak, A. Thaiprakong, G. Thwaites, C. Turner, P. Turner, H. R. van Doorn, S. Velaphi, A. Vongpradith, H. Vu, T. Walsh, S. Waner, T. Wangrangsimakul, T. Wozniak, P. Zheng, B. Sartorius, A. D. Lopez, A. Stergachis, C. Moore, C. Dolecek, M. Naghavi, *Lancet* **2022**, *399*, 629–655.

[2] W. Cong, A. N. Poudel, N. Alhusein, H. Wang, G. Yao, H. Lambert, *Antibiotics* **2021**, *10*, 745.

[3] CDC, “COVID-19 & Antibiotic Resistance,” can be found under https://www.cdc.gov/drugresistance/covid19.html, **2022**.

[4] “Global trends in antimicrobial use in food-producing animals: 2020 to 2030 | PLOS Global Public Health,”

[5] M. S. Butler, I. R. Henderson, R. J. Capon, M. A. T. Blaskovich, *J. Antibiot.* **2023**, *76*, 431–473.

[6] T. C. Johnstone, K. Suntharalingam, S. J. Lippard, *Chem. Rev.* **2016**, *116*, 3436–3486.

[7] C. Zhang, C. Xu, X. Gao, Q. Yao, *Theranostics* **2022**, *12*, 2115–2132.

[8] F. Li, J. G. Collins, F. R. Keene, *Chem. Soc. Rev.* **2015**, *44*, 2529–2542.

[9] K. L. Smitten, S. D. Fairbanks, C. C. Robertson, J. B. de la Serna, S. J. Foster, J. A. Thomas, *Chem. Sci.* **2019**, *11*, 70–79.

[10] H. Sun, Q. Zhang, R. Wang, H. Wang, Y.-T. Wong, M. Wang, Q. Hao, A. Yan, R. Y.-T. Kao, P.-L. Ho, H. Li, *Nat. Commun.* **2020**, *11*, 5263.

[11] C. Wang, Y. Xia, R. Wang, J. Li, C.-L. Chan, R. Y.-T. Kao, P. H. Toy, P.-L. Ho, H. Li, H. Sun, *Nat. Commun.* **2023**, *14*, 5311.

[12] A. Frei, A. D. Verderosa, A. G. Elliott, J. Zuegg, M. A. T. Blaskovich, *Nat. Rev. Chem.* **2023**, *7*, 202–224.

[13] C. Weng, Y. L. K. Tan, W. G. Koh, W. H. Ang, *Angew. Chem. Int. Ed.* **n.d.**, *n/a*, e202310040.

[14] A. Frei, J. Zuegg, A. G. Elliott, M. Baker, S. Braese, C. Brown, F. Chen, C. G. Dowson, G. Dujardin, N. Jung, A. P. King, A. M. Mansour, M. Massi, J. Moat, H. A. Mohamed, A. K. Renfrew, P. J. Rutledge, P. J. Sadler, M. H. Todd, C. E. Willans, J. J. Wilson, M. A. Cooper, M. A. T. Blaskovich, *Chem. Sci.* **2020**, *11*, 2627–2639.

[15] A. Frei, A. G. Elliott, A. Kan, H. Dinh, S. Bräse, A. E. Bruce, M. R. Bruce, F. Chen, D. Humaidy, N. Jung, A. P. King, P. G. Lye, H. K. Maliszewska, A. M. Mansour, D. Matiadis, M. P. Muñoz, T.-Y. Pai, S. Pokhrel, P.

J. Sadler, M. Sagnou, M. Taylor, J. J. Wilson, D. Woods, J. Zuegg, W. Meyer, A. K. Cain, M. A. Cooper, M. A. T. Blaskovich, *JACS Au* **2022**, *2*, 2277–2294.
[16] K. L. Smitten, H. M. Southam, J. B. de la Serna, M. R. Gill, P. J. Jarman, C. G. W. Smythe, R. K. Poole, J. A. Thomas, *ACS Nano* **2019**, *13*, 5133–5146.
[17] A. M. Varney, K. L. Smitten, J. A. Thomas, S. McLean, *ACS Pharmacol. Transl. Sci.* **2021**, *4*, 168–178.
[18] J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackermann, V. M. Tran, A. Chiappino-Pepe, A. H. Badran, I. W. Andrews, E. J. Chory, G. M. Church, E. D. Brown, T. S. Jaakkola, R. Barzilay, J. J. Collins, *Cell* **2020**, *180*, 688-702.e13.
[19] G. Liu, D. B. Catacutan, K. Rathod, K. Swanson, W. Jin, J. C. Mohammed, A. Chiappino-Pepe, S. A. Syed, M. Fragis, K. Rachwalski, J. Magolan, M. G. Surette, B. K. Coombes, T. Jaakkola, R. Barzilay, J. J. Collins, J. M. Stokes, *Nat. Chem. Biol.* **2023**, 1–9.
[20] A. Capecchi, X. Cai, H. Personne, T. Köhler, C. van Delden, J.-L. Reymond, *Chem. Sci.* **2021**, *12*, 9221–9232.
[21] J. P. Janet, H. J. Kulik, *Chem. Sci.* **2017**, *8*, 5137–5152.
[22] J. P. Janet, F. Liu, A. Nandy, C. Duan, T. Yang, S. Lin, H. J. Kulik, *Inorg. Chem.* **2019**, *58*, 10592–10606.
[23] J. P. Janet, S. Ramesh, C. Duan, H. J. Kulik, *ACS Cent. Sci.* **2020**, *6*, 513–524.
[24] J. P. Janet, C. Duan, A. Nandy, F. Liu, H. J. Kulik, *Acc. Chem. Res.* **2021**, *54*, 532–545.
[25] H. Kneiding, R. Lukin, L. Lang, S. Reine, T. B. Pedersen, R. D. Bin, D. Balcells, *Digital Discovery* **2023**, *2*, 618–633.
[27] R. Laplaza, S. Gallarati, C. Corminboeuf, *Chem. Methods* **2022**, *2*, e202100107.
[28] M. J. Chow, C. Licona, D. Yuan Qiang Wong, G. Pastorin, C. Gaiddon, W. H. Ang, *J. Med. Chem.* **2014**, *57*, 6043–6059.
[30] C. Weng, L. Shen, W. H. Ang, *Angew. Chem. Int. Ed.* **2020**, *59*, 9314–9318.
[30] C. Weng, L. Shen, J. W. Teo, Z. C. Lim, B. S. Loh, W. H. Ang, *JACS Au* **2021**, *1*, 1348–1354.
[31] C. Weng, H. Yang, B. S. Loh, M. W. Wong, W. H. Ang, *J. Am. Chem. Soc.* **2023**, *145*, 6453–6461.
[32] M. A. T. Blaskovich, J. Zuegg, A. G. Elliott, M. A. Cooper, *ACS Infect. Dis.* **2015**, *1*, 285–287.
[34] J. C. Brammer, G. Blanke, C. Kellner, A. Hoffmann, S. Herres-Pawlis, U. Schatzschneider, *J. Cheminf.* **2022**, *14*, 66.
[34] M. Krenn, Q. Ai, S. Barthel, N. Carson, A. Frei, N. C. Frey, P. Friederich, T. Gaudin, A. A. Gayle, K. M. Jablonka, R. F. Lameiro, D. Lemm, A. Lo, S. M. Moosavi, J. M. Nápoles-Duarte, A. Nigam, R. Pollice, K. Rajan, U. Schatzschneider, P. Schwaller, M. Skreta, B. Smit, F. Strieth-Kalthoff, C. Sun, G. Tom, G. Falk von Rudorff, A. Wang, A. D. White, A. Young, R. Yu, A. Aspuru-Guzik, *Patterns* **2022**, *3*, 100588.
[29] RDKit: Open-source cheminformatics. https://www.rdkit.org
[36] D. Rogers, M. Hahn, *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
[31] https://github.com/TheFreiLab/RutheniumML
[38] *Methods for Dilution of Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically; Approved Standard—10th Edition*, Clinical And Laboratory Standards Institute, Wayne, PA, **2015**.