



OPEN

Property space mapping of *Pseudomonas aeruginosa* permeability to small molecules

Inga V. Leus¹, Jon W. Weeks¹, Vincent Bonifay¹, Yue Shen², Liang Yang¹, Connor J. Cooper³, Dinesh Nath¹, Adam S. Duerfeldt⁴, Jeremy C. Smith^{2,5}, Jerry M. Parks³, Valentin V. Rybenkov¹✉ & Helen I. Zgurskaya¹✉

Two membrane cell envelopes act as selective permeability barriers in Gram-negative bacteria, protecting cells against antibiotics and other small molecules. Significant efforts are being directed toward understanding how small molecules permeate these barriers. In this study, we developed an approach to analyze the permeation of compounds into Gram-negative bacteria and applied it to *Pseudomonas aeruginosa*, an important human pathogen notorious for resistance to multiple antibiotics. The approach uses mass spectrometric measurements of accumulation of a library of structurally diverse compounds in four isogenic strains of *P. aeruginosa* with varied permeability barriers. We further developed a machine learning algorithm that generates a deterministic classification model with minimal synonymy between the descriptors. This model predicted good permeators into *P. aeruginosa* with an accuracy of 89% and precision above 58%. The good permeators are broadly distributed in the property space and can be mapped to six distinct regions representing diverse chemical scaffolds. We posit that this approach can be used for more detailed mapping of the property space and for rational design of compounds with high Gram-negative permeability.

Pseudomonas aeruginosa is a challenging pathogen that causes a variety of infections in humans and animals. Only a few therapeutic options are available for *P. aeruginosa* infections and those fail against multidrug-resistant isolates^{1,2}. The challenge posed by this bacterium is exacerbated by a lack of success in discovering new small molecules with antibacterial activities against this pathogen.

The major reason why *P. aeruginosa* presents such a formidable therapeutic challenge is that the cells are protected by an effective drug permeability barrier composed of two lipid membranes and reinforced by active multidrug efflux pumps³. Currently, major efforts are being directed toward understanding and overcoming this drug permeation barrier and optimizing drug potencies against *P. aeruginosa* and other Gram-negative pathogens^{4–6}.

Drug permeation studies using artificial membranes such as parallel artificial membrane permeability assays (PAMPA) or cell cultures (e.g., Caco-2 assays) are highly informative about drug oral bioavailability in mammalian hosts and are used broadly during compound optimization studies (see for example^{7,8}). Recently, these approaches have been combined with traditional statistical and evolving machine learning (ML) approaches to develop predictive computational models and to identify physicochemical properties that could help with optimization of compound permeation and intracellular accumulation^{9–11}. Among notable successes are Lipinski's "rule of five" of drug solubility and permeation that is broadly used in medicinal chemistry optimization of oral bioavailability of human therapeutics¹² and a predictive model for drug accumulation in *Caenorhabditis elegans*¹³.

In bacteria, analyses of intracellular accumulation of a broad range of chemical structures have been enabled by liquid chromatography–tandem mass spectrometry (LC–MS/MS), an approach first applied to bacteria by Cai and coworkers¹⁴. Later, Tan and coworkers used LC–MS/MS to analyze concentration-dependent accumulation of salicyl-AMS (adenosine monosulfamate; 5'-O-sulfamoyladenine), an inhibitor of siderophore biosynthesis in *Escherichia coli*¹⁵. Accumulation at a 30-min timepoint was determined for ten acyl-AMS analogues in *E. coli*, *Bacillus subtilis* (Gram-positive), and *Mycobacterium smegmatis*, in the presence or absence of efflux pump inhibitors¹⁵. Pearson pairwise correlation coefficients were determined for accumulation under each condition

¹Department of Chemistry and Biochemistry, University of Oklahoma, 101 Stephenson Parkway, Norman, OK 73019, USA. ²Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN 37996, USA. ³Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. ⁴Department of Medicinal Chemistry, University of Minnesota, 717 Delaware St. SE, Minneapolis, MN 55414, USA. ⁵Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996, USA. ✉email: valya@ou.edu; elenaz@ou.edu

versus a set of 20 calculated physicochemical parameters. This approach identified significant positive correlations between accumulation in *E. coli* and hydrophobicity.

More recently, Hergenrother and coworkers conducted a larger analysis of a diversity library of 100 compounds built around five major natural product-derived scaffolds^{16,17}. Notably, they found only 12 positively charged compounds that accumulated significantly in *E. coli*, and, within this subset, no correlation was observed with hydrophobicity or molecular weight, in contrast to some early studies of radioactively labeled fluoroquinolones^{18,19}. However, all 12 compounds contained amines, 8 of which were primary amines. Calculations of physicochemical descriptors, synthesis of additional analogues, and application of a random forest classification model to 68 compounds revealed that rotatable bonds and globularity were negative predictors of accumulation while the amphiphilic moment, the distance between hydrophobic and hydrophilic portions of a molecule, was a positive predictor.

Along the same lines, Erwin and coworkers evaluated the accumulation of > 100 inhibitors of NAD-dependent DNA ligase (LigA) in *E. coli* using LC–MS/MS detection²⁰. They found no correlations between the size and hydrophobicity and the intracellular accumulation of these inhibitors. In agreement with early studies of small sets of compounds^{21,22}, they also found poor correlation between the overall level of bacteria-associated compound and antibacterial activity in compounds with matched biochemical activities.

The varying significance of physicochemical properties such as hydrophobicity and molecular weight in the intracellular accumulation of compounds is a consequence of the complex nature of the Gram-negative permeability barrier. Active efflux pumps, which transport compounds across both the inner and the outer membranes (OMs), act synergistically with the membranes, together forming a permeability barrier for any given compound^{23,24}. This synergistic relationship manifests as highly non-linear changes in intracellular accumulation levels depending on the extracellular concentration of compounds and the species-specific compositions of efflux pumps and the OM. An additional difficulty is that the above studies analyze small sample sizes with limited diversity compared to the entirety of chemical property space.

For historic reasons, most of the previous drug permeation studies have been limited to *E. coli*. However, these data are not fully applicable to all pathogens, because the cell envelope of this species is relatively permeable to small molecules. In contrast, the cell envelope of *P. aeruginosa* is notably less permeable to certain compounds than that of *E. coli*²⁵ due to the presence of general porins located in the *E. coli* OM that enable diffusion of amphiphilic molecules up to 600 Da in size²⁶. The OM of *P. aeruginosa* contains only substrate-specific porins that limit natural uptake to amino acids, simple sugars, and short peptides²⁷. Interestingly, inactivation of all characterized and predicted porin-like proteins in *P. aeruginosa* led to only minor changes in antibiotic susceptibility, suggesting that almost all antibiotics and diverse hydrophilic nutrients can bypass porins and instead permeate directly through the OM lipid bilayer²⁷. In addition, *P. aeruginosa* cells express several multidrug efflux pumps with partially overlapping substrate specificities that pump out compounds that do manage to permeate slowly across the OM^{28,29}.

The contributions of the OM barrier and active efflux in antibacterial activities of compounds can be separated by comparing these activities in wild-type, efflux-deficient, hyperporinated, and ‘barrierless’ (i.e., both hyperporinated and efflux-deficient) strains^{30,31}. In combination with machine learning techniques, such analyses identified physicochemical properties of compounds predictive of their interactions with proteins and lipids of the OM and active efflux pumps and showed that each pathogen presents its own unique challenges. Both barriers (i.e., OM permeability and efflux avoidance) must be considered during antibiotic development efforts, because optimizing the properties of compounds to overcome one barrier tends to result in detrimental properties against the other barrier^{30,32}.

In this study, we developed an LC–MS approach to analyze the accumulation of compounds in *P. aeruginosa*. The approach avoids a washing step, the modification that enables analyses of compounds with low affinity to the cell, but also limits the approach to compounds with low non-specific binding to the filters. The insight into compound permeation is derived from measurements in four strains of *P. aeruginosa* with differing permeability barriers. We assembled a library of compounds with properties representative of drug-like libraries and analyzed the accumulation of 83 compounds in four *P. aeruginosa* strains at several extracellular concentrations and two time points (Table S1). We defined the “good” permeators as compounds that accumulate in wild type cells linearly with increasing concentration but that are not affected by compromising permeation barriers and have low non-specific retention on control filters. We further developed a machine learning algorithm that selects a descriptor set with the minimal synonymity between its members and uses this set to generate a unique classification model. The high accuracy of the model (ranging between 85 and 97%) indicates that it recapitulates the permeation properties of the compounds. Good permeators into *P. aeruginosa* were broadly variable and could be mapped to six distinct regions in chemical property space representing diverse chemical scaffolds. We posit that our strategy can be used for further, more detailed mapping of property space and for the rational design of compounds with optimized Gram-negative permeability.

Results

A 96-well filter assay for quantifying the contributions of efflux and the OM barrier to compound penetration into *P. aeruginosa*. The kinetics of intracellular accumulation of compounds is expected to be strongly affected by compound affinities to intracellular binding sites²³. To test this notion, we first analyzed the accumulation of the radioactively labeled antibiotics ciprofloxacin and tetracycline, which are known to have strong affinities to intracellular targets in bacterial cells, as well as the human therapeutics oxymetazoline, nepafenac, efavirenz, delavirdine and metoprolol, which have unknown affinities to bacterial cells. As expected, measurements of minimal inhibitory concentrations (MICs) showed that ciprofloxacin and tetracycline inhibit the growth of the wild type *P. aeruginosa* PAO1 at low micromolar concentrations (Table 1). The

Compound	MIC, μM (mg/L)			
	PAO1	PAO1-Pore	P Δ 3	P Δ 3-Pore
Tetracycline	12.5 (5.56)	0.39 (0.17)	0.78 (0.35)	0.78 (0.35)
Ciprofloxacin	0.19 (0.06)	0.09 (0.03)	0.05 (0.02)	0.02 (0.01)
Metoprolol	> 1000	> 1000	> 1000	> 1000
Oxymetazoline	> 1000	> 1000	1000	1000
Nepafenac	> 1000	> 1000	> 1000	> 1000
Efavirenz	> 1000	> 1000	> 1000	25 (7.89)
Delavirdine	> 1000	> 1000	> 1000	> 1000

Table 1. Minimum inhibitory concentration of antibiotics and human therapeutics in four *P. aeruginosa* strains with varying permeability barriers.

MICs of these antibiotics improved in both the hyperporinated PAO1-Pore and the efflux-deficient P Δ 3 cells, which lack the three major efflux pumps $\Delta mexAB$, $\Delta mexCD$, and $\Delta mexXY$. However, the antibacterial activity of ciprofloxacin was mostly affected by efflux, as seen from a modest, two-fold effect of hyperporination, whereas the activity of tetracycline was affected by both the efflux and OM barriers. The “barrierless” P Δ 3-Pore strain was at least 8- to 16-fold more susceptible to both antibiotics than PAO1. In contrast, no measurable antibacterial activities were observed for the other compounds for concentrations below 1 mM, except that efavirenz had an MIC of 25 mM in P Δ 3-Pore cells.

We next optimized conditions and analyzed time- and concentration-dependent accumulation of [^{14}C]-ciprofloxacin and [^3H]-oxymetazoline in four strains of *P. aeruginosa* with variable efflux and OM permeability barriers using a 96-well filter plate assay. We found that washing of filtered cells with the incubation medium differentially affected the retention of [^{14}C]-ciprofloxacin and [^3H]-oxymetazoline (Fig. S1), suggesting that compound variability in affinities to intracellular binding sites could strongly affect the outcomes of the assay. Therefore, washing of filtered cells was omitted and filters were removed from plates to reduce non-specific background (see Methods). As described in Methods, under the conditions of the assay, each filter retained 4.3 ± 0.7 μL of solution, and cells displaced 2.8 ± 0.2 μL of this volume (67%) (Fig. S1). Thus, the assay is expected to report accurate values for compounds that accumulate in cells but will not be informative for compounds unable to accumulate in at least one of the four strains.

To test the validity of the assay, we next measured the accumulation of [^{14}C]-ciprofloxacin and [^3H]-oxymetazoline into otherwise isogenic strains of *P. aeruginosa* that differ in the permeability of their cell envelope. Since the composition of cell membranes does not change significantly with hyperporination and efflux inactivation²⁹, we expect that non-specific binding remains the same for different strains and the differences in the accumulation levels between the strains reflect the differences in the intracellular accumulation. Kinetic experiments showed that both [^{14}C]-ciprofloxacin and [^3H]-oxymetazoline reached steady-state levels within the first minutes of incubation (Figs. S2, S3). The steady-state levels increased linearly with increasing concentrations of the compounds (Fig. 1A,B). Furthermore, these accumulation levels varied between the strains and were higher in the efflux-deficient and hyperporinated cells than in the wild type PAO1. Efflux inactivation or hyperporination resulted in a similar increase in the accumulation of [^3H]-oxymetazoline, and the levels were highest in P Δ 3-Pore cells. In contrast, inactivation of efflux led to significantly higher levels of accumulation of [^{14}C]-ciprofloxacin than did hyperporination but again the levels were highest in P Δ 3-Pore cells. This major contribution of efflux in the intracellular accumulation of ciprofloxacin was also observed in its MICs in different *P. aeruginosa* strains (Table 1).

The above findings show that the developed assay allows analyses of the kinetics of drug accumulation in cells and can distinguish between the contributions of active efflux and the OM barrier. Importantly, in the time-course of the experiment, changes in the intracellular levels in wild type PAO1 were very small for both radioactive compounds, suggesting that the compounds permeate these cells very slowly, even at the highest concentrations considered. In addition, at the highest concentration of [^3H]-oxymetazoline (20 μM), the signal declined during the first 5 min of incubation in PAO1 cells. This result suggests that when compounds do not accumulate in cells, they can be lost due to non-specific binding to plates or aggregation.

Accumulation of other radioactively labeled compounds was also concentration-dependent and varied between the four *P. aeruginosa* strains (Fig. 1). Hyperporination had the strongest effect on the accumulation of [^{14}C]-efavirenz, [^3H]-nepafenac and [^3H]-delavirdine, whereas the “barrierless” P Δ 3-Pore cells accumulated the highest levels of compounds (Fig. 1C,F). The accumulation of delavirdine and tetracycline was equally affected by inactivation of efflux and hyperporination (Fig. 1D,G), whereas metoprolol accumulation was not affected by either of the two barriers (Fig. 1E). When intracellular accumulation of compounds was normalized to that of the PAO1 cells, the ratio of concentration slopes even for P Δ 3-Pore varied between one- and eightfold (Fig. 1H). Except for ciprofloxacin, inactivation of efflux alone and hyperporination of cells alone generated modest slope ratios of 1.5–3. Thus, the assay is sensitive to differences in permeation and can be used with structurally diverse compounds.

Properties of the focused library of compounds for analyses. Previous studies have shown that the physicochemical space of antibiotics is broader than other drugs and includes more hydrophilic compounds,

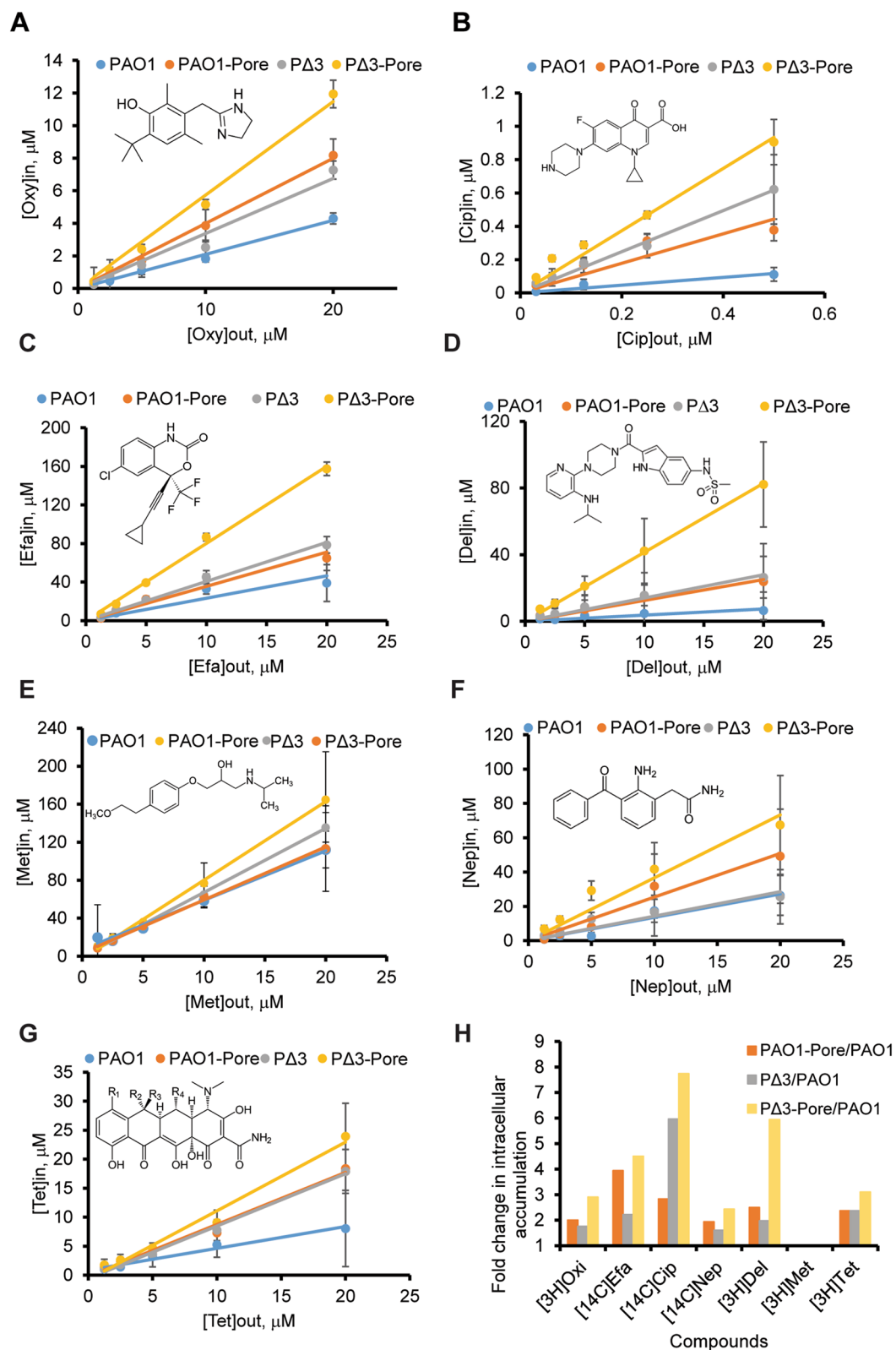


Figure 1. Intracellular accumulation profiles of radioactively labeled compounds in four strains of *P. aeruginosa* after 20 min of incubation: (A) [³H]-oxymetazoline (Oxy); (B) [¹⁴C]-ciprofloxacin (Cip); (C) [¹⁴C]-efavirenz (Efa); (D) [³H]-delavirdine (Del); (E) [3H]-metoprolol (Met); (F) [³H]-nepafenac (Nep); (G) [³H]-tetracycline (Tet). Extracellular (out) and intracellular (in) concentrations were calculated based on specific radioactivities of compounds. (H) Fold change (represented as slope ratios) in the intracellular accumulation of compounds in hyperporinated, efflux-deficient and “barrierless” cells in comparison to the wild type PAO1.

such as beta-lactams, which penetrate the cells through water-filled OM porins, as well as some large compounds, such as vancomycin, which act on cell surfaces and do not need to penetrate the cell envelope³³. In addition, the presence of primary amines and positive charges are in general associated with increased penetration into some bacterial cells³⁴. We next analyzed a focused library of 12,000 commercially available compounds and selected 220 compounds with PSA 50+, MW < 2000 Da, and cLogD_{7.4} < 5 for further analyses.

Approximately 10% of the 220 purchased compounds possessed antibacterial activities in at least one of the four *P. aeruginosa* strains grown in MOPS-M9 medium and Luria–Bertani (LB) broth and these included representatives of fluoroquinolones, sulfonamides, cyclines and linezolid (Tables 2 and S2). After elimination of insoluble compounds and optimization of the LC–MS method, 98 compounds were selected for further analyses, including six synthesized trisubstituted piperazin-2-one derivatives (Fig. 7), out of which the intracellular accumulation of 83 compounds was measured and for 66 compounds it was quantified (see below, Table S1).

The analyzed 66 compounds vary in molecular weight (MW) between 200 and 750 Da and have cLogD_{7.4} values between 5.0 and -3.5 (Fig. 2A). We next calculated nine physico-chemical properties including the molecular weight, logP, the number of hydrogen bond donors and acceptors, clogD_{7.4}, the topological polar surface area, the fraction of sp³ hybridized carbon atoms (Fsp3), the heavy atom count, and the number of rotatable bonds for the analyzed compounds. These properties were also calculated for the SPARK compound library³⁵, FDA-approved drugs³⁶, and compounds whose accumulation in *E. coli* was analyzed in previous studies^{15,20,34}. Principal component analysis showed that the chemical space covered in this work is broader than the space covered by previous studies^{15,20,34} (Fig. 2B–C). The distribution of properties of this library surrounds that of Iyer et al.²⁰ in every feature (Fig. S4) and that of Richter et al.³⁴ in principal component 1 (PC1), which is represented by molecular weight, number of hydrogen bond donors and acceptors, topological polar surface area, rotatable bounds, and heavy atom count (Fig. 2C and S4). The distribution of the compounds analyzed by Richter et al.³⁴ is wider in the Fsp3 property space but narrower in logP and logD, resulting in a slightly narrower distribution in principal component 2 (PC2). The partial overlap in PC1 between our compounds and the Richter et al. set is due to an offset in the range of values for the topological polar surface area and the number of hydrogen bond acceptors. For example, the number of hydrogen bond acceptors ranges from 0–8 in the Richter et al. set and from 2–12 in the present study.

Quantification of the effect of efflux and the OM barrier in compound penetration into *P. aeruginosa*. The 96-well filter assay described above was next used for LC–MS-based quantification of compound penetration into *P. aeruginosa* cells. To reduce variability, non-physiological responses, and matrix effects, cells were grown in the optimized M9 minimal medium (see Methods) and all subsequent steps were carried out at room temperature.

Compound accumulation was analyzed at four concentrations: 10, 20, 40 and 80 μM. Kinetic analyses of the radioactively labeled compounds described above showed that compounds vary in the time needed to reach

Antibiotic	MICs, μM (mg/L)			
	PAO1	PAO1-Pore	PA3	PA3-Pore
Doxorubicin	200 (108.7)	12.5 (6.8)	25 (13.6)	3.2 (1.7)
Levofloxacin	0.5 (0.18)	0.031 (0.011)	0.25 (0.2)	0.008 (0.003)
Norfloxacin	1.25 (0.4)	0.2 (0.06)	0.625 (0.2)	0.1 (0.03)
Sulfaphenazole	>200 (>62.9)	12.5 (3.9)	200 (62.9)	3.2 (1.0)
Sulfameter	>200 (>56.1)	12.5 (3.5)	25 (7.0)	6.25 (1.75)
Sulfamethazine	>200 (>60.1)	50 (15.0)	100 (30.0)	12.5 (3.75)
Sulfathiazole	50 (12.8)	3.2 (0.8)	6.25 (1.6)	1.6 (0.4)
Sulfisoxazole	>200 (>53.7)	>200	>200	>200
Sulfadimethoxine	>200 (62.1)	25 (7.8)	200 (62.1)	1.6 (0.5)
Prulifloxacin	0.5 (0.23)	0.016 (0.01)	0.125 (0.06)	0.004 (0.002)
Nadifloxacin	5 (1.8)	0.0625 (0.02)	0.625 (0.23)	0.0156 (0.01)
Pazufloxacin mesylate	0.5 (0.21)	0.0625 (0.03)	0.125 (0.05)	0.031 (0.01)
Moxifloxacin hydrochloride	2.5 (1.1)	0.0625 (0.03)	0.625 (0.27)	0.0313 (0.01)
Difloxacin hydrochloride	2.5 (1.1)	0.0625 (0.03)	0.313 (0.14)	0.0313 (0.01)
Gatifloxacin	2 (0.8)	0.0625 (0.02)	0.5 (0.2)	0.031 (0.01)
Oxolinic acid	20 (5.2)	0.625 (0.16)	2.5 (0.65)	0.078 (0.02)
Enrofloxacin	2 (0.7)	0.0625 (0.02)	0.5 (0.18)	0.016 (0.01)
Pefloxacin mesylate	2 (0.9)	0.25 (0.11)	0.5 (0.21)	0.0625 (0.03)
Sparfloxacin	2 (0.8)	0.031 (0.01)	0.5 (0.2)	0.016 (0.01)
Lomefloxacin	2.5	0.125	0.625	0.125
Ciprofloxacin	1 (0.33)	0.0625 (0.02)	0.25 (0.08)	0.031 (0.01)
Linezolid	>200 (>67.5)	200	200	25

Table 2. MICs of antibiotics analyzed by LC–MS in four strains of *P. aeruginosa* grown in MOPS-M9 medium.

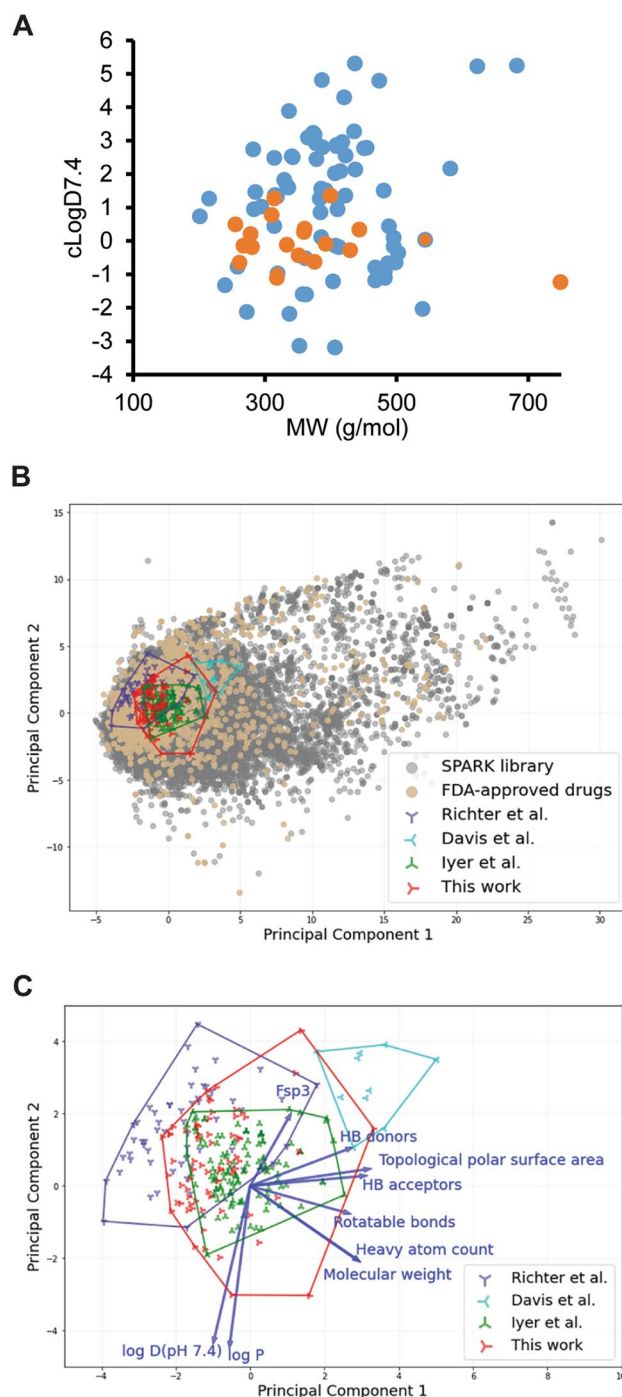


Figure 2. Properties of the library of 66 compounds analyzed in this study. **(A)** Size and lipophilicity of the analyzed compounds as represented by MW and cLogD_{7.4} descriptors. Antibiotics are highlighted in orange. **(B)** Principal component analysis of diversity in the chemical space of molecules from SPARK library³⁵, Richter et al.¹⁶, Davis et al.¹⁵, Iyer et al.²⁰, FDA- approved drugs (extracted from³⁶) and this study. The outermost data points from the four studies are linked to show the relative size of the sampled space. **(C)** Zoom-in view of **(B)** and the loading vectors for this plot. Data points for the SPARK library and FDA-approved drugs are hidden for clarity. Molecular weight, hydrogen bond donors and acceptors, topological polar surface area, rotatable bounds and heavy atom count are the major contributing factors to PC1, whereas LogP, logD and the fraction of sp³ hybridized carbon atoms (Fsp³) are major contributors to PC2.

steady-state levels, with some compounds reaching the steady state in all four strains within the first 1–2 min of incubation and others requiring longer incubation times, especially in the case of PAO1 cells (Fig. 3). In

contrast, for some compounds longer incubation times led to a decrease in compound concentration (Fig. S2). Therefore, we chose to analyze the compound library accumulation in cells incubated with compounds for 1 min and 40 min.

The LC-MS method developed here enabled concentration-dependent quantification of 83 compounds including 22 antibiotics in four strains of *P. aeruginosa*. After subtraction of signals from the excluded volume of empty filters, the positive concentration values were obtained for 66 compounds, which were further analyzed (Table S1). For most of the compounds, the calculated concentration slopes of compound accumulation in four strains at 1 and 40 min correlated well with each other (Fig. 3A). At each concentration, the accumulation levels varied by orders of magnitude depending on the specific compound. However, with few exceptions, the differences between the strains were relatively small. Overall, we found that the accumulation levels in the “barrierless” PΔ3-Pore cells at 1 and 40 min trended positively with the ratios of concentration slopes obtained in PΔ3-Pore and PAO1 cells (Fig. 3B). This result suggests that for certain compounds the high accumulation levels in PΔ3-Pore are due to the removal of the permeation barriers.

To normalize for differences in the affinities of compounds for intracellular targets, we also used: (i) the efflux ratios expressed as the ratios of measured concentrations or concentration slopes in efflux-deficient and efflux-proficient cells with the hyperporinated OM (PΔ3-Pore/PAO1-Pore), (ii) the OM ratios expressed as the respective ratios in efflux-deficient cells with normal and hyperporinated OMs (PΔ3-Pore/PΔ3), and (iii) the total barrier (PΔ3-Pore/PAO1). No significant correlation was found between the efflux and the total barrier ratios, suggesting that the contribution of efflux varied broadly between compounds, and was not the major contributor to the intracellular accumulation (Fig. 3C). In contrast, the OM ratios clearly correlated with the total barrier (Fig. 3D). This result suggests that the permeation across the OM is the dominating factor for intracellular accumulation of the analyzed compounds.

Classification of compounds. The measured levels of compound accumulation were analyzed using machine learning to determine which features of the compounds are most important for enabling their cellular accumulation and to determine their optimal values. For modeling, we used compound concentration slopes of the accumulation levels observed after 1 min (A1) and 40 min (A2) incubations, which were approximated

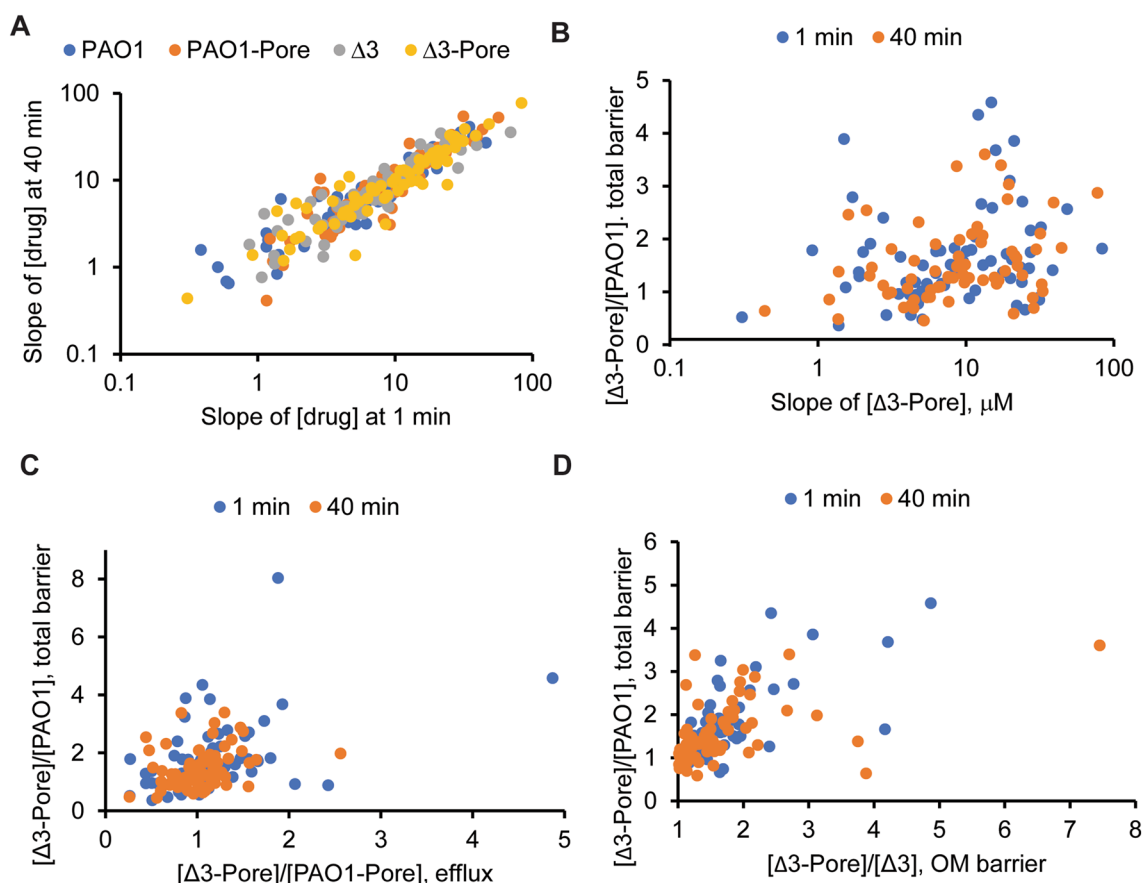


Figure 3. Intracellular accumulation of the analyzed library of 66 compounds. (A) Slopes of intracellular accumulation of compounds at 1 min and 40 min incubation times. (B) Plot of the slopes of measured intracellular accumulation in the “barrierless” strain as a function of the fold difference in accumulation in PAO1 and “barrierless” strain. (C) and (D) Comparison of fold change in the compound accumulation for the PΔ3-Pore/PAO1 (Total barrier) and PΔ3-Pore/PAO1-Pore (Efflux) or PΔ3-Pore/PΔ3 (OM).

as straight lines with y-intercept set to zero. To accentuate the permeability aspect, the accumulation slopes in efflux-deficient and hyperporinated strains were normalized to those in the parental PAO1 strain. In several cases, the slopes were lower in the permeabilized cells than in wild type cells, presumably due to data scatter or low signal (Table S1). Such changes cannot be attributed to the porination state or efflux deficiency of the cell. Therefore, we attributed such cases to the lack of an increase in permeability and set those ratios equal to 1.

To help visualize the data, we performed a principal component (PC) decomposition of the data. The first two principal components were split almost equally between efflux and hyperporination ratios (PC 1) or dominated by the absolute levels of compound accumulation (PC 2; Table S3). Elbow analysis of the unexplained variance suggested the existence of five clusters, which were well separated in the principal component view (Fig. 4A) and in native coordinates (Fig. 4B). All clusters comprised compounds that are affected to a various extent by *P. aeruginosa* permeability barriers. To formalize further analysis, we defined as the target subgroup those compounds that (i) accumulate at or above extracellular levels, (ii) exhibit relatively low signal on empty filters; and (iii) whose uptake does not increase by more than 30% by hyperporination combined with efflux inactivation (boxed in Fig. 4B). A total of 25 compounds satisfied these criteria and were classified as good permeators. Most of the good permeators were found in clusters 1 and 3.

Random Forest classification. The dataset of 66 compounds is relatively small, but it is comparable to previous studies^{16,20}. We first trained a random forest binary classification model using the above LC-MS data and 27 2D and 3D physicochemical descriptors for each compound that were pre-selected from a larger descriptor pool by clustering analysis (see Methods). The model was designed to distinguish between permeators and non-permeators as defined from the principal component analysis (Fig. 4B). The model achieved only 61% accuracy, 48% recall and 46% precision when tested on leave-one-out cross-validation (Fig. 5A). The area under the receiver-operator characteristic (AUROC) curve was only 0.53. The broad confidence interval (Fig. 5B) indicates that the performance of the model is largely dictated by which compounds are included in the training set for a given cross-validation step. This behavior implies that a larger compound set would likely significantly improve the precision of the model. Descriptors derived from molecular dynamics (MD) simulations are prevalent among the top ten, suggesting that molecular shape, conformational flexibility, and dynamics are important for distinguishing permeators from non-permeators (Fig. 5C). Examples include shape-based descriptors such as the Hall-Kier kappa3 molecular shape descriptor, the average acylindricity and asphericity, i.e., the deviation from cylindrical and spherical symmetries, and the radius of gyration (average R_g and average smallest principal R_g). Other top descriptors associated with favorable uptake include lipophilicity (MolLogP) and properties related to molecule topology (second principal moment of inertia, minimum absolute E-state index, and Morgan fingerprint density).

Gnostic classification tree. One of the limitations of the RF analysis is its propensity to use all provided descriptors without any regard for potential overlap between them. To address this issue, we developed a classification algorithm, Gnostic, that helps reduce the number of synonymous descriptors in the constructed classification model and ultimately constructs a single, unique classification tree. The descriptors identified during the construction of the tree define a property space that can be used for mapping compounds with diverse activities.

The algorithm employs several rounds of descriptor selection (Fig. 6A). First, the descriptors are clustered as described in Methods, and a single representative descriptor is selected from each cluster. These representatives are then ranked in the order of their predictability by the other descriptors, using a regression tree, in an iterative procedure that sequentially removes the most predictable descriptor from the list. In this procedure, the misclassification error of the regression tree defines the extent of the information overlap between the tested

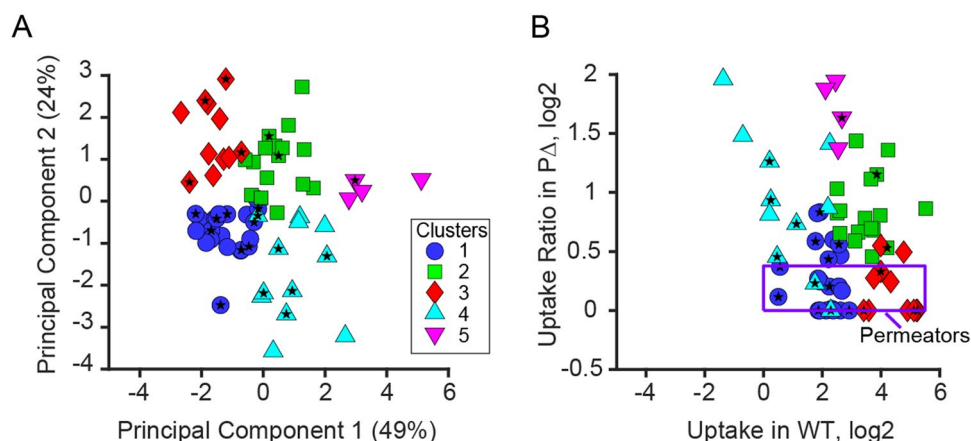


Figure 4. Distribution of uptake data in the activity space in PC (A) and native (B) coordinates. Crosses mark centers of clusters determined using *k*-means clustering. Violet rectangle encompasses compounds defined as good permeators. Stars mark antibiotics.

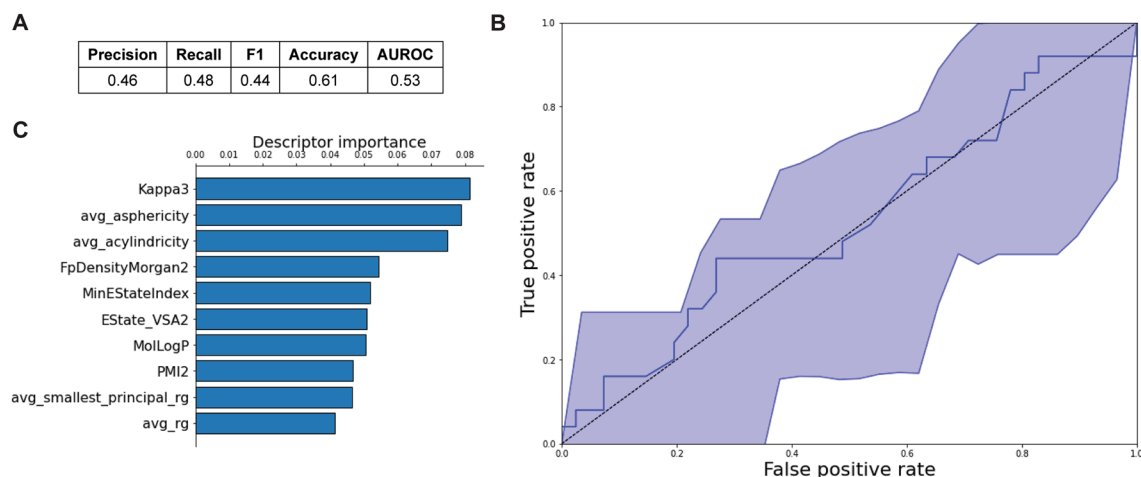


Figure 5. Random forest model performance metrics and diagrams. (A) Performance metrics, (B) ROC curve derived from leave-one-out cross validation with 95% confidence bands, (C) Importance bar plot of top ten features.

descriptors, i.e., the smaller the error, the larger the synonymy. The misclassification error also offers a natural measure of the stringency of the procedure. As a result, the synonymous descriptors are removed from the set prior to modeling, and the stringency of the filtering can be varied simply by changing the selection threshold. Thereafter, a classification tree can be trained on the uptake data at various levels of stringency, thereby finalizing the list of descriptors. To reduce the bias caused by the initial ranking of the descriptors, the order of descriptors was randomly permuted 100 times during the construction of the classification tree at each stringency level.

Using Gnostic, we constructed models that predict the five clusters defined by the permeability of the compounds (Fig. 6). We then selected the model that yields the highest overall precision, 79%. The accuracy of the models varied between 85 and 97% and recall rates between 58% for cluster 4 and 95% for cluster 1 (Fig. 6B, C). The model operated with 7 descriptors, including those reporting on the shape of the molecule, charge and its distribution, topology of the compound, number of rotatable bonds and the presence of aromatic rings (Fig. 6D).

We next visualized the model by plotting the data in the descriptor space (Fig. 6E). Despite the high precision of the model, there was no clear set of descriptor values that defined any of the clusters. Data points from most clusters covered a sprawling, irregularly shaped chemical space that spread throughout the entire sampled region. In contrast, the separation between the clusters was much smaller than their reach.

We next mapped the subset of good permeators defined in Fig. 4B to the property space. The 25 compounds found in this set formed six clusters, one of which included only a single compound (Fig. 6E). Hereafter, we refer to these clusters in the property space as nodes to distinguish them from the clusters in the activity space defined in Fig. 4. A single cluster in the permeability space usually transformed into multiple nodes when viewed in the property space (Fig. 6E) and vice versa. The nodes comprising multiple compounds were mostly devoid

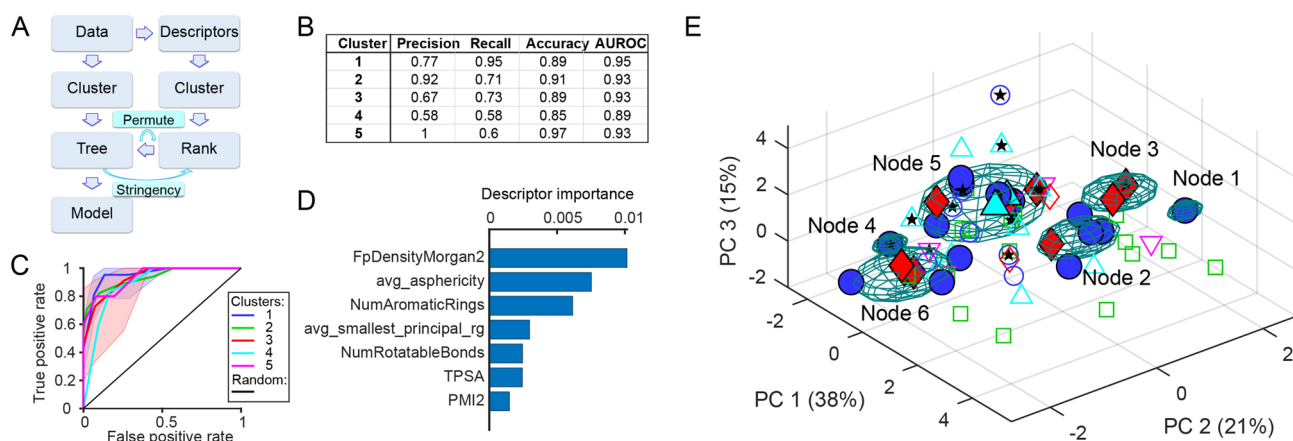


Figure 6. Gnostic model of compound accumulation in *P. aeruginosa*. (A) Gnostic algorithm. (B) Performance metrics for the prediction of each cluster. (C) Receiver operating curve of the classification model. Shaded area marks the confidence bands for cluster 2. (D) Descriptors selected by training the model. (E) Distribution of the clusters in the property space. Color coding the same as in Fig. 4.

of non-permeating molecules, indicating that the constructed map represents at least a fraction of the space attainable to good permeators.

Notably, some known antibiotics were found within or in the vicinity of some, but not all identified regions. Node 5 comprises six fluoroquinolones and linezolid, whereas another fluoroquinolone prulifloxacin is in node 3. Node 4 is sulfamethazine. For these antibiotics, the efflux (PAO1-Pore/PΔ3-Pore) and the OM (PΔ3/PΔ3-Pore) MIC ratios varied broadly (Table 2). This result agrees with previous studies that showed a lack of simple correlations between permeation and the antibacterial activities of antibiotics^{20,23,24}. Compounds from the other nodes have not been previously associated with efficient permeation into bacteria. Thus, this study identifies new property regions in which high permeability into *P. aeruginosa* is expected. The optimal values of the descriptors that define these regions are listed in Table 3.

Discussion

In this study, we have developed an approach to analyze the permeation of small molecules into Gram-negative bacteria and to identify physicochemical properties associated with the trans-envelope permeation. We applied this approach to a library of 66 structurally diverse compounds and identified “good” permeators and their properties. The experimental approach is based on LC–MS quantification of compound concentrations, a technique used previously to measure the intracellular accumulation of antibiotics and series of compounds with antibacterial activities^{14–16,20,37}. A critical bottleneck in studies of intracellular accumulation is the separation of molecules located inside the cells from external compounds. To achieve such separation, the majority of previously reported approaches used either centrifugation through silicon oils followed by freezing and cutting cell pellets^{16,20}, or vacuum filtration onto low-binding filters followed by washing the filters with an ice-chilled buffer or medium^{37–39}. Depending on the approach, these separation steps are completed within several minutes and could strongly affect results for compounds that lack high affinities to extra- or intra-cellular components^{23,40} and equilibrate rapidly across membranes (Figure S1). We evaluated the approach that avoids a washing step and found it informative for analyses of compounds with and without high affinity intracellular sites.

Another important question in such measurements is how to separate “good” permeators from “bad”. The different levels of accumulation in cells with native versus disrupted permeability barriers identify compounds that are affected by specific barriers (Fig. 1) but do not necessarily distinguish between “good” and “bad” permeators into wild type cells. The absolute measured concentrations may be misleading because of non-specific binding to filters or cell debris, even for the high-affinity intracellular binders and methods with washing steps, as well as leakage of compounds from “barrierless” cells. In the presented approach, there is a simple intrinsic control for “bad” permeators: such compounds are identified from their increased intracellular accumulation in hyperporinated and efflux-deficient cells. However, careful measurements and subtraction of non-specific signal was required to identify “good” permeators. We identified compounds with strong non-specific binding to empty filters and such compounds were excluded from further analyses. We further relied on time- and concentration-dependent changes in compound concentrations and defined the “good” permeators as compounds that accumulate in PAO1 cells linearly with increasing concentration but that are not affected by compromising permeation barriers and have low non-specific retention on control filters.

The molecular signature of good permeators consisted of a mixture of distinct properties including the size and shape of the molecule, polarity, topology, and a handful of molecular fingerprints such as the number of aromatic rings or rotatable bonds (Fig. 6). These properties are often found in this type of analysis and can be readily traced to the mechanism of small molecule penetration into bacteria^{15,16,30}. The true challenge was to find numeric values of these parameters that are conducive to penetration.

To solve this problem, we performed two types of analysis. First, we trained an RF model to recognize good permeators. However, due to the limited size and broad coverage of chemical space in the dataset, the performance metrics for the RF model were unsatisfactory (Fig. 5). Thus, a second machine learning approach was designed to address this problem. We developed the Gnostic algorithm (Fig. 6A), which ranks descriptors according to their unpredictability and thereby helps select descriptors with minimal information overlap. We reasoned that this approach would create an acceptable compromise between absorbing unique molecular signatures available in diverse descriptors and avoiding the synonymy between them. Clearly, the use of synonymous descriptors in a model only expands the dimensionality of the solution without improving its predictive power.

Descriptor ^a	Node						AB
	1	2	3	4	5	6	
FpDensityMorgan2	1.5 ± 0	1.8 ± 0.1	2.0 ± 0.0	1.6 ± 0	2.1 ± 0.2	1.9 ± 0.1	2.0 ± 0.2
avg_asphericity	0.52 ± 0	0.40 ± 0.03	0.39 ± 0.00	0.14 ± 0	0.29 ± 0.04	0.22 ± 0.02	0.31 ± 0.06
NumAromaticRings	2.0 ± 0	2.4 ± 0.5	3.0 ± 0.0	2.0 ± 0	1.8 ± 0.4	2.0 ± 0.9	2.0 ± 0.5
avg_smallest_principal_rg	0.22 ± 0	0.23 ± 0.02	0.24 ± 0.02	0.24 ± 0	0.22 ± 0.01	0.29 ± 0.01	0.23 ± 0.03
NumRotatableBonds	4.0 ± 0	3.8 ± 0.4	4.0 ± 0.0	3.0 ± 0	3.1 ± 1.3	6.4 ± 1.0	3.7 ± 1.8
TPSA	93 ± 0	66 ± 8	116 ± 3	98 ± 0	85 ± 14	61 ± 10	98 ± 32
PMI2, × 1000	10.8 ± 0	6.9 ± 1.1	10.7 ± 0.9	1.9 ± 0	3.5 ± 1.1	3.9 ± 1.2	4.5 ± 2.5

Table 3. Values of top descriptors in the center of permeation nodes and their comparison to antibiotics. ^aSee Table S3 for definitions. AB-antibiotics.

Traditionally, ML methods address this issue through the use of curvature tests, which eliminate highly correlated descriptors. The descriptor clustering step builds on this idea and sets an adaptive threshold for the acceptable correlation. However, these steps capture only pairwise correlations between descriptors and often overlook the instances when the information contributed by one descriptor is fully contained in a group of others. Gnostic searches for such instances by building regression models that predict a given descriptor through the rest. A low misclassification error in such models signifies a low informational contribution from said descriptor. This approach can be viewed as an ML equivalent of the linear algebra concept of a linear independence between vectors. Unlike in linear algebra, the predictability can recapture instances of non-linear interaction between descriptors. This feature makes it better suited for the analysis of seemingly stochastic distributions such as the one examined here.

The RF model used all 27 descriptors. In contrast, the Gnostic model identified seven descriptors with non-zero importance (Table S4). These seven descriptors comprise a reference set in the property space for inspecting compounds. The top descriptors of good permeators mainly define topological, constitutional, and molecular properties of compounds (Table S4). Among these, shape (avg_asphericity) and number of rotatable bonds (NumRotatableBonds) were highlighted in previously developed models based on the studies of either intracellular accumulation or antibacterial activities^{15,16,30}. However, a closer inspection shows that the numerical values of these descriptors vary broadly between different permeation nodes and do not clearly distinguish them (Table 3).

The 25 good permeators were broadly distributed in the property space with at least six nodes in diverse scaffolds (Fig. 6E). This distribution suggests that the region occupied by good permeators is not truly fractal but has a continuous fabric at its base. We posit that other compounds that map to these nodes would penetrate *P. aeruginosa*. Moreover, other permeability nodes should be discoverable by further testing. Overall, this approach suggests a framework for mapping compound permeabilities in the chemical space and the development of predictive models for antibiotic design based on them.

The near fractal character of the permeator region explains prior difficulties in the construction of linear regression models of permeation or the discovery of a universal chemical modification that would permeabilize the molecule. Such modifications ought to be scaffold dependent. On the bright side, good permeators can be found in the vicinity of many scaffolds. Some of these scaffolds are already well known. For example, nodes 3, 4 and 5 contain antibiotics. Importantly, antibiotics from the same class, e.g., the fluoroquinolones norfloxacin and prulifloxacin can be found in several nodes because they vary significantly in their structures and properties (Tables 3 and S1). The 2-oxo-1,3-dioxol substitution on the piperazine ring and thiazeto modification on the quinoline ring of prulifloxacin affect size, shape, charge and other properties that are important for permeation (Table 3). Among good permeators are also human therapeutics, e.g., ergotamine and sitagliptin. Both drugs are oral and hence, display good epithelial permeability, and these chemotypes may contain useful structural motifs and/or connectivity patterns that can be leveraged to enhance Gram-negative permeation of other scaffolds.

The inspection of structures of good permeators did not yield a single functional group consistently present in all compounds. However, some of the moieties such as trifluoro- or diaminomethyl- substitutions are known to improve permeation across bacterial membranes^{31,41}. Our results also show that the contribution of such chemical modifications to permeation across *P. aeruginosa* and other Gram-negative cell envelopes will be scaffold dependent. Further analyses of intracellular permeation and modeling of larger chemical libraries around several biologically active scaffolds using hit expansion techniques will lead to a more detailed mapping of the space of good permeators and could potentially reveal scaffold-dependent heuristics.

Optimization of antibiotic activities against Gram-negative bacteria could be strongly facilitated if the effect of a given chemical modifications on the antibacterial activity could be separated from the effect on its efflux and permeation across the OM^{3,32}. In agreement with previous studies^{20,37}, we did not find a numeric correlation of the intracellular uptake of antibiotics and their antibacterial activities (Table 2). This lack of agreement between the two approaches is the result of technical differences: MIC values report on the concentration needed to kill growing bacterial cells, whereas LC/MS data report on time-dependent changes in the intracellular concentrations of compounds in non-growing cells. However, both datasets are usable to derive permeation rules when concentrations are measured in wild type, hyperporinated and efflux-deficient cells^{24,30}. Future studies will be focused on identification of properties that are needed to develop models in which antibacterial activities of antibiotics could be predicted based on the permeation studies.

Methods

Materials. MS grade water, acetonitrile and formic acid were used for the LC-MS method. Radioactively labeled compounds [³H]-oxymetazoline hydrochloride (specific activity 1.5 Ci/mmol), [¹⁴C]-nepafenac (specific activity 50.0 mCi/mmol), [¹⁴C]-efavirenz (specific activity 56.7 mCi/mmol), [³H]-delavirdine (specific activity 1.2 Ci/mmol), [³H]-tetracycline (specific activity 17.9 Ci/mmol), [¹⁴C]-ciprofloxacin (specific activity 58.8 mCi/mmol), [³H]-metoprolol (specific activity 29.5 Ci/mmol) and their unlabeled analogs were purchased from Moravex Biochemicals Inc. (CA, USA). Other compounds were purchased from ChemDiv (CA, USA), MicroSource, Otava and TimTec or requested from National Cancer Institute. Compound registration and assay data management were carried out by Collaborative Drug Discovery.

Strains and growth media. In all experiments, *P. aeruginosa* PAO1-Pore (PAO1 attTn7::P_{LAC}-*fhuA*) and PΔ3-Pore (Δ*mexAB* Δ*mexCD* Δ*mexXY* attTn7::P_{LAC}-*fhuA*) strains were used²⁹. Strains were routinely grown in Luria-Bertani (LB) or MOPS-M9 medium at 37 °C with aeration. MICs shown in Tables 1 and 2 were determined by a two-fold dilution method in LB broth and MOPS-M9 medium, respectively, following the previously described protocol³⁰.

Intracellular accumulation of radioactively labeled compounds. PAO1-Pore and PΔ3-Pore were grown at 37 °C overnight with shaking at 225 rpm in LB medium. Cells were sub-cultured (50–100 mL) 1:100 in fresh LB broth in two flasks per strain. At OD₆₀₀ 0.3 (~1.5 to 2 h of incubation), the expression of the pore was induced with 0.1 mM IPTG (final concentration) in one of the flasks and cells were grown for additional 1 to 1.5 h (OD₆₀₀ 1–1.5). Cells were collected by centrifugation at room temperature (RT) and washed in PMG Buffer (50 mM KPi, pH 7.0, 1 mM MgSO₄, 0.4% glucose, prepared fresh) and concentrated 20x (i.e., 100 mL culture resuspended in 5 mL of PMG Buffer with final OD₆₀₀ ~20–30). Assays were performed in 96-well 1.0 μm Glass Fiber Type B filters plates (Millipore). Cells were incubated with indicated concentrations of compounds, 100 μL cell aliquots were withdrawn after 0.5, 1, 2, 4, 6, 8, 16, and 32 min incubation at RT and collected by vacuum filtration onto filters.

Because compounds that penetrate the cells but do not have intracellular binding sites could be washed out by buffer, we first analyzed how washing of cells affects retention of radioactively labeled compounds (Fig. S1). We found that washing of filtered cells with buffer differentially affects compounds with variable intracellular affinities. Therefore, the washing step was avoided. Filter plates were dried, removed from the vacuum manifold and left in the hood overnight to dry completely. Filters were punched out from the plates, placed into 1 mL of scintillation counter solution and radioactivity counted using Tri-Carb 2810 TR Liquid Scintillation Analyzer (Perkin Elmer).

Preparation of cells for LC–MS analysis. PAO1-Pore and PΔ3-Pore were grown at 37 °C overnight with shaking at 225 rpm in 50 mM MOPS-M9 (pH 7.2) Minimal Medium supplemented with 1% glycerol^{42,43}. Importantly, a vancomycin spot assay showed that hyperporination of the OM was efficient in this medium and cells became highly susceptible to this antibiotic when incubated in the presence of inducer (Fig. S5).

The next day, cells were sub-cultured (1.5:100) in fresh MOPS-M9 medium and grown for 16 h. The cultures were then split into two flasks and one of them was induced with 0.1 mM IPTG. Cells incubated for additional 4 h until OD₆₀₀ reached ~1.0 and harvested by centrifugation at 3,275xg for 20 min in RT, washed and concentrated 10X in the spent medium. Assays were performed in 96-well plates in a final volume of 200 mL. Cells were incubated with 10, 20, 40, 80 μM of compounds and after 1- and 40-min incubation at RT, 100 μL cell aliquots were collected by vacuum filtration onto 1.0 μm Glass Fiber Type B filters. For no-cell controls, compounds were diluted to 10, 20, 40, 80 μM final concentration in the same medium, incubated for 1 min and 40 min and filtered onto the same plates. The cell loaded and cell-free filters were dried, removed from plates and placed into 100% methanol at -80 °C for at least 10 min. Filter-bound material was extracted by water bath sonication for 1 min. Cell and filter debris were separated by ultracentrifugation at 45,000 rpm (TLA55 rotor, Beckman) for 8 min at 16 °C and the pellet re-extracted with 80% methanol in water by sonication for 15 min followed by ultracentrifugation at the same conditions. Supernatants from two extractions were combined. For the compound quantification, 5 μL of solution was analyzed in triplicate.

Calibration curves for MS analyses of compounds were generated in the same experiment by mixing compound at increasing concentrations with the sonicated PAO1 cell extracts. For most of the tested compounds, a linear calibration curve was obtained using the LC–MS approach when compounds were analyzed in a pure solvent at concentrations in the 10 nM to 1000 nM range (Fig. S6). However, in cells extracted from filters, the detection of compounds at the external compound concentration below 10 mM was unreliable. Levofloxacin at 100 nM final concentration was used as an external control and added to each sample after extraction¹⁴. Azithromycin was used as a positive control in all experiments. We found that azithromycin is readily detectable in the presence of levofloxacin and the kinetics of its accumulation is reflective of differences in permeability properties of the four strains of *P. aeruginosa* (Fig. S7).

We found that for 80% of the compounds (66 out of 83) the signal on filters with cells was higher than from empty filters. For the remaining compounds, the signal from empty filters was similar to or higher than from filters with cells and these compounds were excluded from further analyses.

LC–MS quantification of intracellular accumulation of compounds. 200 mL of samples supplemented with 100 nM levofloxacin as external standard were transferred into glass vials and 10 μL of samples and standards were injected into the UPLC in triplicates. A UPLC/HRMS system Agilent 1290 Infinity UPLC coupled to an Agilent 6545 quadrupole time-of-flight mass spectrometer was used to quantify the concentrations of compounds. Samples were processed in positive ion mode. An Eclipse plus C18 Rapid Resolution HD column (1.8 μm, 2.1 × 50 mm, Agilent, USA) was used for separation. The flow rate was 0.6 mL/min using the following gradient: 0 to 1 min (0–20% ACN), 1 to 4 min (20–80% ACN), 4 to 5.1 min (80–100% ACN), 5.1 to 6.30 (100% ACN). Solvents contained 0.1% formic acid to promote positive ion formation in the electrospray. The MS parameters were as follows: ion-source gas temperature, 325 °C; capillary voltage, 4000 V; fragmentor voltage, 180 V; *m/z* range, 50–1100; data acquisition rate, 4 GHz; and 2 spectra were recorded per second. Data were analyzed using Agilent Masshunter workstation quantitative analysis v. B.10 software.

Determination of volumes of loaded cells and empty filters. To determine the cell-free volume of filters and the volume of filters occupied by loaded cells, we used microspheres (20 nm) labeled with Nile Red fluorophore (FluoroSpheres, ThermoFisher Scientific, Cat# F8784). These microspheres are too large to penetrate the cells and bind to cell surfaces and they are expected to be retained only by the cell-free volume of the filters. FluoroSpheres in the final concentrations of 10, 20, 40 and 80 mg/mL were incubated for 40 min without and with PAO1 cells (1.0 OD) in 100 mL of MOPS-M9 medium. FluoroSphere solutions with and without cells were vacuumed onto the filters, filter dried, removed from plates and fluorospheres were extracted twice with 0.3 mL HPLC-grade water. Fluorescence was quantified using a microplate reader and converted into amounts

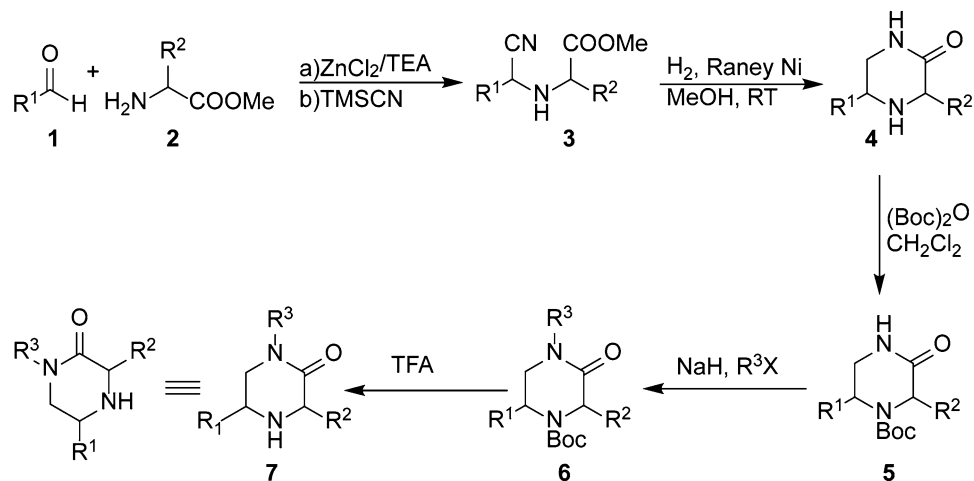


Figure 7. Synthesis of trisubstituted piperazin-2-one.

retained on filters. The retained volumes were determined from the slopes of the amounts of FluoroSpheres retained by filters with and without cells plotted as a function of concentration of solutions (Fig. S1). The difference in volumes is the volume displaced by cells.

Synthesis of trisubstituted piperazin-2-one derivatives. Our synthetic approach for building the trisubstituted piperazin-2-one derivatives involves two key steps: a modified Strecker reaction for the synthesis of cyanomethylamino pseudodipeptides (3), followed by a reductive cyclization to provide the core piperazin-2-ones (4), which could be functionalized accordingly to yield desired analogs. The synthesis commenced by treatment of the aldehyde 1 with the amino acid methyl ester 2 and TMSCN in presence of catalytic $ZnCl_2$ ⁴⁴. The pseudodipeptide 3 was obtained as an epimeric mixture at the stereogenic center of the peptide bond surrogate [(1R)/(1S) 1:3], which could not be resolved by chromatography. Reductive cyclization of pseudodipeptides 3 by catalytic hydrogenation resulted in *in situ* lactamization, in the presence of Raney Ni as the catalyst and led to the 5-substituted piperazin-2-ones 4 (Fig. 7) in 52–76% yield. The diastereomeric ratio remained unchanged in this reductive cyclization [(R)/(S) 1:3]. Boc protection of the secondary amine (5) was followed by alkylation of the secondary amide in the presence of NaH in DMF and the requisite alkyl halides to provide (6). Removal of the Boc-group in the presence of TFA yielded the final trisubstituted piperazin-2-ones 7.

Computational and machine learning methods. *Cheminformatics.* The chemical diversity of our compounds was compared to that of five other libraries and intracellular accumulation studies, including the Shared Platform for Antibiotic Research and Knowledge (SPARK) Public Projects Data¹ (archived by Collaborative Drug Discovery Vault), FDA-approved drugs³⁶, and small molecule permeability in bacteria studies by Richter et al.¹⁶, Davis et al.¹⁵, and Iyer et al.²⁰. In this comparison, nine features were used to describe each molecule: the molecular weight, logP, the number of hydrogen bond donors, the number of hydrogen bond acceptors, log D at pH 7.4, the topological polar surface area, Fsp3, the heavy atom count and the number of rotatable bonds. Molecular features for molecules from the four intracellular accumulation studies were calculated using the Marvin suite from ChemAxon⁴⁵. All molecules were filtered to retain only those with molecular weights between 200 and 1,500 g/mol and to remove ions and peptides. Standardization and principal component analysis were then applied to the data points using scikit-learn version 0.24.1⁴⁶. To illustrate how much each feature contributes to a particular principal component, PCA loadings are labeled in Fig. 2B. PCA loadings are defined as the coefficients of the linear combination of the original variables from which the principal components are constructed. Loading vectors were scaled by a factor of 7 for clarity. Figures were generated by matplotlib (version 3.3.4)⁴⁷ and seaborn (version 0.11.1)⁴⁸.

Physicochemical property calculation. Tautomers and protonation states at pH 7.4 for each compound were predicted using Marvin from ChemAxon⁴⁵. SMILES strings were converted to 3D coordinates and a maximum of 50 low-energy conformers within an energy window of 10 kcal/mol were generated with the *RDKitGenerate-Conformers.py* script from MayaChemTools⁴⁹ using the ETKDG conformer generator⁵⁰ and the MMFF94s force field⁵¹. Both 2D and 3D descriptors were calculated with the *RDKitCalculateMolecularDescriptors.py* script from MayaChemTools and then averaged for use in subsequent machine learning analyses. RDKit version 2020.09.1 was used in all cases.

We also adapted a previously published protocol for computing force field parameters and running molecular dynamics (MD) simulations of antimicrobial compounds to extract physicochemical properties from the resulting trajectories⁵². Using the structures from the tautomer prediction as starting coordinates, we optimized the geometries of each compound at the quantum chemical density functional theory B3LYP/6-31G(d,p)^{53–55} with the polarizable continuum solvent model⁵⁶ using Gaussian16 revision A.03⁵⁷. Partial charges were generated using

the restrained electrostatic potential (RESP)⁵⁸ method in the *antechamber* utility⁵⁹ and 100 ns MD simulations were then performed with AMBER 16⁶⁰ using general AMBER force field (GAFF) parameters and the TIP3P water model⁶¹ with 0.1 M KCl. Additional details on the MD simulations have been described previously⁵². The PLUMED plugin⁶² was used to calculate several shape descriptors, including asphericity, acylindricity, k^2 , and the radius of gyration. These descriptor values were then averaged over the MD simulation and used for machine learning analysis. In total, 217 descriptors were calculated from MD simulations and with RDkit.

Selection of descriptors. The initial list of 217 descriptors was manually inspected to select only those that represent integral physicochemical or topological properties of a molecule or generic molecular fingerprints. The selected 47 descriptors were then clustered and analyzed by a modified elbow-method. Namely, the unexplained variance was calculated as a function of the number of clusters, fit to a three-segmented straight line, and the larger elbow used as a cutoff. Descriptors closest to the centers of the identified 30 clusters were subsequently used in ML analyses.

Random forest classification. A procedure consisting of data standardization, oversampling, and random forest (RF) classification was developed for this task. All procedures were performed with Python 3.8.8 in a Jupyter Notebook environment (Available at https://github.com/yshen25/EPI_ML) with scikit-learn version 0.24.1⁶³ and imbalanced-learn version 0.8.0⁶⁴. Of the 66 compounds, 25 were identified as permeators and the remaining 41 were labeled as non-permeators (Fig. 5). Because the dataset is imbalanced, the use of an oversampling technique was expected to be beneficial. Based on the mapping result (Fig. 6), the permeators belong to six different clusters in feature space, so common techniques such as SMOTE⁶⁵ and ADASYN⁶⁶ may produce noise points because synthetic data points may be situated in overlapping class regions. To overcome this issue, k-means SMOTE⁶⁷ was applied. By giving more weight to sparse minority areas and avoiding overlapping zones, k-means SMOTE is expected to be better at overcoming skewness. Physicochemical properties were standardized to have mean and standard deviation of 0 and 1.

To obtain increased accuracy in the RF model, three hyperparameters were tuned using an exhaustive grid search. Hyperparameters included the number of trees in the random forest ($n_{estimators}$), the minimum number of samples required at a leaf node ($min_samples_leaf$), and the maximum number of features considered at each split ($max_features$). The hyperparameter combination with the highest accuracy was then selected for model training and testing. The performance of the final model was evaluated by leave-one-out cross validation. The 95% confidence interval was calculated by performing five-fold stratified cross validation repeated five times. Model performance was evaluated by the following scoring metrics. The *accuracy* describes the portion of samples that are correctly classified and is defined as follows:

$$accuracy = \frac{true\ positives + true\ negatives}{all\ samples}$$

Precision is the proportion of true positives among all predicted positives, also called *positive predictive value*:

$$precision = \frac{true\ positives}{true\ positive + false\ positives}$$

Recall is the proportion of true positives among all actual positives, also called *sensitivity*.

$$recall = \frac{true\ positives}{true\ positives + false\ negatives}$$

The *F1* score assesses the balance between *precision* and *recall*:

$$F1 = 2 \times (precision \times recall) / (precision + recall)$$

The receiver operating characteristic (ROC) curve plots the true positive rate versus the false positive rate. The area under the ROC curve, AUROC, quantifies the ability of the model to discriminate between classes, with 1.0 corresponding to a perfect classifier and 0.5 corresponding to random.

The Gnostic program was developed using MATLAB 2019a with the Statistics and Machine Learning toolbox⁶⁸.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Received: 10 February 2022; Accepted: 10 May 2022

Published online: 17 May 2022

References

1. Buhl, M., Peter, S. & Willmann, M. Prevalence and risk factors associated with colonization and infection of extensively drug-resistant *Pseudomonas aeruginosa*: A systematic review. *Expert Rev. Anti. Infect. Ther.* **13**, 1159–1170. <https://doi.org/10.1586/14787210.2015.1064310> (2015).
2. Behzadi, P., Baráth, Z. & Gajdács, M. It's not easy being green: A narrative review on the microbiology, virulence and therapeutic prospects of multidrug-resistant *pseudomonas aeruginosa*. *Antibiotics (Basel)* <https://doi.org/10.3390/antibiotics10010042> (2021).

3. Zgurskaya, H. I. & Rybenkov, V. V. Permeability barriers of Gram-negative pathogens. *Ann. N. Y. Acad. Sci.* **1459**, 5–18. <https://doi.org/10.1111/nyas.14134> (2020).
4. Zgurskaya, H. I. An old problem in a new light: Antibiotic permeation barriers. *ACS Infect. Dis.* **6**, 3090–3091. <https://doi.org/10.1021/acinfecdis.0c00780> (2020).
5. Lewis, K. The science of antibiotic discovery. *Cell* **181**, 29–45. <https://doi.org/10.1016/j.cell.2020.02.056> (2020).
6. Silver, L. L. Challenges of antibacterial discovery. *Clin. Microbiol. Rev.* **24**, 71–109. <https://doi.org/10.1128/CMR.00030-10> (2011).
7. Masungi, C. *et al.* Parallel artificial membrane permeability assay (PAMPA) combined with a 10-day multiscreen Caco-2 cell culture as a tool for assessing new drug candidates. *Pharmazie* **63**, 194–199 (2008).
8. Artursson, P., Palm, K. & Luthman, K. Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv. Drug Deliv. Rev.* **46**, 27–43. [https://doi.org/10.1016/s0169-409x\(00\)00128-9](https://doi.org/10.1016/s0169-409x(00)00128-9) (2001).
9. Schaich, M. *et al.* An integrated microfluidic platform for quantifying drug permeation across biomimetic vesicle membranes. *Mol. Pharm.* **16**, 2494–2501. <https://doi.org/10.1021/acs.molpharmaceut.9b00086> (2019).
10. Berben, P. *et al.* Drug permeability profiling using cell-free permeation tools: Overview and applications. *Eur. J. Pharm. Sci.* **119**, 219–233. <https://doi.org/10.1016/j.ejps.2018.04.016> (2018).
11. Avila, Q. P., H.I., Z. & Duerfeldt, A. S. *Recent advances towards rational antibacterial discovery: Addressing permeation and efflux.*, Vol. 52 (MEDI, Inc., 2017).
12. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26. (2001).
13. Burns, A. R. *et al.* A predictive model for drug bioaccumulation and bioactivity in *Caenorhabditis elegans*. *Nat. Chem. Biol.* **6**, 549–557. <https://doi.org/10.1038/nchembio.380> (2010).
14. Cai, H., Rose, K., Liang, L. H., Dunham, S. & Stover, C. Development of a liquid chromatography/mass spectrometry-based drug accumulation assay in *Pseudomonas aeruginosa*. *Anal. Biochem.* **385**, 321–325. <https://doi.org/10.1016/j.ab.2008.10.041> (2009).
15. Davis, T. D., Gerry, C. J. & Tan, D. S. General platform for systematic quantitative evaluation of small-molecule permeability in bacteria. *ACS Chem. Biol.* **9**, 2535–2544. <https://doi.org/10.1021/cb5003015> (2014).
16. Richter, M. F. *et al.* Predictive compound accumulation rules yield a broad-spectrum antibiotic. *Nature* **545**, 299–304. <https://doi.org/10.1038/nature22308> (2017).
17. Huigens, R. W. *et al.* A ring-distortion strategy to construct stereochemically complex and structurally diverse compounds from natural products. *Nat. Chem.* **5**, 195–202. <https://doi.org/10.1038/nchem.1549> (2013).
18. Chopra, I. & Hacker, K. Uptake of minocycline by *Escherichia coli*. *J. Antimicrob. Chemother.* **29**, 19–25. <https://doi.org/10.1093/jac/29.1.19> (1992).
19. McCaffrey, C., Bertasso, A., Pace, J. & Georgopapadakou, N. H. Quinolone accumulation in *E. coli*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **36**, 1601–1605. <https://doi.org/10.1128/aac.36.8.1601> (1992).
20. Iyer, R. *et al.* Evaluating LC-MS/MS to measure accumulation of compounds within bacteria. *ACS Infect. Dis.* **4**, 1336–1345. <https://doi.org/10.1021/acinfecdis.8b00083> (2018).
21. Asuquo, A. E. & Piddock, L. J. Accumulation and killing kinetics of fifteen quinolones for *E. coli*, *Staphylococcus aureus* and *Pseudomonas aeruginosa*. *J. Antimicrob. Chemother.* **31**, 865–880. <https://doi.org/10.1093/jac/31.6.865> (1993).
22. Piddock, L. J. V., Jin, Y. F. & Griggs, D. J. Effect of hydrophobicity and molecular mass on the accumulation of fluoroquinolones by *Staphylococcus aureus*. *J. Antimicrob. Chemother.* **47**, 261–270. <https://doi.org/10.1093/jac/47.3.261> (2001).
23. Rybenkov, V. V. *et al.* The whole is bigger than the sum of its parts: Drug transport in the context of two membranes with active efflux. *Chem. Rev.* **121**, 5597–5631. <https://doi.org/10.1021/acs.chemrev.0c01137> (2021).
24. Saha, P., Sikdar, S., Krishnamoorthy, G., Zgurskaya, H. I. & Rybenkov, V. V. Drug permeation against efflux by two transporters. *ACS Infect. Dis.* **6**, 747–758. <https://doi.org/10.1021/acinfecdis.9b00510> (2020).
25. Tamber, S. & Hancock, R. E. in *Pseudomonas* Vol. 1 (ed J.-L. Ramos) 575–601 (Kluwer Academic/Plenum Publishers, 2004).
26. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593–656. <https://doi.org/10.1128/mmb.67.4.593-656.2003> (2003).
27. Ude, J. *et al.* Outer membrane permeability: Antimicrobials and diverse nutrients bypass porins in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* <https://doi.org/10.1073/pnas.2107644118> (2021).
28. Westfall, D. A. *et al.* Bifurcation kinetics of drug uptake by Gram-negative bacteria. *PLoS ONE* **12**, e0184671. <https://doi.org/10.1371/journal.pone.0184671> (2017).
29. Krishnamoorthy, G. *et al.* Synergy between active efflux and outer membrane diffusion defines rules of antibiotic permeation into gram-negative bacteria. *mBio* <https://doi.org/10.1128/mBio.01172-17> (2017).
30. Cooper, C. J. *et al.* Molecular properties that define the activities of antibiotics in *E. coli* and *pseudomonas aeruginosa*. *ACS Infect. Dis.* **4**, 1223–1234. <https://doi.org/10.1021/acinfecdis.8b00036> (2018).
31. Mehla, J. *et al.* Predictive rules of efflux inhibition and avoidance in *Pseudomonas aeruginosa*. *mBio* <https://doi.org/10.1128/mBio.02785-20> (2021).
32. Zhao, S. *et al.* Defining new chemical space for drug penetration into Gram-negative bacteria. *Nat. Chem. Biol.* **16**, 1293–1302. <https://doi.org/10.1038/s41589-020-00674-6> (2020).
33. O'Shea, R. & Moser, H. E. Physicochemical properties of antibacterial compounds: Implications for drug discovery. *J. Med. Chem.* **51**, 2871–2878. <https://doi.org/10.1021/jm700967e> (2008).
34. Richter, M. F. & Hergenrother, P. J. The challenge of converting Gram-positive-only compounds into broad-spectrum antibiotics. *Ann. N. Y. Acad. Sci.* **1435**, 18–38. <https://doi.org/10.1111/nyas.13598> (2019).
35. Thomas, J., Navre, M., Rubio, A. & Coukell, A. Shared platform for antibiotic research and knowledge: A collaborative tool to SPARK antibiotic discovery. *ACS Infect. Dis.* **4**, 1536–1539. <https://doi.org/10.1021/acinfecdis.8b00193> (2018).
36. in www.collaborativedrug.com (Burlingame, CA).
37. Widya, M. *et al.* Development and optimization of a higher-throughput bacterial compound accumulation assay. *ACS Infect. Dis.* **5**, 394–405. <https://doi.org/10.1021/acinfecdis.8b00299> (2019).
38. Six, D. A., Krucker, T. & Leeds, J. A. Advances and challenges in bacterial compound accumulation assays for drug discovery. *Curr. Opin. Chem. Biol.* **44**, 9–15. <https://doi.org/10.1016/j.cbpa.2018.05.005> (2018).
39. Li, X. Z., Livermore, D. M. & Nikaido, H. Role of efflux pump(s) in intrinsic resistance of *Pseudomonas aeruginosa*: Resistance to tetracycline, chloramphenicol, and norfloxacin. *Antimicrob. Agents Chemother.* **38**, 1732–1741. <https://doi.org/10.1128/aac.38.8.1732> (1994).
40. Kojima, S. & Nikaido, H. Permeation rates of penicillins indicate that *E. coli* porins function principally as nonspecific channels. *Proc. Natl. Acad. Sci. U S A* **110**, E2629–2634. <https://doi.org/10.1073/pnas.1310333110> (2013).
41. Mansbach, R. A. *et al.* Machine learning algorithm identifies an antibiotic vocabulary for permeating gram-negative bacteria. *J. Chem. Inf. Model* **60**, 2838–2847. <https://doi.org/10.1021/acs.jcim.0c00352> (2020).
42. Poblete-Castro, I., Wittmann, C. & Nikel, P. I. Biochemistry, genetics and biotechnology of glycerol utilization in *Pseudomonas* species. *Microb. Biotechnol.* **13**, 32–53. <https://doi.org/10.1111/1751-7915.13400> (2020).
43. Nikel, P. I., Romero-Campero, F. J., Zeidman, J. A., Goñi-Moreno, Á. & de Lorenzo, V. The glycerol-dependent metabolic persistence of *Pseudomonas putida* KT2440 reflects the regulatory logic of the GlpR repressor. *mBio* <https://doi.org/10.1128/mBio.00340-15> (2015).

44. Valdivielso, A. M., Ventosa-Andrés, P., García-López, M. T., Herranz, R. & Gutiérrez-Rodríguez, M. Synthesis and regioselective functionalization of piperazin-2-ones based on phe-gly Pseudopeptides. *Eur. J. Org. Chem.* **155–161**, 2013. <https://doi.org/10.1002/ejoc.201201221> (2013).
45. 4521.2.0, M. *ChemAxon*, (2021).
46. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
47. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95. <https://doi.org/10.1109/mcse.2007.55> (2007).
48. Waskom, M. L. seaborn: Statistical data visualization. *J. Open Source Softw.* **6**, 3021. <https://doi.org/10.21105/joss.03021#> (2021).
49. Sud, M. MayaChemTools: An open source package for computational drug discovery. *J. Chem. Inf. Model.* **56**, 2292–2297. <https://doi.org/10.1021/acs.jcim.6b00505> (2016).
50. Wang, S., Witek, J., Landrum, G. A. & Riniker, S. Improving conformer generation for small rings and macrocycles based on distance geometry and experimental torsional-angle preferences. *J. Chem. Inf. Model.* **60**, 2044–2058. <https://doi.org/10.1021/acs.jcim.0c00025> (2020).
51. Halgren, T. A. M. M. F. V. I. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **20**, 720–729. [https://doi.org/10.1002/\(SICI\)1096-987X\(199905\)20:7%3c720::AID-JCC7%3e3.0.CO;2-X](https://doi.org/10.1002/(SICI)1096-987X(199905)20:7%3c720::AID-JCC7%3e3.0.CO;2-X) (1999).
52. Mallocci, G. *et al.* A database of force-field parameters, dynamics, and properties of antimicrobial compounds. *Molecules* **20**, 13997–14021. <https://doi.org/10.3390/molecules200813997> (2015).
53. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652. <https://doi.org/10.1063/1.464913> (1993).
54. Kim, K. & Jordan, K. D. Comparison of density functional and MP2 calculations on the water monomer and dimer. *J. Phys. Chem.* **98**, 10089–10094. <https://doi.org/10.1021/j100091a024> (1994).
55. Pople, J. A. Quantum chemical models (Nobel Lecture). *Angew. Chem. Int. Ed.* **38**, 1894–1902. [https://doi.org/10.1002/\(SICI\)1521-3773\(19990712\)38:13/14%3c1894::AID-ANIE1894%3e3.0.CO;2-H](https://doi.org/10.1002/(SICI)1521-3773(19990712)38:13/14%3c1894::AID-ANIE1894%3e3.0.CO;2-H) (1999).
56. Tomasi, J., Mennucci, B. & Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **105**, 2999–3093. <https://doi.org/10.1021/cr9904009> (2005).
57. Gaussian 16 Rev. A.03 (Wallingford, CT, 2016).
58. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **97**, 10269–10280. <https://doi.org/10.1021/j100142a004> (1993).
59. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260. <https://doi.org/10.1016/j.jmgn.2005.12.005> (2006).
60. Case, D. *et al.* AMBER 2016, (2016).
61. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935. (1983).
62. Bonomi, M. *et al.* PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **180**, 1961–1972. <https://doi.org/10.1016/j.cpc.2009.05.011> (2009).
63. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
64. Lemaitre, G., Nogueira, F. & Aridas, C. K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* **18**, 559–563 (2017).
65. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Int. Res.* **16**, 321–357 (2002).
66. Haibo, H., Yang, B., Garcia, E. A. & Shutao, L. in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. 1322–1328.
67. Douzas, G., Bacao, F. & Last, F. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. *Inf. Sci.* **465**, 1–20. <https://doi.org/10.1016/j.ins.2018.06.056> (2018).
68. Version 2019b (The MathWorks Inc., Natick, Massachusetts, 2019).

Acknowledgements

This work was supported by Defense Threat Reduction Agency grant HDTRA1-14-1-0019 (H.I.Z. and V.V.R) and National Institutes of Health grant RO1-AI052293 (H.I.Z., J.M.P., J.C.S. and V.V.R), RO1-AI136795 (H.I.Z., V.V.R and A.S.D) and RO1-AI132836 (H.I.Z.). The content of the information does not necessarily reflect the position or the policy of the federal government, and no official endorsement should be inferred. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. We thank Rosemary O’Shea, James Aggen and John K. Walker for help in selection of compound libraries.

Author contributions

All authors read the manuscript and contributed to its writing. I.V.L. and J.W.W. measured antibacterial activities and performed accumulation experiments, V.B. developed LC–MS method and carried out compound quantification experiments, Y.S. carried out cheminformatics analyses and Random Forest classification, L.Y. performed accumulation experiments with radioactive compounds, C.J.C. calculated physico-chemical descriptors and carried out machine learning analyses, D.N. synthesized compounds, A.S.D. designed compounds and synthetic routes, J.C.S. and J.M.P. designed and supervised cheminformatics and Random Forest experiments, V.V.R. designed and supervised the studies, developed Gnostic algorithms, carried out machine learning experiments, H.I.Z. designed and supervised the studies and performed intracellular accumulation studies and antibacterial analyses.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-12376-1>.

Correspondence and requests for materials should be addressed to V.V.R. or H.I.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022, corrected publication 2023