Supplementary Information

Mechanism of Action-Based Classification of Antibiotics using High-Content Bacterial Image Analysis.

Kelly C. Peach,† Walter M. Bray,¥ Dustin Winslow,‡ Peter F. Linington,§ Roger G. Linington†,*

†Department of Chemistry and Biochemistry, University of California Santa Cruz, Santa Cruz, California, USA.

¥Chemical Screening Center, University of California Santa Cruz, Santa Cruz, California, USA.

‡Department of Earth and Planetary Sciences, University of California Santa Cruz, Santa Cruz, California, USA.

§School of Computing, University of Kent, Canterbury, Kent, CT2 7NF, UK.

*To whom correspondence should be addressed. Tel.: (831) 459-3014. Fax: (831) 459-2935. E-mail: rliningt@ucsc.edu
Table of Contents

SUPPLEMENTARY RESULTS

Figure S1. $^1$H NMR spectrum (600 MHz) and structure of cosmomycin D isolated from 1498D and 1498E.

Figure S2. $^1$H NMR spectrum (600 MHz) and structure of novobiocin isolated from 1565D. Related to Figure 3 in the text.

Figure S3. $^1$H NMR spectrum (600 MHz) and structure of pentachloropseudulalin isolated from 1726D.

Figure S4. $^1$H NMR spectrum (600 MHz) and structure of cycloprodigiosin isolated from 2001E.

Figure S5. EIC traces for [M+H]$^+$ adduct of tetracycline in the S. aureofaciens C fraction.

Table S1. Summary of $R_G$ values calculated from model misfits for training set, prefractions, and pure compounds.

SUPPLEMENTARY METHODS

Cultivation of Marine Bacteria ................................................................. S8
Liquid Culture, Extraction, and Prefractionation Protocol .................................. S8
Peak Library Preparation and Bioactivity Screening Protocol ............................ S9
Purification and isolation of cosmomycin D ...................................................... S9
Purification and isolation of cycloprodigiosin ............................................... S9
Purification and isolation of pentachloropseudulalin ...................................... S10
Purification and isolation of novobiocin ....................................................... S10
Definitions of Image Analysis Metrics .......................................................... S11
Figure S1. $^1$H NMR spectrum (600 MHz) and structure of cosmomycin D isolated from 1498D and 1498E.
Figure S2. $^1$H NMR spectrum (600 MHz) and structure of novobiocin isolated from 1565D.

novobiocin
in DMSO-$d_6$ (600 MHz)
Figure S3. $^1$H NMR spectrum (600 MHz) and structure of pentachloropseudilin isolated from 1726D.
Figure S4. $^1$H NMR spectrum (600 MHz) and structure of cycloprodigiosin isolated from 2001E.
The prefraction sample was analyzed using RP-HPLC (Phenomenex Synergi Fusion-RP, 250 x 4.6 mm, 10 µm) using a solvent system of methanol and H₂O with 0.02% formic acid. The sample was prepared as a 1.0% formic acid solution, and run on a gradient from 30%-100% methanol at 1.0 mL min⁻¹ for 40 min. These conditions yield an EIC showing two characteristic tetracycline peaks, representing the keto and enol -forms.¹
Table S1. Summary of $F_G$ values calculated from model misfits for training set, prefractions, and pure compounds.

### SUPPLEMENTARY METHODS

**Cultivation of Bacteria**

Marine sediment samples were transferred onto media plates using our standard protocol. Six different solid agar media were used for microbial isolation: actinomycete isolation agar (AIS and AIF by Difco),\(^2\) NTS, NTF, HVF, and HVS.\(^3\) All isolation plates were prepared with sterile seawater and supplemented with 50 mg/L of both cyclohexamide and nalidixic acid. Sediments were serially stamped onto solid agar with a sterile swab. Cultures were incubated at room temperature and bacterial colonies displaying desired morphologies were subcultured on Difco Marine Broth solid agar plates until pure. Typical incubation times for the appearance of colonies from isolation plates ranged from 10–90 days.

**Liquid Culture, Extraction, and Prefractionation Protocol**

Selected colonies were inoculated into 10 mL of modified saline SYP (mSYP) media (10 g starch, 4 g peptone, 2 g yeast extract and 31.2 g instant ocean in 1 L of distilled water) and shaken for 3 days before being stepped up to large-scale in stages, firstly by adding 1.5 mL of the 10 mL cell cultures (small-scale) into 50 mL of mSYP (medium-scale), followed by inoculation of 15 mL of these medium-scale cell cultures into 1 L of the same culture medium (large-scale) with 10 days at each interval. All cultures were incubated at 26°C and shaken at 200 rpm.

Large-scale cultures were fermented in 2.8 L Fernbach flasks containing a stainless steel spring for 10 days prior to chemical extraction. 20 g of pre-washed Amberlite XAD-16 resin ($\text{CH}_2\text{Cl}_2$ MeOH and water) was added to each large-scale culture, shaken for 2 hrs (200
rpm), and the resulting slurry filtered under vacuum through a glass microfiber filter (Whatman). The cells, resins and filter paper were extracted with 1:1 CH$_2$Cl$_2$/MeOH (250 mL) and the suspension shaken at 200 rpm for 1 hr. Organic extracts were filtered and concentrated to dryness in vacuo. Dried crude extracts were pre-fractionated by solid phase extraction chromatography (5 g C$_{18}$ cartridge, Supelco, USA) using a stepwise MeOH/H$_2$O gradient: 40 mL of 10%, 20% (fraction A), 40% (fraction B), 60% (fraction C), 80% (fraction D), 100% MeOH (fraction E) then 100% EtOAc (fraction F). Fractions A – F were concentrated to dryness in vacuo, then resuspended in DMSO (1 mL) and aliquots of these DMSO stock solutions reformatted to 384-well plates prior to screening. Active hits were serially diluted (two-fold dilutions) and re-screened to analyze for concentration-dependent phenotype trajectories.

**Peak Library Preparation and Bioactivity Screening Protocol**
A gradient was prepared with an average 50% increase in MeOH over 80 minutes. The eluent from each minute was collected in a single well of a 96-well plate (2 mL per well). Chromatography data and mass spectrometry data were simultaneously acquired. The collected eluent plates were concentrated to dryness in vacuo, and resuspended in 10 μL of DMSO. 96-well plates were reformatted into 384-well format, and rescreened. Only optical density data was collected after incubation (no image data required). Optical density output was used to identify active wells, and this was compared to the corresponding peak from the HPLC trace in order to identify active constituents for isolation.

**Purification and identification of cosmomycin D**
Prefraction 1498D was analyzed by LCMS and found to contain a complex mixture of anthracyclines as determined by diagnostic UV absorbances. This prefraction was prepared by standard extraction and solid phase separation protocols. Cosmomycin D was isolated using RP-HPLC (Phenomenex Jupiter C$_{18}$, 250 x 4.6 mm, 5 μm) using a solvent system of acetonitrile and H$_2$O with a 20 mM phosphate buffer (pH = 7.0). The prefraction was run on a gradient from 20%-60% acetonitrile at 2 mL min$^{-1}$ for 40 min, and washed with H$_2$O for 10 min to afford cosmomycin D as an amorphous purple solid.

Cosmomycin D was analyzed by (+)-HRESITOFMS (obsd [M + 2H]$_2^+$ at m/z 595.2996) to give an exact mass of 1188.5834 (calcd. 1188.5829, Δppm= 0.42). $^1$H and gCOSY NMR spectra were obtained to verify the structure. Cosmomycin D has a 4,6,11-trihydroxy substitution pattern, resulting in three coupled aromatic protons on the A ring from 7.32-7.93 ppm. Analysis of the gCOSY spectrum confirmed their connectivity (Supplementary Fig. S1 inset), and differentiated cosmomycin D from two other 1,4,6,11-tetrahydroxy anthracyclines that share the same mass, but lack these three coupled aromatic protons. The structure was further confirmed by comparison of the $^1$H NMR data to published data for cosmomycin D.

**Purification and identification of cycloprodigiosin**
Prefraction 2001E was analyzed by LCMS and found to contain a highly pigmented compound as determined by its diagnostic UV-visible absorbance at 536 nm. Subsequent HPLC purification isolated cycloprodigiosin from prefraction 2001E. This prefraction was prepared by standard extraction and solid phase separation protocols. Cycloprodigiosin was isolated using RP-HPLC (Phenomenex Synergi Fusion-RP, 250 x 4.6 mm, 10 μm) using a solvent system of methanol and H$_2$O with 0.02% formic acid. The prefraction was run on a gradient from 52%-65% methanol at 2 mL min$^{-1}$ for 30 min. A second purification was
required using RP-HPLC (Phenomenex Kinetex XB-C_{18} 100Å, 100 x 4.6 mm, 2.6 μm) using a solvent system of methanol and H_{2}O with 0.02% formic acid. The sample was prepared for purification as a 1.0% formic acid solution, and run on a gradient from 50%-80% methanol at 0.8 mL min^{-1} over 10 min to yield pure cycloprodigiosin as an amorphous bright pink solid.

Cycloprodigiosin was analyzed by (+)-HRESITOFMS (obsd. [M+H]+ at m/z 322.1928) to give an exact mass of 321.1856 (calcd. 321.1841, Δppm= 4.67). One-dimensional ¹H NMR was obtained to verify the structure. The structure was further confirmed by comparison of the ¹H NMR data to published data for cycloprodigiosin.⁶,⁷

**Purification and identification of pentachloropseudilin**

Pentachloropseudilin was isolated from prefracion 1726D. This prefracion was prepared by standard extraction and solid phase separation protocol. Pentachloropseudilin was isolated using RP-HPLC (Phenomenex Synergi Fusion-RP, 250 x 4.6 mm, 10 μm) applying an isocratic separation (67% MeOH, 33% H_{2}O with 0.02% formic acid, 2 mL min^{-1}, t_R= 28.2 min.) to afford the pure compound as an off-white solvent.

Analysis by (-)-HRESITOFMS (obsd. [M - H]^- at m/z 327.8663) to give an exact mass of 328.8645 (calcd. 328.8736, Δppm = 2.74; m/z (%) = 329.8643 (100.0%), 331.8615 (61.6%), 327.8673 (58.4%), 333.8584 (17.3%), 330.8676 (9.4%), 328.8704 (5.8%), 332.8643 (5.6%), 335.8559 (2.7%), 334.8616 (1.8%)), with an isotope pattern indicative of a pentachlorinated compound. This mass data, coupled with the assignment and comparison of the ¹H NMR data to the established literature values, established the structure as pentachloropseudilin.⁸

**Purification and identification of novobiocin**

Novobiocin was found to be present in both prefracion 1565D and 1565E (sequential prefractions from the same crude extract), but was isolated and characterized from the 1565D prefracion. This prefracion was prepared by standard extraction and solid phase separation protocol. Novobiocin was isolated using RP-HPLC (Phenomenex Synergi Fusion-RP, 250 x 4.6 mm, 10 μm) applying an isocratic separation (66% MeOH, 34% H_{2}O with 0.02% formic acid, 2 mL min^{-1}, t_R= 33.5 min.) to afford the pure compound as an off-white solid.

Analysis by (+)-HRESITOFMS (obsd. [M+H]+ at m/z 613.2407) gave an exact mass of 612.2335 (calcd. 612.2319, Δppm= 2.61). One-dimensional ¹H NMR was obtained to verify the structure, which was further confirmed by comparison of the ¹H NMR data to published data for novobiocin.⁹
Definitions of Image Analysis Metrics

**image mean intensity**
Average intensity in complete image, before black correction. This is the sum of all pixel intensities over the total number of pixels.

**feature halfwidth**
Using radial projection of autocorrelation function, find the radius for which the central correlation peak drops to 50% of the initial value. This value is used as a starting tile size for black correction. Rules are applied to keep this value within bounds, no matter what the image content.

**top threshold**
Intensity value at the knee between the top and bottom feature regions. All features with sum of intensity values (also referred to as their weight) above this point are in the top group.

**top threshold tolerance**
This is an indication of how robust the threshold determination is. If the two linear regions are considered to form two sides of a triangle, the tolerance is the perpendicular distance to the third side divided by the square of the side opposite the knee. The value will be larger as the regions are more populated and the gradients more different.

**peak size**
Number of features in the top (bright) feature set.

**peak interior count**
Total number of interior pixels in the peak (top) features divided by the total number of pixels in the image; i.e. the fraction of the image classified as peak interior (yellow).

**peak boundary count**
Total number of boundary pixels in the peak (top) features divided by the total number of pixels in the image; i.e. the fraction of the image classified as peak boundary (red).

**film size**
Number of features in the film (or bottom, dim) feature set.

**film interior count**
Total number of interior pixels in the film (or bottom, dim) features divided by the total number of pixels in the image; i.e. the fraction of the image classified as film interior (cyan).
**film boundary count**
Total number of boundary pixels in the film (or bottom, dim) features divided by the total number of pixels in the image; i.e. the fraction of the image classified as film boundary (dark blue).

**back size**
Number of features in the background feature set.

**back interior count**
Total number of interior pixels in the background features divided by the total number of pixels in the image; i.e. the fraction of the image classified as background interior (dark purple).

**back boundary count**
Total number of boundary pixels in the background features divided by the total number of pixels in the image; i.e. the fraction of the image classified as background boundary (even darker purple).

**clustered fraction**
This is just the fraction of the image in all the top features, either boundary or interior.

**top mean depth**
This is the unweighted average over all the top features of their individual depths; the depth of a feature is the sum of its pixel intensities divided by the number of pixels.

**top mean dispersion**
This measure is larger if most of the intensity is near the boundary, and smaller if it is concentrated near the middle. A relative intensity is used, which is the actual intensity minus the smallest intensity in the feature. The sum of the relative intensity times the square of the radius (distance from the pixel to the center of gravity). This is divided by the sum of the relative intensities, and the square root taken (analogous to the radius of gyration in mechanics). Finally it is normalized with respect to feature size by dividing by the mean radius.

**top mean fanout**
This counts the number of features (of whatever type) that touch this feature. Each touching feature is counted only once, even if it touches distinct parts of the perimeter. The final metric is the unweighted average of the individual values over all the top features.

**top mean fence**
This counts the number of pixels belonging to other features that touch the boundary of a particular feature. The calculation is the sum of foreign neighbors for all the boundary
pixels, so that if a single foreign pixel touches several boundary pixels of this object, it will be counted several times. The final metric is the unweighted average of the individual values over all the top features.

**top mean eccentricity**

For each feature, this calculates the number of boundary pixels divide by the square root of the total number (boundary plus interior) of pixels. A circular feature would have the lowest value. The final metric is the unweighted average of the individual values over all the top features.

**top mean rough**

This is a measure of how irregular the intensity of the feature is. If it has a single peak and decreases smoothly out from it, the value will be low, and if it has many internal peaks, the value will be higher. It is calculated during feature extraction. Whenever two touching features are merged, the working value for the result is the sum of the working values of the two source features plus the difference in intensity between the lower of the two features and the saddle point where they have been found to touch. The final value is scaled by dividing by the total number of pixels in the feature. For background, all these inequalities are reversed. The final metric is the unweighted average of the individual values over all the top features.

**top mean environment correlation**

This metric used the angular distribution calculated for each feature as a starting point. It calculates a background distribution for each feature by scanning a square window centered on the center of gravity and considering all points outside the feature in constructing an angular distribution of intensities. It then correlates these two distributions. To do this it first calculates scaling factors by finding the maximum and minimum values in each distribution and offsetting values so that (max+min)/2 becomes zero and max = 1, min = -1. The correlation is the average value over all directions of featureDistribution(direction)*backgroundDistribution(direction). Note that the limited window (currently +/-50 pixels) means that the correlation for large features is zero because no background is seen. The final metric is the unweighted average of the individual values over all the top features.

**top mean radial spread**

The radial distribution of a feature is formed by examining all the pixels (interior and boundary) of a feature that fall within a sampling window centered on its center of gravity. Pixels outside this region are ignored. The density function is built from the distance from the center of gravity and the intensity values offset from the feature minimum. The quartile points of the cumulative distribution are found and the feature metric is calculated from the 75% radius divided by the 25% radius. The final metric is the unweighted average of the individual values over all the top features.
**top mean angular peaks**
The angular distribution of a feature is formed by examining all the pixels (interior and boundary) of a feature that fall within a sampling window centered on its center of gravity. Pixels outside this region are ignored. The density function is built from the angle (from north) of the pixel relative to the center of gravity and the intensity values offset from the feature minimum. This is smoothed to retain only significant peaks. The number of cycles is used as the feature metric. The final metric is the unweighted average of the individual values over all the top features.

**top mean angular variation**
The smoothed angular distribution is calculated as in the top mean angular peaks metric. The feature metric is the lowest trough value divided by the highest peak. The final metric is the unweighted average of the individual values over all the top features.

**top mean skeleton fraction**
The feature metric is the number of skeleton pixels divided by the total number of pixels. The final metric is the unweighted average of the individual values over all the top features.

**top mean marrow fraction**
The feature metric is the number of marrow pixels divided by the total number of pixels. The final metric is the unweighted average of the individual values over all the top features.

**bottom mean depth**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**bottom mean dispersion**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**bottom mean fanout**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**bottom mean fence**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**bottom mean eccentricity**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.
**bottom mean rough**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**bottom mean environment correlation**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**bottom mean radial spread**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**bottom mean angular peaks**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**bottom mean angular variation**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**bottom mean skeleton fraction**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**bottom mean marrow fraction**
The metric is calculated in the same way as the equivalent top metric, but using the bottom feature set.

**film correlation plus fraction**
This is one of two metrics introduced primarily to detect out of focus images. The film correlation is a non-standard autocorrelation calculated by summing the contribution of all possible interior pixel pairs within each separate feature (including duplicates). For each pair, the contribution to the sum is the product of the intensities offset by that feature’s depth. The metric is the fraction of pixels in the autocorrelation map, which are positive and above 10% of the maximum value.

**film correlation minus fraction**
Using the same autocorrelation map as the plus fraction, this metric is the fraction of pixels in the autocorrelation map which are negative and below 10% of the minimum value.
**mean dark fanout**

The metric is calculated in the same way as the equivalent top metric, but using the background feature set. Signs and inequalities are reversed where appropriate.

**mean dark isolation**

This is the reciprocal of the mean dark fanout.
References


