HALO384: A Halo-Based Potency Prediction Algorithm for High-Throughput Detection of Antimicrobial Agents

Marcos H. Woehrmann, Nadine C. Gassner, Walter M. Bray, Joshua M. Stuart and Scott Lokey

J Biomol Screen 2010 15: 196 originally published online 19 January 2010
DOI: 10.1177/1087057109355060

HALO384: A Halo-Based Potency Prediction Algorithm for High-Throughput Detection of Antimicrobial Agents

MARCOS H. WOEHRMANN,[1] NADINE C. GASSNER,[2,3] WALTER M. BRAY,[2,3]
JOSHUA M. STUART,[1] and SCOTT LOKEY[2,3]

A high-throughput (HT) agar-based halo assay is described, which allows for rapid screening of chemical libraries for bioactivity in microorganisms such as yeast and bacteria. A pattern recognition algorithm was developed to identify halo-like shapes in plate reader optical density (OD) measurements. The authors find that the total growth inhibition within a detected halo provides an accurate estimate of a compound's potency measured in terms of its EC_{50}. The new halo recognition method performs significantly better than an earlier method based on single-point OD readings. An assay based on the halo algorithm was used to screen a 21,120-member library of drug-like compounds in *Saccharomyces cerevisiae*, leading to the identification of novel bioactive scaffolds containing derivatives of varying potencies. The authors also show that the HT halo assay can be performed with the pathogenic bacterium *Vibrio cholerae* and that liquid culture EC_{50} values and halo scores show a good correlation in this organism. These results suggest that the HT halo assay provides a rapid and inexpensive way to screen for bioactivity in multiple microorganisms. (*Journal of Biomolecular Screening* 2010:196-205)

Key words: halo assay, high throughput, optical density, chemical library, agar

INTRODUCTION

THE CLASSIC DISK DIFFUSION, OR HALO, ASSAY is commonly used to evaluate the antimicrobial activity of small molecules and natural product extracts. An agar-plated lawn of microorganism is exposed to a filter disk soaked in a test solution. Lethal or growth-inhibitory compounds cause visible halos, representing a concentration-dependent decrease in growth surrounding the disk. This assay has the advantage of simplicity, and halos provide unmistakable visual confirmation of bioactivity. In addition, because halo size is correlated with potency, the halo assay can be used as a simple and effective way to compare activities among groups of compounds. To accurately determine the inhibition of a compound, we calculate a halo score by measuring the optical density at multiple points across the diameter of the halo and integrating across the area of inhibition. This gives a much more reliable indication of a compound's effect than using either the optical density at the center or the diameter of the zone of death. Indeed, we show that the halo score is accurate enough to estimate a compound's EC_{50}. We recently developed a high-throughput version of the classic halo assay, in which compounds are delivered robotically to agar plates seeded with a microorganism using a 384-pin tool.[1] Here we describe a computational algorithm to score and quantify potency. We used the algorithm to screen 21,120 compounds in the yeast *Saccharomyces cerevisiae* and identified 590 bioactive compounds from 30 structural classes. The method generalizes across species; we show EC_{50} can also be predicted in a pathogenic bacterium *Vibrio cholerae*.

[1]Department of Biomolecular Engineering and [2]Department of Chemistry and Biochemistry, UC Santa Cruz, Santa Cruz, California.
[3]UCSC Chemical Screening Center, Santa Cruz, California.

Received Aug 17, 2009, and in revised form Oct 2, 2009. Accepted for publication Oct 6, 2009.

Journal of Biomolecular Screening 15(2); 2010
DOI: 10.1177/1087057109355060

MATERIALS AND METHODS

Quantification of drug toxicity from soft-agar pinning using a "halo score"

In the high-throughput (HT) halo assay described here, trays are filled with agar seeded with microorganism, and compound stock solutions are deposited from library plates into the agar using a robotically driven 384-pin array (**Fig. 1A**). Active (i.e., lethal or growth-inhibitory) compounds generate halos, or zones of growth inhibition, in which the effect decays as a function of the distance from deposition.

OD measurements are then taken with a plate reader, scanning 9 points around each well (4 on each side and 1 centered on the site of compound transfer; **Fig. 1B**). When viewed in cross section, a halo gives a characteristic bowl-shape pattern of optical density (OD) that reaches a minimum at the site of compound addition. To aid visual inspection of the results, we produce an "OD line plot" to summarize all of the readings on

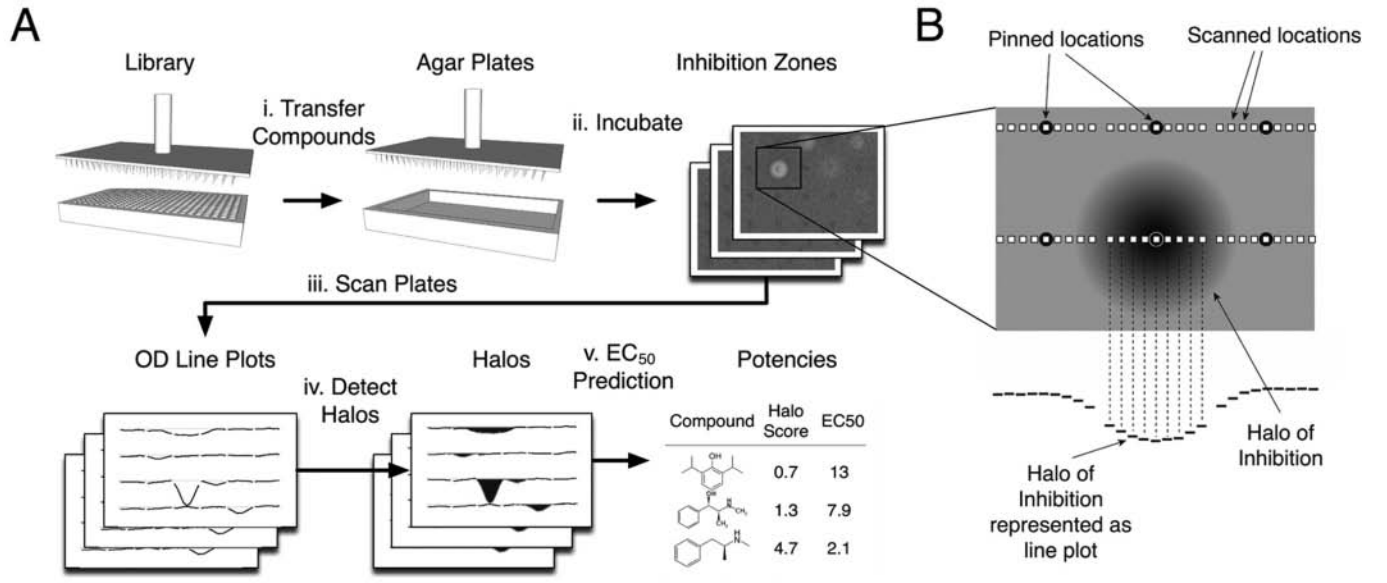


FIG. 1. (**A**) High-throughput pipeline for drug potency prediction. A library of compounds is transferred from standard 384-well plates into agar using a pinning robot (step i). Following incubation at room temperature (step ii), optical density measurements are scanned from the resulting plates, producing optical density (OD) line plots (step iii) from which halos are detected and quantified (step iv). EC_{50} potencies are predicted for all compounds producing halos and recorded (step v). (**B**) Overview of agar-based pinning, reading, and halo detection strategy. Compounds are pinned into soft agar in a grid pattern (small black circles). A plate reader records 9 optical density readings across each pinning location (white squares) that can be viewed in cross section to visualize the pattern of growth inhibition as a function of the distance from pinning (OD line plot; bottom). Toxic compounds show a zone of clearing centered on the location where a compound has been pinned into soft agar (large shaded circle).

a plate in cross section. A raw "halo score" is then calculated for any detected halo-like patterns in the agar by estimating the volume of growth inhibition originating from a single compound (described below). Compound potencies, in the form of EC_{50}s, are then estimated and recorded.

Strains and chemicals

Yeast strain BY4741 was obtained from Open Biosystems (Huntsville, AL). Smooth FY_Vc_1, *V. cholerae* O1 El Tor A1552 was a gift from Fitnat Yildiz.[2] Growth media reagents were purchased from Sigma (St. Louis, MO). OmniTrays were purchased from Fisher (12565450; Fisher Scientific, Hampton, NH). Library compounds were obtained from the National Cancer Institute's Developmental Therapeutics Program and ChemDiv, Inc. (San Diego, CA).[3]

High throughput yeast and cholera halo assay

Media were prepared as previously described.[1] The warm media were inoculated with overnight culture diluted to give a final concentration of $A_{600} = 6 \times 10^{-2}$ and poured into an OmniTray. The tray was set on a flat surface to cool for 15 min and dried in a biological safety cabinet for 15 min.

Compounds were pin-transferred from DMSO stocks plated in 384-well polypropylene trays (Fisher AB1056) into the cooled agar with a pin-tool robot (JANUSMPD; PerkinElmer, Waltham, MA) using notched pins that deliver 200 nL (±8%) each (VP 384FP3S100; V&P Scientific, San Diego, CA). Before and between applications, pins were cleaned by submersion in 70% ethanol (3×), 50% DMSO sonication bath (3×), and finally a 95% ethanol circulating 384-channel bath (3×). Between each wash step, the pins were applied to blotting paper (V&P Scientific VP540D-100) to absorb excess solvent. At the end of the cleaning cycle, pins were dried in an air drier manifold. The soft agar plate was incubated at 24°C for yeast and 37°C for cholera for 14 h, and then the A600 was read in an EnVision plate reader (PerkinElmer 2104-0010). Each of the 384-pin array points was scanned in a 9-by-1 horizontal line. The data were saved in CSV file format for use as input for the halo detection software.

Data preprocessing

The density of cells in agar, the amount of agar, and other effects can vary across a plate. To mitigate the influence of these local fluctuations of cell density, we normalized the optical density readings by subtracting out 3 main location effects due to the solid agar assay. These effects included the orientation of a reading relative to the site of pinning, the tilt of a plate that may cause systematic differences in cell density across a plate, and whether a reading was taken near a plate edge where cell density can increase because of adherence of the media onto the plastic. These final normalized quantities were then used to detect the presence of bioactive compounds on a plate.

Halo detection and quantification

Using the normalized optical density readings, the presence of bioactive compounds producing characteristic halos of inhibited cell growth was detected and quantified. The halo results from the diffusion of compounds into the agar, which can be used to identify compounds over a wide range of toxicity. Compounds found to produce a zone of inhibition consistent with the shape of a halo were subsequently quantified using a score that reflects the overall amount of inhibition produced.

Intuitively, the algorithm detects hits in agar in an analogous fashion to the way humans identify hits—instead of only inspecting the level of microorganism at the point where the compound was pinned, it searches for circles of reduced growth in a neighborhood around the site. Potent compounds can inhibit cell growth spanning multiple wells in the plate. Therefore, the algorithm first detects multiwell halos, flags any wells that are included in any of these large halos, and then searches for single-well halos within the remaining wells. Because it is unlikely that a circular pattern would be produced by chance and because integrating multiple readings can mitigate the noise present in any single reading, this shape-based approach has the potential to be much more accurate than a single-reading-based approach.

Once a halo is detected, the amount of total inhibition is quantified for wells detected to be centered on the halo. A *raw halo score* is computed for all multiple- and single-well halos. The raw halo score sums up all of the readings to the left, right, and center of a detected halo. The final *halo score* is computed by dividing the raw halo score by the concentration of the compound that was pinned onto the plate.

Growth inhibition measurement in liquid culture

Yeast (wild-type haploid strain BY4741, $A_{600} = 6 \times 10^{-2}$) or cholera (smooth FY_Vc_1, *V. cholerae* O1 El Tor A1552,[2] $A_{600} = 1 \times 10^{-4}$) were incubated with 6 two-fold dilutions of each compound in 200-µL cultures in 96-well plates, in addition to DMSO controls. ODs were read every 45 min using an EnVision plate reader, and the plate was agitated just prior to reading to suspend the cells. Yeast or bacteria doubling times at each concentration were calculated and compared to the doubling time in DMSO.

Determination of EC_{50} from liquid culture

To estimate EC_{50} from liquid culture, we fit a dose-response curve to the liquid culture optical density readings for a single compound using the GraphPad software (GraphPad, San Diego, CA). We then used the concentration (independent coordinate) at which the fitted curve passed through the midpoint of the optical density readings (dependent coordinate) as the most likely EC_{50}. For compounds with steep Hill slopes, as was the case for many of the NSC compounds, GraphPad either produced no confidence interval or output an excessively wide range. Therefore, rather than use a confidence interval estimated by GraphPad, we determined a range of possibility (ROP) for the EC_{50}, consisting of the minimum and maximum concentrations that must bracket the most likely EC_{50} value based on the data. First, we identified the 2 points that straddled the 50% inhibition point lying nearest to the fitted sigmoid. Then, the concentrations of these 2 points were used as the minimum and maximum of the ROP. For each compound, we report the ROP along with the most likely EC_{50} estimate.

Molecular similarity analysis

The molecular similarity component of Pipeline Pilot (Accelrys Software, San Diego, CA) was used to calculate the similarity between molecules in the ChemDiv collection using SciTegic's molecular fingerprints (FCFP = 4). Similarity was calculated using Tanimoto coefficients to generate up to 2000 compound clusters with an average of 50 members.

RESULTS

The halo algorithm detects a broad range of compound toxicity

To determine the halo algorithm's utility for predicting a compound's potency, we evaluated the correlation between the raw halo score and the stock solution concentration for a series of known drugs that span a wide range of potencies: rapamycin, disulfram, and ciclopiroxolamine (EC_{50}s: 14 nM, 94 µM, and 39 µM, respectively; **Fig. 2**). We used a constrained linear regression in which fitted lines were forced to pass through the origin so that compound concentrations of zero were matched with halo scores equal to zero. Raw halo scores and compound concentrations were strongly correlated for rapamycin ($R^2 = 0.93$, $p < 1.2 \times 10^{-4}$) and disulfram ($R^2 = 0.79$, $p < 0.0175$) and not significantly correlated for ciclopiroxolamine ($R^2 = 0.10$, $p < 0.45$) where the raw halo scores were very small and influenced by excess noise. These results are conservative as the R^2 values are underestimated during the constrained linear regression.

The HT halo assay reported previously[1] used a single-point OD (SPOD) reading per compound, rather than the set of 9 readings used to calculate the raw halo score used here. To quantify whether the new halo score approach improves the detection range compared to the previous method, we plotted the SPOD readings against an increasing concentration of rapamycin and compared it to the results obtained for the raw halo score (**Fig. 3A**). For compounds that are less potent, the raw halo score and SPOD readings are both able to discriminate between halos of different diameter. However, above a critical concentration of pinned stock solution (for rapamycin around 15 µM), the SPOD readings flatten out while the raw halo scores continue to increase linearly. The current method takes advantage of the spatial pattern created by compound deposition, expanding the upper limit of potencies predicted for toxic compounds.

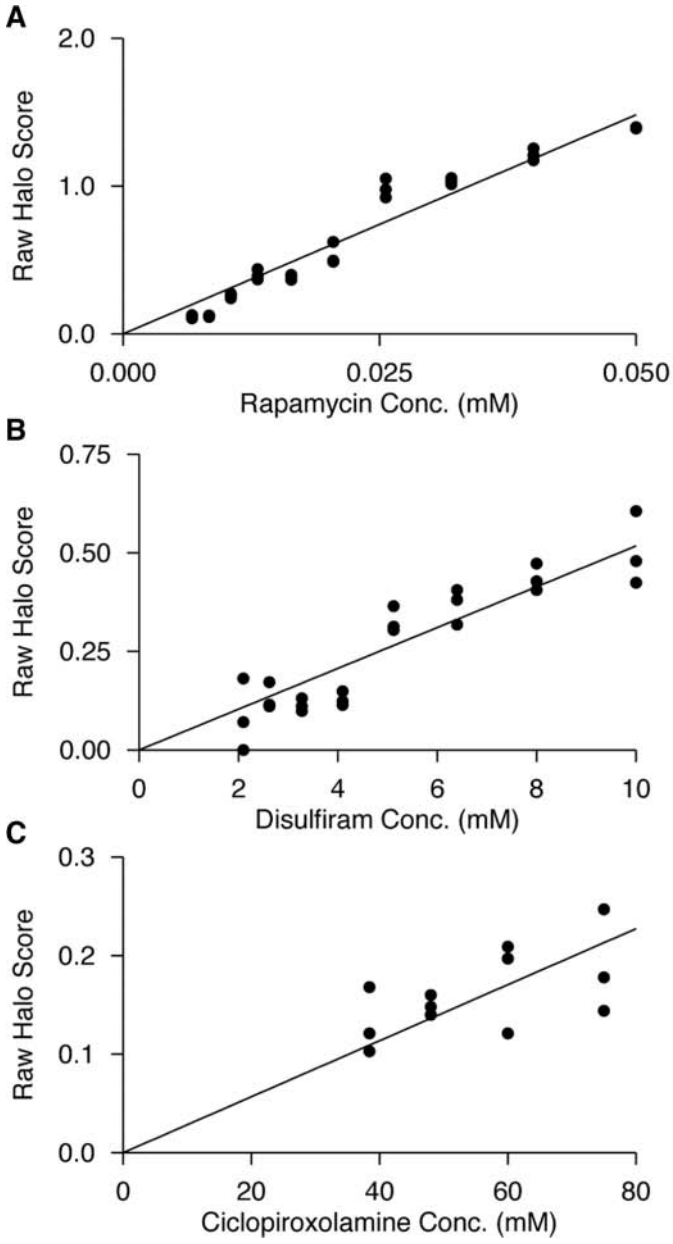


FIG. 2. Linear correlation of halo score with compound concentration. (**A-C**) Scatter plots of stock concentration (*x*-axis) against raw halo score (*y*-axis) for different compounds. Points represent compounds pinned into a different well on a plate. Plates were run in triplicate for each of 3 compounds, including (**A**) rapamycin, (**B**) disulfram, and (**C**) ciclopiroxolamine.

Because the raw halo score increases linearly with the pinned stock concentration of a compound, we calculated a final halo score (*H*) by dividing the raw halo score by the compound's stock concentration. The final halo score allows for direct comparison of measurements from different compounds or from the same compound run on different plates and, as we show next, can be used to predict a compound's EC_{50}.

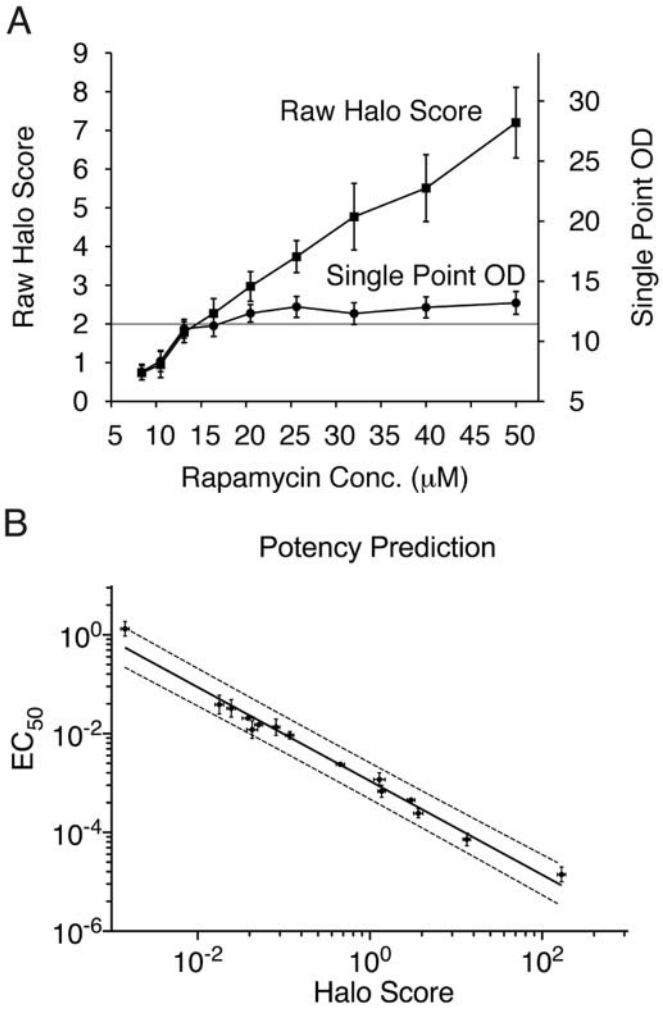
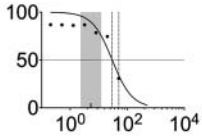
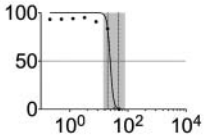
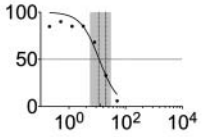
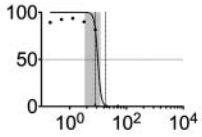
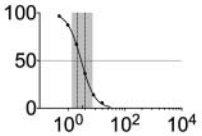
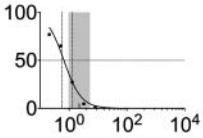


FIG. 3. (**A**) The dynamic range of the halo score is larger than single-point optical density (SPOD). Raw (prenormalized) halo scores (left *y*-axis) and a well's SPOD reading (*right* y-axis) plotted against the pinned concentration of rapamycin (*x*-axis). Both the halo score and SPOD increase linearly with rapamycin concentration, but the raw halo score is linear over a wider range than SPOD. Bars represent standard errors calculated from 3 replicates. (**B**) Halo scores predict EC_{50} in *Saccharomyces cerevisiae*. Log-log plot of EC_{50} measurements obtained from liquid culture (*y*-axis) plotted against the halo score obtained in agar (*x*-axis) for 19 chemicals of varying toxicity against *S. cerevisiae*. The solid line represents the least squares regression line; dashed lines show the upper and lower bounds of a 95% confidence interval. Linear regression with 95% prediction intervals was performed using GraphPad Prism v5.0b software. The 2 estimates for a compound have an R^2 of 0.98 in log-log space, computed over EC_{50} potencies ranging from 14.4 nM to 1.32 mM.

An intuitive and widely used measure of drug toxicity is the EC_{50} or the *effective concentration* that causes a reduction of 50% in cell population. In microorganisms, EC_{50}s are generally determined in liquid culture using a low-throughput measurement of

Table 1. Potency Prediction of Uncharacterized Compounds[a]

Compound	Liquid Culture		Agar	
	Dose Response Curve	EC_{50} (range) μM	Halo Profile (Raw Halo Score)	Pred. EC_{50} (range) μM
NSC 371777		29.6 (28.6-50.0)	 (1.92)	5.27 (2.27-12.3)
NSC 17383		24.6 (20.0-47.4)	(0.275)	33.40 (14.2-78.9)
NSC 638432		11.7 (11.2-20.0)	(0.745)	12.96 (5.55-30.3)
NSC 65238		9.79 (8.00-18.5)	 (1.25)	7.93 (3.41-18.4)
NSC 207895		2.88 (2.12-3.91)	 (3.14)	3.31 (1.42-7.67)
NSC 301460		0.66 (0.57-1.30)	 (4.73)	2.24 (0.96-5.20)

[a]Six compounds from the National Cancer Institute (NCI) mechanistic, diversity, and natural product (MDNP) libraries were compared with the EC_{50} values predicted by the halo score method. The columns show the compound identity (column 1), the EC_{50} determined by liquid culture (column 2; gray demarks the 95% confidence interval estimated by GraphPad), the EC_{50} value (column 3), the optical density (OD) line plot (column 4), the predicted EC_{50} value, and associated range of possibility (column 5; see Materials and Methods).

growth rate as a function of compound concentration across a dilution series. We investigated whether the halo score, generated from a single concentration of pinned stock solution, could be used directly to predict EC_{50} values. If successful, this would enable drug potency characterization in a high-throughput setting. To do this, we determined EC_{50}s in liquid culture for 19 chemicals of varying toxicity (from 1000 μM to 0.1 μM) in *S. cerevisiae*. We asked whether H could predict EC_{50} determined from liquid culture. We first plotted the EC_{50}s against the halo scores in log-log space (**Fig. 3B**). Regression analysis revealed that the logarithm of the EC_{50} was linearly correlated with log H ($R^2 = 0.98$; $p < 1.4 \times 10^{-4}$). These results suggest that an EC_{50} estimate (E) can be calculated from a compound's halo score using the equation $E = \alpha H\beta$, where α and β were estimated from the intercept and slope of the linear regression. In particular, for *S. cerevisiae*, $\alpha = 10^{-3.0}$ and $\beta = -1.0$ so that $E = 10^{-3}H^{-1}$. Using a sliding window across the log H values, the standard deviation of the log EC_{50} values was calculated from which 95% prediction intervals were derived.

We next evaluated the halo score's ability to predict the potencies of unknown compounds. We selected 6 test compounds from the 3081-member National Cancer Institute (NCI) mechanistic, diversity, and natural product (MDNP) libraries

that were known to have a range of potencies in *S. cerevisiae*. EC_{50}s were predicted for each compound using the EC_{50}s estimated from the regression in log-log space. The predicted EC_{50}s ranged from 0.61 μM for NSC-301460 up to 16.2 μM for NSC-371777. We determined EC_{50}s using standard methods in liquid culture and compared these values to the EC_{50}-based predictions (**Table 1**). The EC_{50}s determined in liquid culture showed good agreement with those predicted based on *H*. Five of the 6 compounds had EC_{50}s within the ROP calculated from the data (see Materials and Methods). One compound fell outside the range, NSC-371777. Repeated attempts to assess the EC_{50} in liquid confirmed the difference in the liquid- versus the agar-based EC_{50} estimation (data not shown). In this case, the disagreement could be due to the compound's differential effects on cells in liquid versus agar. Apart from NSC-371777, 4 of the remaining estimated EC_{50}s differed by no more than 50% from the EC_{50}s determined in liquid, and all 5 had ranges of possibilities that overlapped with the 95% confidence interval for EC_{50}s from the halo score. Thus, the agar-based halo score produces highly comparable EC_{50}s to the more laborious liquid-based approach in 80% of the cases and is applicable to a broad range of hit potencies.

Accuracy of halo scores for high-throughput screening

For the purpose of screening large chemical libraries, a trade-off exists between recall, the sensitivity to detect even moderately toxic compounds, and precision, the proportion of true positives among the detected hits. To measure the utility of the halo score in a screening setting, we compared its ability to detect hits to the previously deployed SPOD method.

We conducted 2 tests to measure the accuracy of the halo and SPOD methods. First, a "bioactives" test was performed in which true positives were defined as those wells pinned with one of the potent compounds listed in **Figure 2**. Various concentrations of these compounds were used as true positives. To measure the sensitivity of the methods at low yet biologically relevant concentrations, we set the minimum pinned concentration to 20-fold higher than the EC_{50} determined in liquid culture. Second, a "screening" test was performed on the NCI MDNP library. In this test, a human expert using visual inspection aided by OD line plots defined the true positives halos. The true positives in this case were defined based on symmetry and alignment to the site of pinning.

For both tests, we plotted the precision as a function of recall by sweeping through a set of cutoff values for both the halo score and SPOD methods (**Fig. 4A**). In both the bioactives and screening tests, the halo score method had a higher precision than the SPOD method for recall levels in the range of practical application. For the purpose of library screening, high recall rates are desirable, even if a few false positives are allowed because these can be discarded in follow-up screens. Therefore, the precision at a recall of 90% and higher is of particular interest. At the 90% recall rate for both tests, the precision of the halo score method was significantly higher than SPOD. In the bioactives test, SPOD had a 40-fold higher false-positive rate than the halo method at the 100% recall rate. Thus, the halo score is predicted to drastically reduce the number of potential secondary screens should the primary screen be conducted at a desired maximum sensitivity. Upon inspection, the false positives called by the SPOD method at the 100% recall rate, which were correctly excluded by the halo score, were caused by edge effects or overlap with neighboring hits. Interestingly, in the screening test, the halo score method did not achieve 100% recall. This is due to the fact that many of the hits called by the human expert did not meet the shape criteria imposed by the halo score method. We checked all 204 of these expert-defined hits and found that 4 of them had no pinned compound. Thus, a recall of approximately 98% is optimal for the bioactive test, and the best recall achieved by the halo score is 94% after excluding the erroneous positive calls made by the human expert. The results of these 2 tests confirm the utility of the halo score as an automated method that is sensitive across a range of known activities and exceeds the accuracy of expert visual scoring.

To further assess the quality of the halo score–based high-throughput assay, we used Z factor analysis.[4] The Z factor measures the degree of separation in the reported scores between the positive and negative calls relative to standard deviations of the scores. Higher Z factors indicate that fewer false positives and false negatives can be expected. A general rule of thumb for commercial applications is that a screen has a minimum Z factor of 0.50, which corresponds to a separation between signal and background of 3 standard deviations.

To measure the halo score's applicability for use in an HT setting, we estimated Z factors for 2 drugs, rapamycin and disulfiram, pinned at stock concentrations. To obtain a conservative estimate of the assay's Z factor, we pinned stock concentrations reflecting potencies at the low end of detection (5-20 μM rapamycin, 2.5-5 mM disulfiram) at 11 sites on the plate. We repeated this analysis 3 times on 3 different plates and plotted the Z factors (**Fig. 4B**). The separation and corresponding Z factors were reproducible across plates. As expected, rapamycin at 20 μM, corresponding to the most potent positive control, obtained the highest Z factor of 0.68, well within the accepted range of a high-quality HT assay. On the other hand, the weakest controls had more borderline Z factors, and 2 were outside the suggested range of HT. Nonlinear regression revealed that a raw halo score of 1.2 corresponded to a Z factor 0.5 (**Fig. 4C**), corresponding to a raw halo score cutoff predicted to provide a robust measure of potency when performing a primary screen. However, because the Z factor analysis is overly conservative, we have found that screening with raw halo scores around 0.30 provides sufficiently accurate calls. For standard commercial libraries with stock concentrations of 10 mM, this halo score allows detection of compounds with EC_{50} values of 30 μM or lower, and we expect the assay to be applicable to compounds with much higher EC_{50} values (e.g., in the 200-μM range).

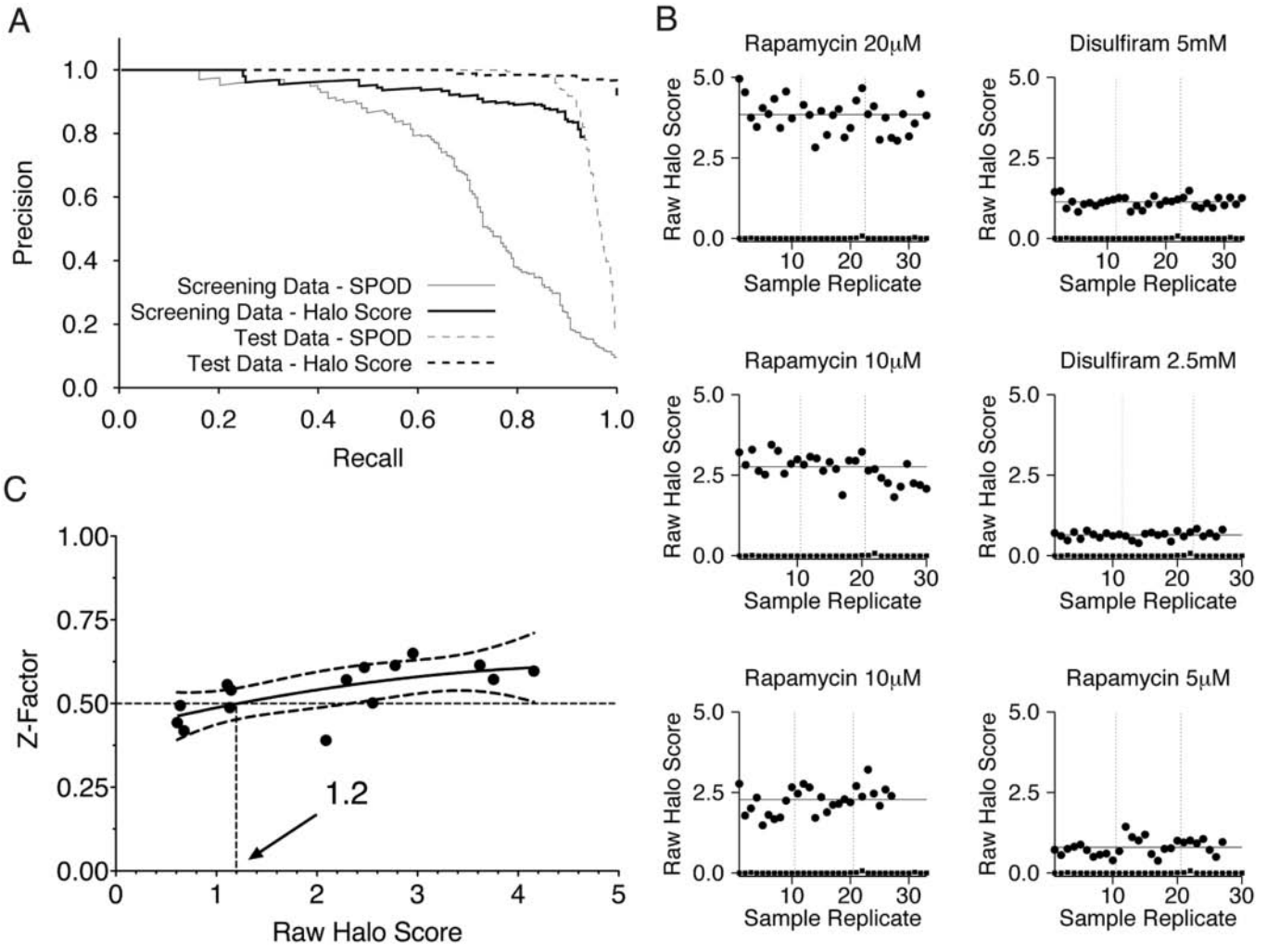


FIG. 4. (**A**) Precision recall plot comparing the halo score and single-point optical density (SPOD) methods. The performance of the halo score (black lines) and SPOD methods (gray lines) was measured in terms of precision (*y*-axis) and recall (*x*axis) in a small screen in which halo score calls were compared to a human expert (Screening Data; solid lines) and in a second setting in which calls were made for a set of compounds with known activity (Test Data; dashed lines). (**B**) Separation of positive controls from background. Raw halo scores of positive controls for 30 or 33 replicates (circles) were plotted alongside raw halo scores estimated for background wells containing no pinned compound (triangles). Two separate rounds were performed for rapamycin at 10 μM. (**C**) Assessment of halo score performance as a screening method. The Z factor for potency prediction (*y*-axis; see Materials and Methods) is plotted for several compounds with different halo scores.

Application of the pipeline to S. cerevisiae *and* V. cholerae

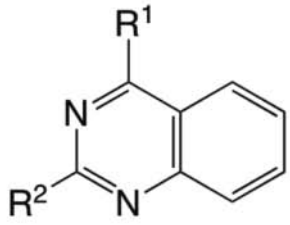
We used the algorithm to screen 21,120 compounds from a commercial library of drug-like compounds (ChemDiv, San Diego, CA).[3] A screen of this library against *S. cerevisiae* resulted in the identification of 590 bioactive compounds comprising 30 distinct structural scaffolds, out of 1056 scaffolds in the library. Activity and structural trends for a cluster of 2,4-diaminoquinazolines were determined (**Fig. 5**). EC_{50} values were predicted from the halo scores, and 95% confidence intervals were determined (**Table 2**). Repeated pinning of the same compound showed a standard deviation of 10% in EC_{50} prediction from day to day.

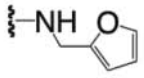
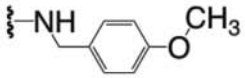
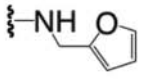
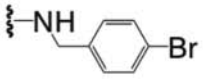
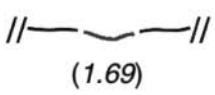
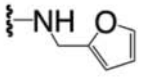
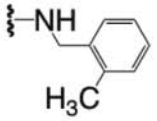
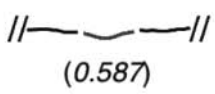
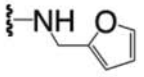
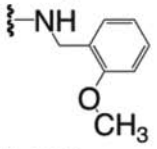
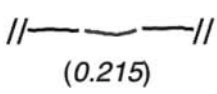
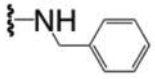
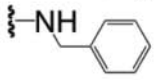
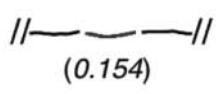
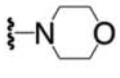
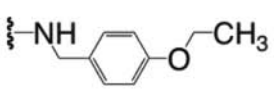
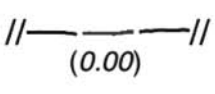
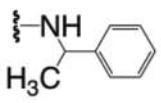
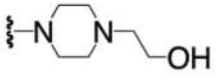
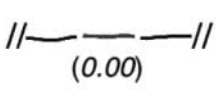
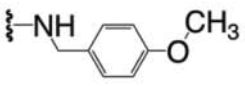
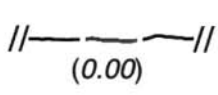
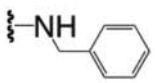
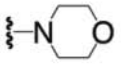
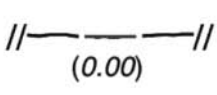
DISCUSSION

We have developed an automated method for identifying antimicrobial agents that is rapid, sensitive, and accurate. The key component of this algorithm is a "halo score" that uses multiple OD readings at different intervals from the site of pinning. In this way, the method makes use of the symmetric decrease in OD as a function of distance from its point of deposition. Correlations with liquid culture EC_{50} measurements allow for an estimation of potencies over a broad range.

This method significantly increases the number of hits that can be detected relative to visual or single-point OD

Table 2. Structure-Activity Relationships of 2,4-Diaminoquinazolines in *Saccharomyces cerevisiae*



Compound	R^1	R^2	Halo[a] (*raw halo score*)	EC_{50} (μM)
4408-0546			 (*2.51*)	4.09 (1.76 - 9.49)
4408-0539			 (*1.69*)	5.94 (2.56 - 13.8)
4408-0537			 (*0.587*)	16.26 (6.95 - 38.0)
4408-0549			 (*0.215*)	42.2 (17.8 - 100)
5940-0056			 (*0.154*)	57.94 (24.3 - 138)
4408-0144			 (*0.00*)	ND[b]
7765-0010			 (*0.00*)	ND[b]
C301-5029			 (*0.00*)	ND[b]
4408-0520			 (*0.00*)	ND[b]

[a] Each line in the Halo image represents a plot of the OD600 across the site of pinning.
[b] Compound showed no detectable activity.

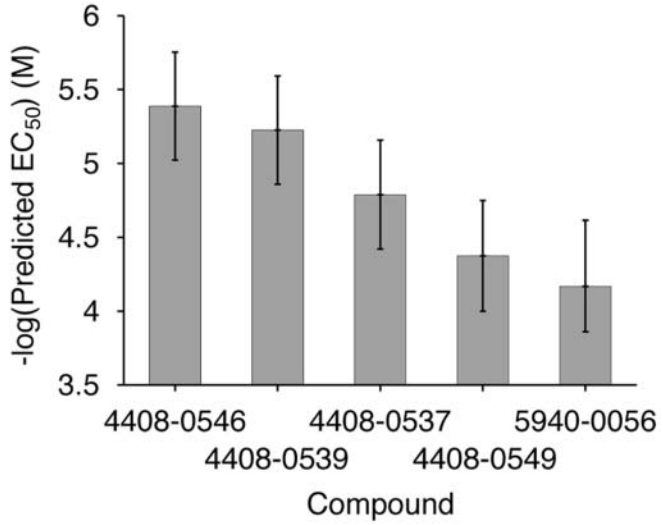


FIG. 5. Potency predictions for 2,4-diaminoquinazoline scaffold in *Saccharomyces cerevisiae*. The halo score–based EC_{50} predictions (*y*-axis) are shown for 5 modifications of the 2,4-diaminoquinazoline scaffold (ChemDiv compound code; *x*-axis). The uncertainty in the EC_{50} prediction is depicted by 95% prediction intervals for each compound (black error bars).

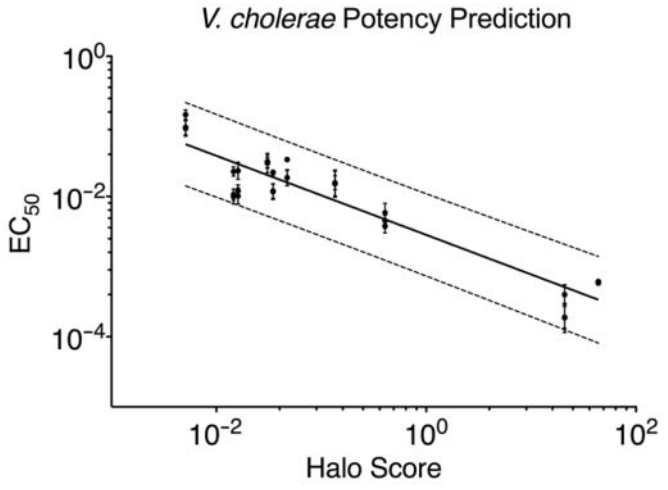


FIG. 6. Halo scores predict EC_{50} in *Vibrio cholerae*. Log-log plot of EC_{50} measurements obtained from liquid culture (*y*-axis) plotted against the halo score obtained in agar (*x*-axis) for 19 chemicals of varying toxicity against *V. cholera*. The solid line represents the least squares regression line; dashed lines show the upper and lower bounds of a 95% prediction interval. Linear regression with 95% prediction intervals was performed using GraphPad Prism v5.0b software. The 2 estimates for a compound have an R^2 of 0.88 in log-log space, computed over EC_{50} potencies ranging from 186 nM to 150 μM.

methods. A potential reason for this increase in hit rate is that the halo score is able to pick up compounds with weak effects that still produce characteristic halos. In addition, the current method picks up many hits that are obscured by edge and neighboring compound effects, indicating that the local background correction built into the halo score method helps deconvolute the signal from the noise for these cases.

A 21,120-member commercial library was screened, resulting in the identification of 590 active compounds in *S. cerevisiae*. Among the most active hits in yeast, several have known activities from previously reported screens in other organisms. The algorithm allowed quantification of structure-activity relationships (SAR), and trends were found for a cluster of 2,4-diaminoquinazolines (**Fig. 5** and **Table 2**), a structural class with no previously reported antifungal activity.

The 4 most potent compounds in the 2,4-diaminoquinazoline structure/activity series (4408-0546, 4408-0539, 4408-0537, and 4408-0549) have been shown to modulate hepatocyte growth factor activity, suggesting that they may be useful in the treatment of cancer.[5] The second-most bioactive (4408-0539) has been identified as having antimicrobial activity against *Escherichia coli* and *Pseudomonas aeruginosa*.[6] In addition, 4408-0539 was identified in a cell-based assay as an inhibitor of the tyrosine kinase DYRK1A, a protein encoded on the critical region of chromosome 21 thought to be involved in learning and memory deficits associated with Down syndrome.[7] The least potent compound in the series (5940-0056) is reported to be toxic to mycobacteria.[8]

SAR analysis provides insights that may be useful in the synthesis of more potent derivatives and/or affinity reagents for future efforts aimed at target identification. In the series of compounds active in yeast, replacement of the 2-furanylmethyl group at the R1 position of the most potent 2,4-diaminoquinazoline, 4408-0546, results in a complete loss of activity for 4408-0144 (tertiary amine) and C301-5029 (primary amine), suggesting that a secondary aromatic amine in this position is important for bioactivity against yeast. There is a 2-fold reduction in potency when the 4-methoxy group of the phenyl in the R2 position (4408-0546) is replaced with bromine (4408-0539) and a 4-fold decrease in potency when the *p*-methoxybenzyl group is replaced with an *o*-methylbenzyl group (4408-0537). Changing the *p*-methoxy group of 4408-0546 to an *o*-methoxy substituent (4408-0549) results in a 10-fold decrease in activity, suggesting that para substitution in R2 is critical for activity. When both the 2-furanylmethyl of R1 and *p*-methoxybenzyl of R2 are replaced with benzyl groups (5940-0056), activity is reduced 14-fold. The SAR trends identified with the help of the halo score algorithm allow rapid determination of synthetic directions to take with hits against yeast. We are employing genetic and genomic approaches to identify molecular targets.

Identification of the target(s) of these compounds in yeast may lead to investigation of homologous targets in higher eukaryotes. In addition, their newly discovered antifungal

activity may be due to a novel biological mechanism. The algorithm was equally successful at identifying compound toxicity in microorganisms other than yeast. A calibration in wild-type *V. cholerae* resulted in accurate EC_{50} estimation (see **Fig. 6**). The algorithm can be used to identify compounds with antibacterial activity with novel scaffolds. In addition, the screen can be used to predict the potency of natural products now that several new technologies have emerged for the expansion of libraries containing both crude and purified extracts.[9] Thus, we expect the halo score–based method to generalize to many diverse organisms, as we have shown that it is useful for both a eukaryote and prokaryote.

ACKNOWLEDGMENTS

We thank F. Yildiz for providing *V. cholerae* stocks, K. Bilecen for help with *V. cholerae* agar plate preparation, L. Rocha for EC_{50} determinations, and T. Wipke and J. Davis for assistance with chemical library clustering. RSL, NCG, and WCB were supported by the US Civilian Research and Development Foundation Grant Assistance Program (GTR-G7-044). MHW and JMS were supported by a fellowship from the Alfred P. Sloan Foundation.

REFERENCES

1. Gassner NC, Tamble CM, Bock JE, Cotton N, White KN, Tenney K, et al: Accelerating the discovery of biologically active small molecules using a high-throughput yeast halo assay. *J Nat Prod* 2007;70:383-390.
2. Yildiz FH, Schoolnik GK: Vibrio cholerae O1 El Tor: identification of a gene cluster required for the rugose colony type, exopolysaccharide production, chlorine resistance, and biofilm formation. *Proc Natl Acad Sci USA* 1999;96:4028-4033.
3. Balakin KV, Kozintsev AV, Kiselyov AS, Savchuk NP: Rational design approaches to chemical libraries for hit identification. *Curr Drug Discov Technol* 2006;3:49-65.
4. Zhang JH, Chung TD, Oldenburg KR: A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 1999;4:67-73.
5. Zembower J, Mishra R: 2,4-Diaminoquinazoline compound modulators of hepatocyte growth factor/c-met activity, and use in the treatment of cancer and other dysproliferative diseases. US Patent Application Publication, 2006.
6. De La Fuente R, Sonawane ND, Arumainayagam D, Verkman AS: Small molecules with antimicrobial activity against *E. coli* and *P. aeruginosa* identified by high-throughput screening. *Br J Pharmacol* 2006;149:551-559.
7. Kim ND, Yoon J, Kim JH, Lee JT, Chon YS, Hwang M-K, et al: Putative therapeutic agents for the learning and memory deficits of people with Down syndrome. *Bioorg Med Chem Lett* 2006;16:3772-3776.
8. Wynne O, Johnson PD, Vickers R: Preparation of 2,4-diaminoquinazolines and analogs, and their use for the treatment of mycobacterial infections, especially tuberculosis. PCT International Application, 2008.
9. Koehn FE: High impact technologies for natural products screening. *Prog Drug Res* 2008;65:175, 177-210.

Address correspondence to:
Joshua M. Stuart
Department of Biomolecular Engineering, University of California
1156 High St., Santa Cruz, CA 95064

E-mail: jstuart@soe.ucsc.edu

Scott Lokey
Department of Chemistry and Biochemistry, University of California
1156 High St., Santa Cruz, CA 95064

E-mail: lokey@chemistry.ucsc.edu