xCELLanalyzer: A Framework for the Analysis of Cellular Impedance Measurements for Mode of Action Discovery

Raimo Franke*[a], Bettina Hinkelmann[a], Verena Fetz[a], Theresia Stradal[b], Florenz Sasse[a], Frank Klawonn[c,d], and Mark Brönstrup*[a,e]

[a] Dr. Raimo Franke, Bettina Hinkelmann, Dr. Verena Fetz, Dr. Florenz Sasse & Prof. Dr. Mark Brönstrup
Department of Chemical Biology, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany
E-mail: raimo.franke@helmholtz-hzi.de, mark.broenstrup@helmholtz-hzi.de
[b] Prof. Dr. Theresia Stradal
Department of Cell Biology, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany
[c] Prof. Dr. Frank Klawonn
Biostatistics Group, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany
[d] Department of Computer Science, Ostfalia University, Salzdahlumer Str. 46/48, 38302 Wolfenbuettel, Germany
[e] Center of Biomolecular Drug Research (BMWZ), Institute of Organic Chemistry, Leibniz Universität, 30159 Hannover, Germany

Keywords: mode of action • impedance spectroscopy • target identification • natural products • actinomycin D

Abstract: Mode of action (MoA) identification of bioactive compounds is very often a challenging and time-consuming task. We used a label-free kinetic profiling method based on an impedance readout to monitor the time-dependent cellular response profiles for the interaction of bioactive natural products and other small molecules with mammalian cells. Such approaches have been rarely used so far due to the lack of data mining tools to properly capture the characteristics of the impedance curves. We developed a data analysis pipeline for the xCELLigence Real-Time Cell Analysis detection platform to process the data, assess and score their reproducibility, and provide rank-based MoA predictions for a reference set of 60 bioactive compounds. The method can reveal additional, previously unknown targets, as exemplified by the identification of tubulin-destabilising activities of the RNA synthesis inhibitor actinomycin D and the effects on DNA replication of vioprolide A. The data analysis pipeline is based on the statistical programming language R and is available to the scientific community through a GitHub repository.

Introduction

Cell-based phenotypic assays constitute a highly relevant approach for discovering novel bioactive small molecules and for the consecutive development of novel therapeutic modalities[1,2]. However, the optimization of phenotypically active compounds is hampered by the fact that the identification of their molecular target(s) and of the associated downstream pathways is tedious, as a generally applicable strategy does not exist.[1] In addition, it has become clear that many known drugs exert their effects via modulation of more than one target,[3] a phenomenon coined 'polypharmacology' [4]. An important question for quantitative biology is to determine whether such an imperfect selectivity results is a desirable, enhanced overall effect, or whether unwanted, detrimental off-target effects prevail.

In consequence, there is a strong demand for analytical methods that (i) capture the overall cellular response to small molecules and (ii) give hints to their molecular mode of action by comparing the induced cellular responses to those of known drugs.

A large variety of such compound profiling methods based on molecular signatures derived from transcriptomes[5, 6], proteomes[5, 7], metabolomes[8], based on high-content cellular imaging (HCS)[9, 10], or on bioactivity profiles[11] have been developed and successfully applied[12]. However, a drawback shared by most of them is their destructive nature, i.e. the cellular integrity needs to be disrupted prior to sample analysis. Therefore, usually only single time point measurements are carried out that provide only a snapshot of the cellular state, with the risk that a crucial period of response is missed. One exception is the IncuCyte system developed by Sartorius that offers real-time, automated live-cell imaging even over longer time periods, but still demands the use of labeling reagents[13].

A label-free method that offers continuous, nondestructive monitoring of the cellular state over days is realized by electrochemical impedance spectroscopy (EIS), an application pioneered by Giaever and Keese[14]. EIS measures the resistance that an electrochemical system exhibits upon application of an alternating voltage. In systems that contain living cells, this parameter is influenced by cellular alterations, such as proliferation, morphology, adhesion, migration, volume and apoptosis. As these alterations are in turn modulated by small bioactive molecules, impedance measurements lend themselves for bioactivity profiling applications. Abassi *et al*. were the first to demonstrate that impedance measurements of growing cells over time in the presence of bioactive compounds could be used to infer their mode of action, as compounds with the same mode of action produced similar impedance curves[15].

Further advantages of EIS are that the method is label-free, thereby avoiding artifacts from the introduction of reporters or fluorophors, and that it captures an overall cellular response induced by primary and secondary targets at different time points.

Nevertheless, the method has not often been used in the context of compound bioactivity profiling due to limited tools to analyze the time-dependent cellular response profiles for this particular application.[16] The published studies that demonstrate the usefulness of impedance measurement for compound MoA prediction were carried out in cooperation with the company ACEA Biosciences and used proprietory methods without giving access to raw data and analysis scripts preventing other researchers from using the analysis methods presented in these papers[15, 17, 18].

For example, Atienzar *et al.* assessed the impedance curves by visual inspection due to the lack of an interpretation/classification algorithm.[16] Consequently, they emphasize the necessity of the "development of a reliable analysis tool […] to standardize the interpretation of RTCA patterns".

With the present contribution, we want to address this need by a publicly available pipeline for the analysis and interpretation of impedance curves for the mode of action analysis of small molecules. Elements of the pipeline have been successfully applied by us in the past to provide valuable hints for the elucidation of the mode of action of several natural products[19, 20].

All experiments were performed on a commercial instrument, the xCELLigence Real Time Cell Analyzer (RTCA; ACEA Biosciences), that measures impedance at a single frequency in 96 well plates equipped with gold microelectrodes. The plot of the normalized impedance, called cell index, over time is termed the 'time-dependent cellular response profile' (TCRP).

It is the primary objective of this contribution to address the need for a freely available data analysis pipeline that can be used to analyze TCRP data for MoA prediction of bioactive compounds. We implemented methods in the data analysis workflow to enhance and evaluate inter-assay reproducibility, to identify outliers, and to analyze TCRPs using cubic smoothing splines. The data analysis pipeline was implemented in the statistical programming language R and can be freely accessed via a GitHub repository (https://github.com/raimofranke/xCELLAnalyzer).

A reference set consisting of 59 compounds that cover a wide range of biological activities and the natural product vioprolide A with previously unknown mode of action were analyzed in 16 independent experiments on two xCELLigence machines. Finally, the utility of the pipeline to provide insights into the mode of action

of bioactive compounds is demonstrated by two case studies on the natural products actinomycin D and vioprolide A.

Materials and Methods

Cultivation conditions, cell line and compounds

L929 mouse fibroblasts (DSMZ No. ACC-2) and HeLa cells (DSMZ No. ACC-57) were obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures). DMEM medium, FBS, PBS, Trypsin-EDTA (0,25%), phenol red, were purchased from Gibco/Life Technologies (now Thermo Fisher Scientific, Waltham, MA, USA). Compounds were either purchased from Sigma-Aldrich (St. Louis, MO, USA) or were obtained from the HZI inhouse natural products collection. Stock solutions were prepared using dimethyl sulfoxide (DMSO). L929 cells were grown in DMEM medium with 10% FBS for three days at 37 °C until confluency (inoculated 1:1 from pre-culture that was grown for 3 weeks (1:100 inoculation)). Cells were washed with PBS and trypsinated at 37 °C.

Single-frequency impedance measurement

60 µl of medium were added to each well of the E-plate 96 (ACEA, USA). After waiting for 30 min., the background impedance of the cell culture medium was measured with the xCELLigence RTCA SP (single plate) instrument (ACEA, USA). The analyzer with the E-plate was placed in an incubator at 37°C with 10% CO_2 for impedance measurements. To calculate the cell index, the background was substracted from the measurements that followed by setting the cell index to zero at t = 0.

After pausing the system, the E-plate was removed from the incubator, and 120 µl L929 cell culture (with 100.000 cells / ml, prepared from the 3-day pre-culture) were added to each well of the E-plate. After waiting for 30 min, the E-plate was placed in the incubator again, and cells were grown for 24h. The measurement was halted

again to take out the E-plate and add 1.8 µl compound solution (prepared form a 1000fold concentrated stock solution in DMSO, diluted 1:10 with medium) to each well, resulting in a 1:1000 total diluation of the compound in the well. To observe the maximal effect of the compound on the cellular phenotype, we used end concentrations in the wells corresponding to the IC_{90} determined in an MTT proliferation assay. In cases where~~unless~~ **the concentration was too high, resulting in issues with dissolving the compound or too high toxicity in the assay,~~:~~ ~~Then~~ a concentration between IC_{50} and IC_{90} was used (see Suppl. Table S1).**

Outer wells of the E-plates were not used for measurements to avoid any edge effect and compound replicates were randomly distributed to avoid batch effects (see supplemental material, experimental design considerations and **Suppl. Fig.1**).

After compound addition, the impedance signal was recorded every 5 min for ~66.5 h, corresponding to 800 data points. For the recording of the TCRP the proprietary software RTCA (Version 1.2, ACEA, USA), was used. The system supplies the cultered cells with an alternating voltage at low frequency and measures the current in real time.

The RTCA software records the impedance as a dimensionless parameter called cell index (CI). The CI is calculated by dividing the change in the electrical impedance R at a certain frequency by the nominal impedance value and thereby represents a relative value. Eq. 1 shows how the CI at a given time point t (CI(t)) is calculated.

$$CI(t)=[\frac{R(f_n,t)-R(f_n,t_0]}{Z_n}] \quad (1)$$

f_n is the frequency at which the impedance measurement is carried out; $R(f_n,t)$ is the measured impedance at frequency f_n at time point t; $R(f_n,t_0)$ is the measured

impedance at frequency f_n at time point t_0 (usually the time point when the background is measured) and Z_n is the corresponding nominal impedance value of f_n. For the experiments, a frequency of 10 kHz was used, corresponding to a nominal impedance of the xCELLigence device of 17 Ohms. The raw cell index values were exported from the RTCA software as a csv-file for further analysis in R.

Calculation of the normalized cell index (NCI)

Calculation of the normalized cell index (NCI) was carried out by dividing the cell indices at each time point after compound addition by the cell index at a reference time point (Eq. 2).

As reference time point the last measurement before compound addition was taken

$$NCI_{i,t} = \frac{CI_{i,t}}{CI_{i,t_{ref}}} \quad (2)$$

$NCI_{i,t}$ is the normalized cell index for compound i at timepoint t, $CI_{i,t}$ is the cell index of compound i at time point t and $CI_{i,tref}$ is the cell index of compound i at the reference time point t_{ref} (last measurement before compound addition).

Vioprolide Treatment of Cells, Cell Lysates and Western Blotting

L929 and HeLa cells were seeded subconfluently into six-well plates and the following day treated with the compounds for 16h. Cells were detached by scratching on ice and washed once with PBS before lysis with low salt lysis buffer (50mM Tris pH 8.0, 150 mM NaCl, 5 mM EDTA, 0.5 % NP-40, 1 mM DTT, 1 mM PMSF, Complete EDTA-free from Roche Diagnostics, Mannheim, Germany) for 20min on ice. 25µg of total protein per lane were mixed with ¼ Roti®-Load 1 (Carl Roth GmbH, Karlsruhe, Germany), boiled for 10min and loaded onto precast SDS protein gels (4–

20% Mini-PROTEAN TGX Precast Protein Gels, 10-well, 50 µl, Bio-Rad Laboratories Inc., Hercules, CA, USA). Proteins were transferred by semi-dry blotting (45min, 15 V) onto PVDF-membranes (Merck Millipore, Burlington, MA, USA) and detected with antibodies specific to: γH2AX (A300-81A, Bethyl Laboratories Inc., Montgomery, TX, USA), GAPDH (sc-25778, Santa Cruz Biotechnology, Dallas, TX, USA), p-mTOR (Ser2448, 2971, Cell Signaling Technology, Danvers, MA, USA), mouse IgG, HRP-linked Antibody (7076, Cell Signaling Technology, Danvers, MA, USA), rabbit IgG, HRP-linked Antibody (7074, Cell Signaling Technology, Danvers, MA, USA). HRP-signals were detected using Clarity™ Western ECL Substrate in a ChemiDoc™ Imaging System (both Bio-Rad Laboratories Inc., Hercles, CA, USA). Signal intensities were quantified using Image Lab software (Bio-Rad Laboratories Inc., Hercles, CA, USA).

Data Mining Pipeline

The whole data processing and mining workflow was implemented in the statistical programming language R, Version 3.4.3[21]. The implemented functions and details of the data analysis are explained in the supplemental information and in the R markdown document “xCELLanalyzer.Rmd” that can be accessed via the github repository (https://github.com/raimofranke/xCELLanalyzer).

Results

The premise of the methodology for compound MoA prediction, first introduced by Abassi *et al.*[15], is that compounds with the same MoA produce similar impedance curves (TCRPs). And the prerequisite for MoA prediction would be that the similarity between TCRPs of compounds that have the same MoA is higher than the similarity between TCRPs of compounds with different MoAs. This was found to be indeed the case, as illustrated by cell index curves for the two actin binders cytochalasin D and chondramide C and the two proteasome inhibitors MG132 and bortezomib (**Figure 1**). In the following, the workflow from impedance curve to MoA prediction is discussed in detail.

Workflow of the Data Processing Pipeline

Figure 2 shows an overview of the data processing pipeline that was implemented in the programming language R[21]. It consists of the following steps: the raw cellular impedance data recorded with the ACEA Biosciences RTCA Software (Version 1.2) are imported into R (Version 3.4.3). Quality control and outlier detection is done by the median polish algorithm and plotting of the normalized TCRPs. Then, median TCRPs are calculated from multiple normalized cell indices that correspond to technical replicates in the same plate. To combine biological replicates from multiple independent measurements, a local normalization is carried out followed by outlier detection and a calculation of median cell indices (“medians of medians”). Each of these cell indices represents a descriptor that can be used for subsequent data mining approaches. To achieve a dimension reduction of the high-dimensional data set, cubic smoothing splines were applied for TCRP approximation. The basis spline coefficients spanned a distance matrix that was subsequently used for supervised

and unsupervised data mining approaches. In this study, hierarchical clustering and a rank-based MoA prediction were conducted. In the following, the single steps are described individually.

Pre-Processing of the imported data

Data from the vendor software could be imported without manual formatting, because all parsing and reformatting steps were implemented in the read_xcell-function of our R script.

A global normalization of the raw cell index data was carried out by dividing the cell indices for each time point after compound addition by the cell index that was last measured before compound addition to calculate the normalized cell index (NCI, see Eq.2). The impedance curves were recorded in 2-4 technical replicates. To detect and remove outliers, which could be caused by defective electrodes, a median polish procedure was applied. The sum of the absolute residuals of the median polish fit was used to judge if a technical replicate was an outlier. Replicates, where the sum of the residuals were above of a threshold of 90 were removed from the dataset. If only technical duplicate measurements of a compound existed, both replicates were removed if one of the two impedance curves came above the 90 threshold.

Median TCRPs and local normalization

For each compound the median cell index was calculated at each time point to produce median TCRPs. The median was favored over the mean because of its robustness against outliers.

As a 96 well plate does not contain enough wells to accomodate all reference and test compounds in multiple replicates, datasets from independent experiments have

to be fused. To make biological replicates comparable, a local normalization was established that takes differences in cell growth behavior between plates into account. We used the growth curve of the DMSO solvent blank as a reference, and normalized by subtracting the median cell index of the solvent-treated cells from the median cell index of the compound-treated cells for each time point.

The beneficial effect induced by normalization of TCRPs is exemplified for Wortmannin in **Suppl. Fig. S2** that shows TCRPs before and after normalization with the corresponding DMSO solvent controls. Pairwise plots for all reference compounds are shown in **Suppl. Fig. S3**.

We also observed that the replicates generated on different xCELLigence machines have the highest deviations, despite using the exact same experimental protocol for both machines (datasets 1 to 10 were generated on machine A and 11 to 16 were generated on machine B, see **Suppl. Fig. S3**). This deviation can partly be accounted for using the local normalization.

Dimension reduction using cubic smoothing splines

A central challenge for the automated analysis of impedance data is to provide a mathematical description of the highly complex curve shapes of TRCPs. If a high dimensional space is constructed by considering more and more curve attributes, a distance function for subsequent data mining approaches loses its ability to separate points well in this high-dimensional space, and thereby suffers from a so-called "curse of dimensionality". In order to reduce the number of attributes for each curve, we searched for an appropriate curve approximation model. Due to the complex nature of the TCRPs, normal fitting with polynomials is only satisfactory using high degree polynomials, with the danger of over-fitting and oscillation at the edges of the

interval (Runge phenomenon). We thus decided to use cubic smoothing splines for a piecewise polynomial approximation, that achieved a good fit of the TCRPs while avoiding approximation to noise. The basis spline coefficients were then used as attributes in the subsequent data mining approaches.

To the best of our knowledge, we were the first to apply cubic smoothing splines for dimension reduction of the high dimensional data sets from impedance measurements.[20, 22] Xi *et al.* also reported the use of smoothing splines for functional data analysis of TCRPs, but their approach differs considerably from the one presented here, since they performed a principal component analysis first and applied the B splines basis functions to the principal component scores.[23] In a more recent publication, Zhang *et al*. applied wavelet transforms to TCRP curves at multiple concentrations to infer compound MoA via supervised methods, such as neural networks and support vector machine algorithms.[24] For both publications, the source code is not available.

A rank-based score to evaluate reproducibility

To evaluate the inter-assay reproducibility between biological replicates and also the responsiveness of the assay, we established a score based on rank position. Using the basis spline coefficients as an input, a distance function was applied to generate a matrix that contained all pairwise distances between the biological replicates and between different compounds. The columns of the matrix were then sorted by increasing distance for each compound replicate individually. The premise is that the distances within the group of biological replicates for a given compound of interest should be smaller than the distances between different compounds. The biological replicates should therefore be found on the first ranks of the sorted distance matrix.

As a score for reproducibility, the rank sum is calculated column-wise, for each compound replicate individually. A normalized score that is independent from the number of replicates is obtained by division of the best possible rank sum by the obtained rank sum. For instance, if the three biological replicates of a given compound are on rank 1, 2 and 4, a score of 7 would result. To calculate the normalized score, the ideal score of 6 (ranks 1, 2, 3) is divided by the actual score of 7, resulting in a normalized score of 0.86. The closer the normalized score to 1, the better the reproducibility. In contrast to measures such as the coefficient of variation, the rank-based score does not only give a measure of how reproducible the TCRPs are from experiment to experiment, but also how well they are separated from the curves that other compounds yield. Uniqueness of the TCRP pattern and responsiveness of the assay influence the rank-based score. A good score means that the biological replicates of a compounds are reliably more similar to each other than to other compounds. A compound that induces a strong impedance phenotype (e.g. actin modulators, see below) can have more deviation from replicate to replicate, but is still grouped reliably correctly, because the profile is distinct from others. In contrast, ~~whereas~~ a compound inducing only a weak impedance phenotype can have less deviation between the replicates, but is more likely to be grouped with other less responsive compounds. The score was not only used to assess the inter-assay reproducibility and responsiveness for a given compound, but also as a measure of overall performance of the data processing approach. The latter was obtained by dividing the sum of the normalized scores by the number of profiles. For example, a comparison of different distance functions and data pretreatment methods showed that Euclidean distance and Maximum distance performed better than the Manhatten distance (**Table 1**).

To assess the effect of the local normalization (see above), the total score was calculated for the same dataset with and without local normalization, resulting in average normalized rank scores of 0.479 and 0.274, respectively, using Euclidean distances and scaled data. This clearly demonstrates the improvement in reproducibility when local normalization is used.

For this first part, a score for each biological replicate was calculated. To judge how a group of biological replicates performed overall, a compound group score was calculated by dividing the sum of all replicate scores of all compounds belonging to one group by the number of replicates (**Suppl. Table S2**).

The group score can be used to judge the reproducibility for a certain compound. Chelerythrine, H89, saframycin Mx1, staurosprine and wortmannin have a perfect score of 1, thus are highly reproducibly, whereas neopeltolide[25] and archazolid B have a very low score. TCRPs can fluctuate more from experiment to experiment for certain compounds, and to account for TCRPs that deviate strongly from the other biological replicates, the normalized rank score for each individual replicate can be used to identify and filter outliers. Comparison of the results of the score based approach with the plotted median TCRPs of the biological replicates shows that the smaller the normalized rank score, the stronger the deviation of the TCRP of this particular replicate from the other ones. We used the normalized rank score to identify and remove outliers using a defined threshold of 0.1. Outliers below this threshold were removed. Following the removal of outlier replicates, recalculation of the average normalized rank score showed a considerable improvement from 0.479 to 0.647. Correspondingly, the group scores also improved (see **Suppl. Table S2**). They can be used to judge which compounds behave very reproducible in the xCELLigence assay. In our set of 60 compounds, 15 have a perfect group score of

1.0 after removal of outliers: These are apicularen, argyrin A, cerulenin, chelerythrine, colchicine, H89, LY294002, methotrexate, neopeltolide, oxamflatin, saframycin mx1, staurosporine, trichostatin, tubulysin B and wortmannin.

Hierarchical clustering of the reference data set and MoA prediction

We next wanted to assess whether the optimized processing protocol for TCRPs enabled a correct classification of the reference compounds according to their MoA.

The data mining workflow described above was followed by applying cubic smoothing spline approximation of the filtered median TCRPs of the 59 reference compounds and one compound with unknown MoA (vioprolide A) with local normalization. The basis spline coefficients were scaled by dividing them with their root mean square and then used as attributes to span a distance matrix. Both the compounds and the attributes were subjected to hierarchical cluster analysis. For the clustering of the compounds the dendsort function of the dendsort package was also applied. It optimizes ordering of nodes in a dendrogram, without affecting the meaning of the dendrogram[26]. Clustering of the attributes, denoted as “c1” to “c22”, reveals that the first two coefficients are in separate clusters and show unique patterns, whereas coefficients 10 to 22 show less variability (**Figure 3**). Most information is therefore captured in the first nine basis spline coefficients, corresponding to the early events after compound addition.

Most compounds cluster according to their mode of action, but there are also some exceptions and unintuitive results: Starting from the top, a cluster of five compounds can be identified including four compounds that have an effect on the respiratory chain (cruentaren A, oligomycin, myxothiazol A and CCCP). For the anti-cancer drug doxorubicin, inhibition of cellular respiration has also been described[27].

Oxamflatin and scriptaid, both HDAC inhibitors, can be found in one cluster, and a third HDAC inhibitor, trichostatin is in close proximity, although in another cluster.

A23187 (calcium ionophore) and Indirubin-3-monoxime (CDK inhibitor) are found within one cluster, which could be explained by the fact that cell cycle progression is regulated by calcium dependent pathways[28]. The compounds in the next cluster point in the same direction: cyclosporine A modulates calcium dependent pathways, purvalanol A is a CDK inhibitor and for puromycin, a translation inhibitor, an effect on cell cycle progression was described [29].

The next cluster contains simvastatin and mevastatin, both HMG-CoA inhibitors. The two p38-kinase inhibitors SB202190 and SB203580 are found in one cluster. Also compounds acting on tubulin are found within one cluster, namely griseofluvin, taxol, vinblastine and nocodazol; epothilone B is also found in close proximity. Colchicine and podophyllotoxin, both also acting on tubulin, are found in a cluster a bit further away.

The proteasome inhibitors bortezomib and MG132 are found in one cluster, as well as the protein synthesis inhibitors cycloheximide, myriaporone and anisomycine.

One cluster that is clearly separated contains chondramide C, cytochalasin D and rhizopodin A. These three compounds act on actin, either as depolymerizer (rhizopodin and cytochalasin D) or as an actin stabilizer (chondramide C). While the clusters described above can be interpreted with ease, there are also false positive and false negative groupings. For instance, a separated cluster consists of apicularen, cerulenin and chelerythrine. Apicularen is a V-ATPase inhibitor, but does not cluster with the other V-ATPase inhibitor archazolid B, although their profiles are similar according to a visual inspection. As the effect of apicularen on the TCRP is more pronounced, this difference could be concentration dependent. Chelerythrine is

a PKC inhibitor, but has not been grouped with the other (unselective) PKC inhibitor staurosporine. Also here, the profiles have a similar shape, but for chelerythrine the effect is more pronounced. On the other hand, some compounds cluster closely together although they do not share mechanistic similarities (as far as known today). This concerns, for example, the CRM1 inhibitor ratjadon C and the ACC inhibitor soraphen A, or the topoisomerase inhibitor doxorubicin is placed closely to the ATP synthase inhibitor oligomycin.

In summary, the hierarchical clustering analysis led to many correct groupings of compounds with similar modes of action, thereby proving its utility with respect to giving important hints for the putative mode of action. On the other hand, the hierarchical clustering failed to recognize all similarities and moreover, it generated 'false positive' associations. Such limitations of the hierarchical clustering method have been reported before in the context of other applications[30].

Rank-based mode of action prediction

Because the dendrogram generated by a hierarchical clustering algorithm cannot accurately represent all pairwise distances stored in the distance matrix, it is useful to explicitly investigate the distances calculated for a given compound with unknown mode of action to the reference compound set. We therefore implemented a function in our script that ranks all compounds of the reference set according to their calculated distance to a given query compound (**Suppl. Table S3**).

To examplify the utility of this method, we included the natural product vioprolide A, a cyclic peptide isolated from the myxobacterium *Cystobacter violaceus* with a hitherto unknown mode of action, in the dataset. Vioprolide A exhibits potent cytotoxic effects

and also inhibits Type I interferon signaling, which may be due to the modulation of more than one target[31,32]. In hierarchical clustering and in the rank order list, the compound was grouped closely to rapamycin, a natural product that gave the hint and the name to its target, the kinase mTOR (**Table 2** and **Figure 3**). mTOR, a serine/threonine protein kinase, is part of protein complexes involved in signal transduction (mTORC1 and mTORC2) and thereby a central regulator involved in control of cell growth, proliferation and survival. Only within mTORC1, mTOR is sensitive to inhibition by rapamycin, that binds in a complex with FKBP12 apart from the catalytic kinase domain to mTOR[33]. Rapamycin has been shown to inhibit autophosphorylation of mTOR occurring during mTORC1 activation[34]. For vioprolide A we were also able to detect an inhibiting effect on mTOR phosphorylation status by western blotting (**Figure 4A**). This effect was dose-dependent in HeLa cells, but could not be reproduced in L929 cells, which were used for the xCELLigence assay. The exact mode of action of vioprolide A remained unclear. We therefore considered aphidicolin as a candidate for MoA prediction on rank 2 of the ranked distances. Aphidicolin is known to inhibit DNA polymerase-α and to induce replication stress that leads to DNA damage[35, 36].

In the dendrogram in **Figure 3**, aphidicolin is not in immediate proximity to vioprolide A. Another small cluster in close proximity to vioprolide A consists of the two proteasome inhibitors MG-132 and bortezomib, and the p38 MAPK inhibitor PD169316, for which proteasome modulatory properties have also been described[37].

To investigate whether vioprolide A could induce DNA damage, we checked the level of phosphorylated histone H2AX (phosphorylated at Ser139) by western blot analysis. Detecting such γH2AX (the phosphorylated form of H2AX) is a common indicator of DNA double-strand breaks[38]. Inhibition of the proteasome is also linked

to an increase of the γH2AX level[39], which would provide an explanation why the proteasome inhibitors were close to DNA damage inducers in the rank order list.

In HeLa and L929 cells we indeed observed an effect of vioprolide A on γH2AX level, for HeLa cells even at low concentrations of 5 ng/ml or 6 nM (**Figure 4B**). In L929 cells, γH2AX is detected after treatment with higher vioprolide A concentration (500 ng/ml or 600 nM). These results provide evidence that vioprolide A treatment indeed induces DNA double strand breaks, as its near neighbors in the impedance curve analysis via rank-based MoA prediction.

In a second application example, we paid closer attention to actinomycin D, a peptide antibiotic that also exhibits anti-tumor activities. Actinomycin is known and widely applied as a DNA intercalator and inhibitor of transcription and translation[40]. But surprisingly, actinomycin D was found in a cluster with tubulysin A, an inhibitor of tubulin polymerization, in the dendrogram.[41] In close proximity, the protein synthesis inhibitor emetine and the DNA synthesis inhibitor saframycin mx1 are found. In the ranked-based MoA prediction, tubulysin B and taxol are on the top of the list (**Table 2**). Indeed, Rajagopalan *et al*. observed that actinomycin D inhibited tubulin assembly into microtubules in *in-vitro* assays[42]. Whereas this was observed when actinomycin D was applied to tubulin in equimolar amounts, we here used approx. 10E3 fold lower concentrations. However, actinomycin D was also described to have profound effects on tubulin mRNA production and protein synthesis. Ben-Ze'ev *et al*. showed that tubulin protein levels in cells were highly sensitive to treatment with actinomycin D, along with low levels of translatable tubulin mRNA[43]. Shortage of tubulin in cells in turn results in reduced microtubule turnover and increased microtubule stabilization,

Discussion

A potential problem of compound profiling methods that involve reference compounds concerns the difficulty to find probes that selectively modulate only one target. Many bioactive compounds that are used as potent inhibitors of specific targets have off-targets and/or show additional downstream effects that are not even known or fully understood. This hampers the definition of clear-cut MoA classes, also within the set of 59 reference compounds used for this study.

In addition, the impedance readout is not unbiased in terms of an equally strong response to every mode of action. For instance, compounds acting on the cytoskeleton have a more pronounced effect on the cellular impedance than compounds modulating intracellular pathways. For the latter group, not the direct effect on a target is sampled, but rather the secondary, downstream events become manifested by the impedance curve. These effects might also overlap with compounds that have a different direct target, thereby giving rise to similar curve shapes and to close groupings in the clustering or the rank order tables, and eventually to false positive ‘assignments’.

Another limitation is given by the finite number of MoA classes represented in the reference compound set. If the MoA of a compound of interest is not covered by the reference set, it is *a priori* not possible to find it by this (or any other) correlation method. Even worse, the clustering and ranking will always give a closest match, and the quantitative distances often do not clearly discern such a closest match as a false positive. Therefore, an independent biochemical or biophysical validation of a prediction generated by a profiling method is always necessary. Mode of action

elucidation will most often not be achieved by a single methodology, but rather by an integration of diverse, complementary methods[45].

Nevertheless, the method has proven to be effective in generating new and valid hypothesis on the cellular effects of a compound. For the natural product vioprolide A, an important cellular effect (DNA damage) was predicted and verified in independent experiments. The method is also well suited to discover previously unknown or unexpected compound ("off target") effects, as shown for actinomycin D, that in addition to its effect on transcription and translation modulates microtubule stability.

Our data mining pipeline makes high content screening using impedance measurements accessible. We introduced several improvements to the data mining workflow (median polish, local normalization, rank based score) that lead to an improved reproducibility and easier interpretability of results.

Combination of descriptors derived from other profiling technologies such as automatic microscopy and transcriptomics with the TCRP-derived descriptors is a promising strategy for future investigations to improve the quality of the MoA predictions. Moreover, it would be interesting to generate data in different cell lines at a range of different concentrations to compare the outcomes and eventually also combine the derived descriptors for higher predictive power in future studies.

All the data and scripts used for this study can be accessed via GitHub: https://github.com/raimofranke/xCELLanalyzer.

Acknowledgements

We would like to thank Dr. Herbert Weich for helpful discussions.

This work has been supported by the President's Initiative and Networking Funds of the Helmholtz Association of German Research Centres (HGF) under contract number VH-GS-202, and by the EU-funded European Marine Biological Research Infrastructure Cluster (EMBRIC, code 654008).

References

1. Ziegler, S.; Pries, V.; Hedberg, C.; et al. Target identification for small bioactive molecules: finding the needle in the haystack. *Angew. Chem. Int. Ed.* **2013,** *52*, 2744-92.
2. Moffat, J. G.; Vincent, F.; Lee, J. A.; et al. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* **2017,** *16*, 531-543.
3. Yildirim, M. a.; Goh, K.-I.; Cusick, M. E.; et al. Drug-target network. *Nat. Biotechnol.* **2007,** *25*, 1119-26.
4. Tan, Z.; Chaudhai, R.; Zhang, S. Polypharmacology in Drug Development: A Minireview of Current Technologies. *ChemMedChem* **2016,** *11*, 1211-8.
5. Knezevic, C. E.; Wright, G.; Rix, L. L. R.; et al. Proteome-wide Profiling of Clinical PARP Inhibitors Reveals Compound-Specific Secondary Targets. *Cell Chem. Biol.* **2016,** *23*, 1490-1503.
6. Lamb, J.; Crawford, E. D.; Peck, D.; et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006,** *313*, 1929-35.
7. Zhou, Y.; Li, W.; Zhang, X.; et al. Global profiling of cellular targets of gambogic acid by quantitative chemical proteomics. *Chem. Commun.* **2016,** *52*, 14035-14038.
8. Zampieri, M.; Szappanos, B.; Buchieri, M. V.; et al. High-throughput metabolomic analysis predicts mode of action of uncharacterized antimicrobial compounds. *Science translational medicine* **2018,** *10*, 1-12.
9. Nonejuie, P.; Burkart, M.; Pogliano, K.; et al. Bacterial cytological profiling rapidly identifies the cellular pathways targeted by antibacterial molecules. *Proc. Nat. Acad. Sci. U.S.A.* **2013,** *110*, 16169-74.
10. Bray, M. A.; Singh, S.; Han, H.; et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **2016,** *11*, 1757-74.
11. Xu, T.; Zhu, R.; Liu, Q.; et al. Quantitatively integrating molecular structure and bioactivity profile evidence into drug-target relationship analysis. *BMC Bioinformatics* **2012,** *13*, 75.
12. Young, D. W.; Bender, A.; Hoyt, J.; et al. Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat. Chem. Biol.* **2008,** *4*, 59-68.
13. Single, A.; Beetham, H.; Telford, B. J.; et al. A Comparison of Real-Time and Endpoint Cell Viability Assays for Improved Synthetic Lethal Drug Validation. *J Biomol Screen* **2015,** *20*, 1286-93.
14. Giaever, I.; Keese, C. R. Monitoring fibroblast behavior in tissue culture with an applied electric field. *Proc. Nat. Acad. Sci. U.S.A.* **1984,** *81*, 3761-4.
15. Abassi, Y. A.; Xi, B.; Zhang, W.; et al. Kinetic Cell-Based Morphological Screening : Prediction of Mechanism of Compound Action and Off-Target Effects. *Chemistry & Biology* **2009,** *16*, 712-723.
16. Atienzar, F.; Gerets, H.; Tilmant, K.; et al. Evaluation of Impedance-Based Label-Free Technology as a Tool for Pharmacology and Toxicology Investigations. *Biosensors* **2013,** *3*, 132-156.

17. Fu, H.; Fu, W.; Sun, M.; et al. Kinetic cellular phenotypic profiling: prediction, identification, and analysis of bioactive natural products. *Analytical chemistry* **2011,** *83*, 6518-26.
18. Xi, Z.; Khare, S.; Cheung, A.; et al. Mode of action classification of chemicals using multi-concentration time-dependent cellular response profiles. *Computational Biology and Chemistry* **2014,** *49*, 23-35.
19. Lu, H. H.; Raja, A.; Franke, R.; et al. Synthesis and biological evaluation of paleo-soraphens. *Angew. Chem. Int. Ed.* **2013,** *52*, 13549-52.
20. Sasse, F.; Muthukumar, Y.; Hinkelmann, B.; et al. Deciphering intracellular targets of organochalcogen based redox catalysts. *MedChemComm* **2012,** *3*, 784-787
21. R Core Team. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing: 2017.
22. Ushakov, D. B.; Raja, A.; Franke, R.; et al. Total Synthesis of (±)-Moluccanic Acid Methyl Ester. *Synlett* **2012,** *23*, 1358.
23. Xi, Z.; Khare, S.; Cheung, A.; et al. Mode of action classification of chemicals using multi-concentration time-dependent cellular response profiles. *Comput. Biol. Chem.* **2014,** *49*, 23-35.
24. Zhang, Y.; Wong, Y. S.; Deng, J.; et al. Machine learning algorithms for mode-of-action classification in toxicity assessment. *BioData Min.* **2016,** *9*, 19.
25. Vintonyak, V. V.; Kunze, B.; Sasse, F.; et al. Total synthesis and biological activity of neopeltolide and analogues. *Chemistry* **2008,** *14*, 11132-40.
26. Sakai, R.; Winand, R.; Verbeiren, T.; et al. dendsort: modular leaf ordering methods for dendrogram representations in R. *F1000Res.* **2014,** *3*, 177.
27. Tao, Z.; Withers, H. G.; Penefsky, H. S.; et al. Inhibition of cellular respiration by doxorubicin. *Chem. Res. Toxicol.* **2006,** *19*, 1051-8.
28. Kahl, C. R.; Means, A. R. Regulation of cell cycle progression by calcium/calmodulin-dependent pathways. *Endocr. Rev.* **2003,** *24*, 719-36.
29. Davidoff, A. N.; Mendelow, B. V. Unexpected cytokinetic effects induced by puromycin include a G2-arrest, a metaphase-mitotic-arrest, and apoptosis. *Leuk. Res.* **1992,** *16*, 1077-85.
30. Uygun, S.; Peng, C.; Lehti-Shiu, M. D.; et al. Utility and Limitations of Using Gene Expression Data to Identify Functional Associations. *Plos Comput. Biol.* **2016,** *12*, e1005244.
31. Schummer, D.; Forche, E.; Wray, V.; et al. Antibiotics from gliding bacteria, LXXVI. Vioprolides: New antifungal and cytotoxic peptolides from Cystobacter violaceus. *Liebigs Ann.* **1996,** 971-978.
32. Koester, M.; Pulverer, J.; Sasse, F.; et al. System for the Determination of Molecules Altering the Function of Interferon, Method Therefor and Compounds Altering Interferon Activity. *US Patent Application 2011/0173707 A1* **2011**.
33. Ballou, L. M.; Lin, R. Z. Rapamycin and mTOR kinase inhibitors. *J. Chem. Biol.* **2008,** *1*, 27-36.
34. Jacinto, E.; Loewith, R.; Schmidt, A.; et al. Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive. *Nat. Cell Biol.* **2004,** *6*, 1122-8.
35. Glover, T. W.; Berger, C.; Coyle, J.; et al. DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum. Genet.* **1984,** *67*, 136-42.

36. Mazouzi, A.; Velimezi, G.; Loizou, J. I. DNA replication stress: causes, resolution and disease. *Exp. Cell Res.* **2014,** *329*, 85-93.
37. Leestemaker, Y.; de Jong, A.; Witting, K. F.; et al. Proteasome Activation by Small Molecules. *Cell Chem. Biol.* **2017,** *24*, 725-736 e7.
38. Kuo, L. J.; Yang, L. X. Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In Vivo* **2008,** *22*, 305-9.
39. Krum, S. A.; la Rosa Dalugdugan, E.; Miranda-Carboni, G. A.; et al. BRCA1 Forms a Functional Complex with gamma-H2AX as a Late Response to Genotoxic Stress. *J Nucleic Acids* **2010,** *2010*.
40. Hurwitz, J.; Furth, J. J.; Malamy, M.; et al. The role of deoxyribonucleic acid in ribonucleic acid synthesis. III. The inhibition of the enzymatic synthesis of ribonucleic acid and deoxyribonucleic acid by actinomycin D and proflavin. *Proc. Nat. Acad. Sci. U.S.A.* **1962,** *48*, 1222-30.
41. Khalil, M. W.; Sasse, F.; Lunsdorf, H.; et al. Mechanism of action of tubulysin, an antimitotic peptide from myxobacteria. *Chembiochem : a European journal of chemical biology* **2006,** *7*, 678-83.
42. Rajagopalan, R.; Gurnani, S. Inhibition of microtubule assembly by actinomycin D, an anti-tumour drug. *Chem.-Biol. Interact.* **1986,** *60*, 201-206.
43. Ben-Ze'ev, A.; Farmer, S. R.; Penman, S. Mechanisms of regulating tubulin synthesis in cultured mammalian cells. *Cell* **1979,** *17*, 319-25.
44. Schulze, E.; Kirschner, M. Dynamic and stable populations of microtubules in cells. *J. Cell Biol.* **1987,** *104*, 277-88.
45. Schenone, M.; Dančík, V.; Wagner, B. K.; et al. Target identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.* **2013,** *9*, 232-240.

Tables and Table Legends

Table 1. Average normalized rank scores for different distance measures and data preprocessing methods.

Data Preprocessing	Euclidean Distance	Maximum Distance	Manhattan Distance
None	0.3780	0.4038	0.3424
Scaled	0.4793	0.4729	0.4119
Centered and Scaled	0.4788	0.4725	0.4102

Table 2. Rank order list for MoA prediction of vioprolide A and actinomycin D

Rank	**Compound (increasing distance to vioprolide A)**	**Compound (increasing distance to actinomycin D)**
1	Rapamycin	Tubulysin B
2	Aphidicolin	Taxol
3	Bortezomib	Saframycin Mx1
4	SaframycinMx1	Griseofulvin
5	Emetine	Emetine
6	PD169316	SB202190
7	Etoposide	Soraphen
8	MG132	Oxamflatin
9	Amanitin	SB203580
10	Apicidin	Vinblastin

Figure Legends

Figure 1. TCRPs of two compounds acting on the actin cytoskeleton (left) and two proteasome inhibitors (right). NCI = normalized cell index.

Figure 2. Flowchart summarizing the data mining workflow of the TCRP data analysis.

Figure 3. Hierarchical clustering and heatmap representation resulting from processing of the TCRPs of 59 reference compounds and one compound with unknown MoA (vioprolide A). The color scale represents the values of the scaled basis spline coefficients derived from the cubic smoothing splines fitting. They can be viewed as descriptors representing the phenotype induced by the compound treatment.

Figure 4. Effects of vioprolide A on the level of phosphorylated mTOR and γH2AX (histone H2AX phosphorylated at Ser139). **A**. HeLa cells were treated as indicated for 16h, and lysates were subjected to western blotting and detection with a phospho-mTOR (p-mTOR) specific antibody. GAPDH was used for normalization of equal loading. Graph depicts mean values of three independent experiments for the ratio of p-mTOR/GAPDH intensity. **B**. HeLa or L929 cells were treated as indicated for 16h and subjected to western blotting and detection with a γH2AX-specific antibody.