NPR

Target Identification by Image Analysis

SCHOLARONE™
Manuscripts

NPR

REVIEW

Target identification by image analysis

V. Fetz,[a, b, †] H.-P. Prochnow,[a, †] M. Brönstrup[a, b*] and F. Sasse[a*]

Each biologically active compound induces phenotypic changes in target cells that are characteristic for its mode of action. These phenotypic alterations can be directly observed under the microscope or made visible by labelling structural elements or selected proteins of the cells with dyes. A comparison of the cellular phenotype induced by a compound of interest with the phenotypes of reference compounds with known cellular targets allows predicting its mode of action. While this approach has been successfully applied to the characterization of natural products based on a visual inspection of images, recent studies used automated microscopy and analysis software to increase speed and to reduce subjective interpretation. In this review, we give a general outline of the workflow for manual and automated image analysis, and we highlight natural products whose bacterial and eucaryotic targets could be identified through such approaches.

Introduction

Using bioactive natural compounds for medical applications is part of the human cultural heritage. The most successful strategy until today for the identification of their bioactivity and their medical application has been coined 'forward pharmacology'.[1] In this approach, natural products are tested in phenotypic assays of high relevance for *in vivo* pharmacology. For example, the assay may probe their ability to inhibit the growth of bacterial pathogens or cancer cells, or to induce cellular differentiation. The major drawback of phenotypic assays is that the molecular interaction partner(s) of the bioactive compound remain unknown, rendering a rational optimisation of their target affinity and selectivity difficult. While a profound knowledge of the molecular target(s) of a compound is indispensable for its use as a tool in biochemical research[2] and/or its development as a therapeutic agent,[3] the elucidation of the target(s) starting from phenotypic observations is still a tedious and time consuming process, as a widely applicable, generic protocol does not exist.[1, 4, 5] The large number of possible binding partners render the search for a target similar to a quest for the "needle in a giant haystack".[4]

One way to generate a specific hypothesis about the mode of action of a compound of interest is to correlate the cellular phenotype induced upon compound treatment to the phenotype of reference compounds with known mode of action. This approach is based on the assumption that the modulation of a particular target results in a specific phenotype that is characteristic for such a modulation; therefore compounds that induce highly similar phenotypes may share the same molecular target. The cellular phenotypes can be molecular signatures (e.g. the cellular transcriptome, proteome or metabolome),[6, 7] bioactivity patterns,[8] or changes in cellular morphology.[5, 9, 10] An important strength of techniques based on such correlation signals is that they reflect 'global' compound effects that capture direct target interactions as well as the downstream consequences of the interactions in a hypothesis-free manner. In addition, they usually do not require labelling or chemical modification of the compound of interest, thereby avoiding an artificial perturbation of its cellular interactions. On the other hand, a major limitation is that they can only detect matches to known modes of action of the reference set; therefore, they are not suited to disclose targets that are not addressed by the reference set. Moreover, as various algorithms always propose 'most similar' signatures, there is an inherent danger of not clearly discerning false positive matches. Therefore, all target predictions derived from correlation of cellular phenotypes need to be verified by subsequent biochemical or biophysical experiments.

Here, we will review how imaging techniques that monitor changes in cellular morphology have been applied to the mode of action analysis of natural products and other small molecules. The analysis can be done visually by comparing microscopic images. The information content of such images can greatly be enhanced by staining specific cellular components, especially with immunofluorescence techniques. The introduction of immunofluorescence has been a decisive step in elucidating modes of action by image analysis, as they are able to visualize almost every potential target in the cell.[11] The second important step was the development of microscopes that are able to record large amounts of

a. Helmholtz Centre for Infection Research, Department of Chemical Biology, Inhoffenstrasse 7, D-38124 Braunschweig
b. German Centre for Infection Research (DZIF), partner site Hannover-Braunschweig, Germany.
† V.F. and H.-P. Prochnow contributed equally to this work.
* mark.broenstrup@helmholtz-hzi.de, florenz.sasse@helmholtz-hzi.de

microscopic images automatically. Using specific image analysis software, the information stored in these images serves to construct detailed phenotypic profiles of compounds that are compared by statistical analysis.[9]

In the following we will give an overview of target identification by image analysis starting from a visual inspection of cellular morphology and of immunofluorescence pictures, followed by a general outline of the workflow for an automated image analysis. Natural products whose bacterial and eukaryotic targets could be identified through such approaches are highlighted along this path.

Phenotypic changes related to specific modes of action

Morphological phenotypes

Each biologically active compound induces phenotypic changes in target cells that are more or less characteristic for its mode of action. The simplest cases are striking alterations in the morphology of incubated cells that can easily be observed under the microscope. Myxothiazole, a strong inhibitor of complex III of the respiratory chain[12] induced striking changes in the morphology of L-929 mouse fibroblasts. The cells became bigger and more circular in shape with an outspread cytoplasm. Their shape resembled that of "fried eggs" (Figure 1). This fried eggs effect, easily seen directly under the microscope or after Giemsa staining, was found to be characteristic for this type of inhibitors. The same morphological changes were observed again with neopeltolide and helped to elucidate the mode of action of this marine natural product, which is unrelated to myxothiazole in terms of structure and phylogenicity of the producer (Scheme 1).[13]
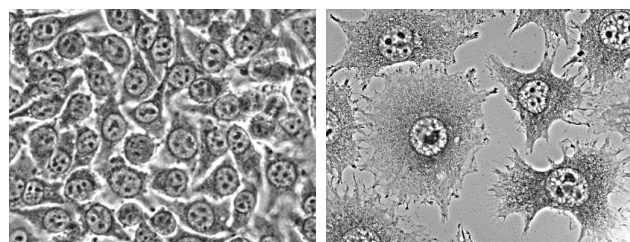

Figure 1. L-929 mouse fibroblasts show a "fried eggs effect" when incubated with inhibitors of complex III of the respiratory chain.[13] The cells in the right image were incubated with neopeltolide (50 ng/mL) for 1 day. The left picture shows control cells. Cells were stained with Giemsa.

Alterations in the shape and size of the cellular nucleus and the number of nuclei in a cell can be observed either visually or after staining. Propidium iodide and DAPI are sensitive fluorescent dyes that can easily be applied to stain nuclei and chromosomes in alcohol-fixed cells. Compounds interfering with mitotic spindle formation induce a characteristic multimininucleation (Figure 2) that is due to mitotic slippage.[14] DAPI staining was used to screen for spindle interfering compounds like paclitaxel and epothilone, which are used as anticancer drugs.[15] The active principle of some positive extracts of the myxobacterium *Sorangium cellulosum* was shown to be disorazol, which had been detected ten years earlier as a highly active compound.[16] The induction of multimininucleation gave first hints to elucidate its mode of action. In subsequent studies, it was shown that disorazol induced microtubule depletion in cells and inhibited tubulin polymerisation in vitro.[17]
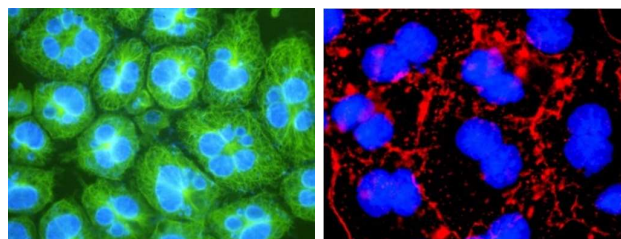

Figure 2. The morphology of cellular nuclei and their number per cell can easily be visualized by DAPI staining. Left: Compounds interfering with spindle assembly induce the formation of multimininucleated cells.[17] L-929 cells were incubated with disorazol A1 (50 pg/ml) for 2 d. Right: Compounds interfering with actin polymerisation induce cells with double nuclei. Human A-431 epidermoid cancer cells were incubated with chivosazole F (200 ng/ml) for 16 h. (Microtubules are stained green, microfilaments red).

Compounds interfering with actin polymerisation lead to an increase of nuclei in the cell, as they inhibit the function of the contractile ring, which consists of myosin and actin. In the example depicted in Figure 2, the cytokinesis at the end of mitosis was inhibited, which resulted in many cells having a double nucleus. This phenotype can clearly be discriminated from the multimininucleation effect. The observation of double nuclei led to the elucidation of the modes of action of the myxobacterial products chondramide and chivosazole.[18, 19]

Due to their bigger size, it is much easier to observe specific morphological changes in eukaryotic cells than in prokaryotes. But in principle it should also be possible to find changes in bacterial morphology that are typical for a certain mode of action. A striking phenotype was observed with acyldepsipeptides (ADEPs), which target ClpP, the core unit of a major bacterial protease. ADEPs induced an uncontrolled proteolysis which led to inhibition of bacterial cell division. As a consequence, a filamentation of *Bacillus subtilis* was observed when incubated with ADEPs.[20] An unusual morphology of altered, elongated mycobacteria was also the first hint to a novel mode of action for griselimycins, cyclic peptides produces in *Streptomyces* with a strong activity against *Mycobacterium tuberculosis*.[21] Half a century later, genome sequencing techniques enabled the discovery that inhibition of the DNA sliding clamp DnaN caused the unusual phenotype.[22]

Due to the undercritical size of a well-characterized reference set, a prediction of bacterial mechanisms based on morphology alone is not possible yet.
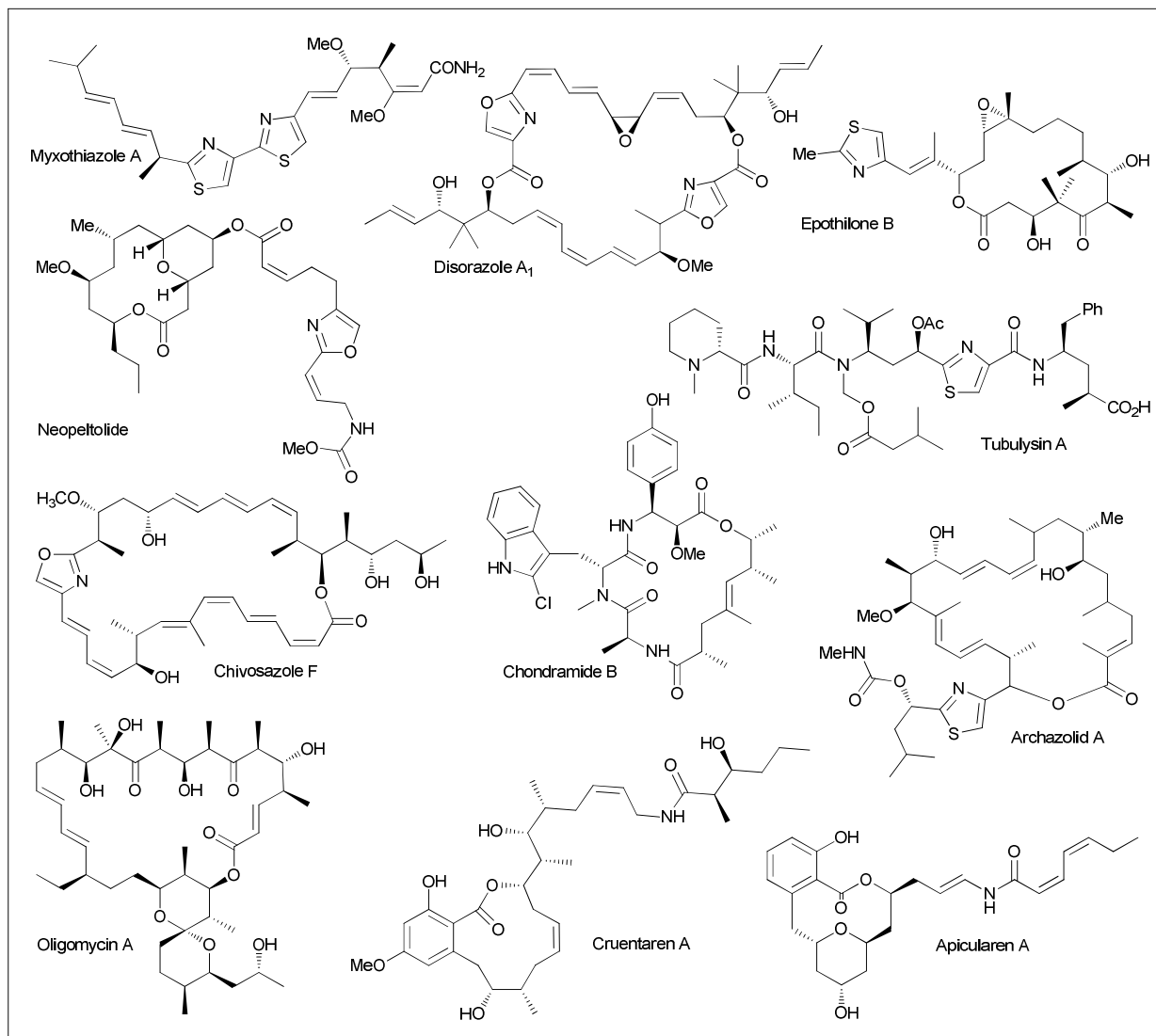
Scheme 1: Natural products that induce specific cellular phenotypes

Phenotypes based on cellular protein patterns

Fluorescent and especially immunofluorescence techniques are able to specifically visualize almost each protein in the cell.[23] They highly increase the number of phenotypes that can be distinguished in cells when incubated with bioactive compounds. Alterations of the main structures of the cytoskeleton can easily be made visible. F-Actin can directly be stained by fluorescently labelled phalloidin, a toxin isolated from *Amanita phalloides* mushrooms. Microtubules are stained by an immunostaining protocol using a primary antibody against tubulin and a secondary, fluorescently labelled antibody that recognizes the primary one.[11] Recently, far-red fluorogenic probes for live-cell imaging of the cytoskeleton were designed that show minimal cytotoxicity with excellent brightness and photostability. Silicon-rhodamine was conjugated to docetaxel and desbromo-desmethyl-

jasplakinolide, which bind to microtubules and F-actin, respectively. The interaction with the polar protein surfaces switched the fluorophores into the ON state.[24] Photostatins are microtubule inhibitors that can be switched on and off in living cells by visible light to optically control microtubule dynamics.[25]

Staining F-actin of cells incubated with the myxobacterial compound chivosazole A or F isolated from *Sorangium cellulosum* showed a depletion of microfilaments within 15 min, resulting in short pieces and small spots of F-actin. The spottiness of the actin cytoskeleton after one day of incubation could be quantified by image analysis. Follow-up experiments proved that the chivosazoles inhibit actin polymerization.[19] On the contrary, chondramides isolated from *Chondomyces crocatus* induced stronger actin filaments with knots and finally big F-actin clumps (Figure 3).[18] *In vitro* experiments with isolated actin showed an enhancement of actin polymerisation by chondramides.
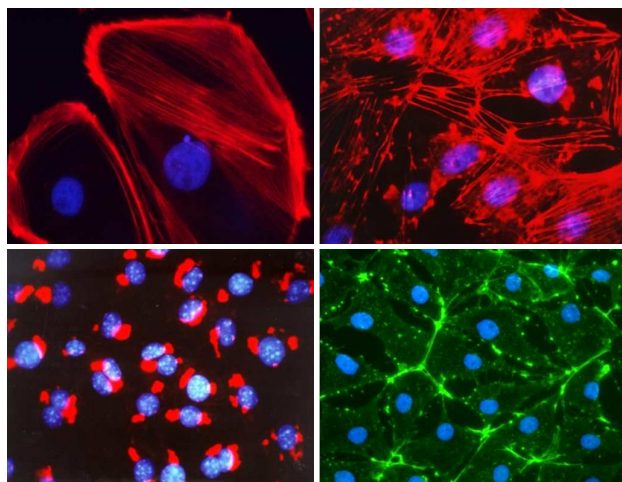
Figure 3. Phenotypic changes in the F-actin cytoskeleton stained by fluorescently labelled phalloidin (red or green). Top left: PtK2 cells treated with the vehicle only; top right: Ptk2 cells incubated with chondramide B (200 ng/ml) for 2 h; bottom left: F-actin of L-929 cells treated with chondramide B (200 ng/mL) overnight; bottom right: PtK2 cells incubated with chivosazole F (200 ng/ml) for 15 min. Nuclei were stained blue.

Compounds acting on tubulin through inhibition or enhancement of tubulin polymerisation induce a depletion of microtubules in the cell, or they give rise to microtubule bundling, respectively.[26] The observation of microtubule depletion led directly to the elucidation of the mode of action of tubulysins.[27, 28] Both categories of compounds interfere with the high dynamics of these structures, which are particularly sensitive during mitotic spindle formation. Abnormal spindles with multipolar configuration were reported for compounds interfering with tubulin polymerisation like tubulysins and disorazols.[17, 27] Multiple asters are typical phenotypes of microtubule stabilizing natural products like paclitaxel, epothilones, and taccalonolides (Figure 4).[29, 30] Detailed studies with GFP-ß-tubulin expressing HeLa cells also showed differences due to the compound's specific mode of action. Paclitaxel-induced asters often coalesced over time resulting in fewer, larger asters whereas numerous compact asters persisted once they were formed in the presence of the taccalonolides.[30]
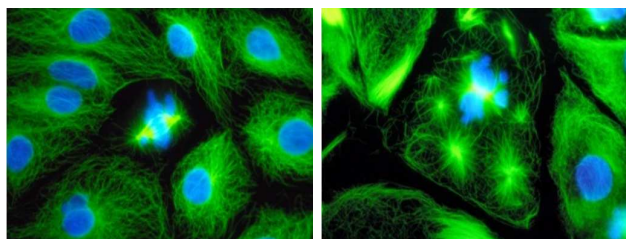

Figure 4. Microtubules in interphase and mitotic PtK2 cells stained in green by a tubulin specific antibody. Control cells show a normal bipolar mitotic spindle (left), cells incubated with epothilone B (50 ng/mL) overnight show multiple asters instead (right). Nuclei and chromosomes were stained blue by DAPI.

Also compounds that do not target tubulin directly can induce specific phenotypic changes of spindle formation. A prime example is monastrol, an inhibitor of the motor protein kinesin-5, which is needed to separate the centrosomes.[31] Monastrol induces monopolar spindle formation.

The endoplasmic reticulum (ER) is an organelle that forms a network that spans over the whole eukaryotic cell. Its structure can be visualised by staining HSP90B1 (also known as endoplasmin, gp96, grp94 and ERp99), a chaperone protein that is located in the ER membrane.[32] Observing phenotypic changes in the ER structure, *i.e.*, in the HSP90B1 distribution, helped elucidating the modes of action of the myxobacterial products archazolid and apicularen, which showed the same phenotype as concanamycin, a known inhibitor of V-ATPase.[33] It also gave a valuable hint for the mode of action of cruentaren, whose phenotype resembled oligomycin, a known inhibitor of F-ATPase (Figure 5).[34]
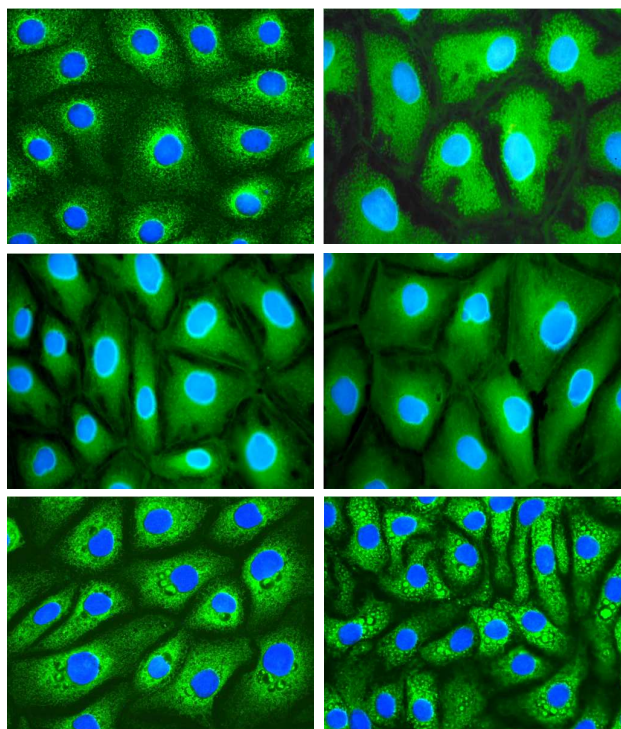
Figure 5. PtK2 cells were incubated with different ATPase inhibitors overnight, fixed and stained green for HSP90B1, a marker for the ER. The cells were treated with (from top to bottom) the vehicle only, oligomycin A, apicularen A (left panel) and thapsigargin, cruentaren A, and archazolid A (right panel). Each type of inhibitor induced a different phenotype, and inhibitors with the same mode of action showed the same phenotype. The images obtained with F-ATPase inhibitors oligomycin and cruentaren are similar, as well as those with V-ATPase inhibitors apicularen and archazolid. Thapsigargin, an inhibitor of Ca-dependent ATPase, induced a third phenotype.

High throughput screening (HTS): HTS describes the process of testing large numbers of chemical compounds (in the order of >10^5/week) for biological activity in pre-designed testing systems, which usually are biochemical *in vitro* assays. Due to the focus on throughput and speed mostly single read out measurements in 384- or 1536-well microtiterplates are performed.[43, 58, 59] HTS is mostly used for the identification of bioactive compounds from libraries of synthetic small molecules and/or natural products.

High content screening (HCS): HCS is regarded as a combination of high throughput screening with cellular imaging. The data obtained are multiple image-based measurements derived from a cell-based assay. Phenotypic screens for cellular effects of bioactive compounds usually draw on HCS. In a HCS experiment assay handling and data evaluation are more complex, thus going along with a lower throughput of tested compounds (in the order of 10^4/week). HCS requires robotic handling platforms and an automated imaging system for the arrayed cell sample (384 well) as well as specialized image analysis software and bioinformatics data management for the interpretation of the multidimensional results. Cellular morphology or alterations in the amount of cellular components (proteins, RNAs, ions) are most commonly visualized by using fluorescent protein-tags, fluorescent proteins or physiological indicator dyes. A special benefit of HCS is the possibility to monitor effects on a single cell level. The stored images provide the opportunity to visually inspect the cellular morphology induced by hit compounds and to discriminate from false positives.

High content analysis (HCA): HCA bears on the same instrumentation and methods as HCS. In contrast to HCS, which usually aims at screening medium sized (>10^4) compound collections, HCA is generally performed on a lower number of compounds. Instead, the total parameters (descriptors) extracted from each image are higher (up to 100). When multiplying the number of single cells analyzed with the serial dilutions of each compound, a HCA campaign can easily generate billions of single data-points.[9] Hence, HCA requires considerably more efforts regarding data processing and data reduction as compared to HCS.[55, 59, 60]

Cellular Profiling: Cellular profiling is applied for comparing cellular reactions to bioactive compounds with each other. In general, profiling refers to the generation of distinct profiles or footprints from datasets in order to identify or predict certain patterns or correlations. In a biological context, methods like transcriptional profiling and proteomic profiling start from molecular measurement of cellular responses to different perturbations. Data gathered are in turn used to generate profiles which allow for receiving information on compound activity and target mechanisms. However, the aforementioned biological profiling techniques are limited as they can only measure an average from a population of treated cells. Cellular profiling circumvents this problem by considering data obtained from single cells and therefore bears heavily on HCA. Evaluation of HCA incorporates the use of descriptors that were calculated from image analysis for creating a multi-dimensional cellular profile (*e.g.*, intracellular protein translocation, organelle structure changes, overall morphology changes, three dimensional structure modifications),[38] reflecting the phenotypic signature of a cell treated with a given compound.

High content image analysis

Automated microscopy as a basis for high content analysis and cellular profiling

As illustrated above, the visual assessment of phenotypical changes of cells can add a significant measure to target-/mode of action-predictions of yet mechanistically uncharacterized compounds. However, comparing greater numbers of microscopy images taken from differently treated cells by the naked eye and evaluating various phenotypical parameters makes the process extremely time-consuming, with only a relative small number of cells being visualized at a time. Eventually, such analyses bear the risk of introducing a bias to the evaluation due to subjective estimations made by the experimenter, especially when comparing minute variations in fluorescence staining.

Thus, to combine more subtle and unbiased approaches with a higher throughput, an automation of the whole image-acquisition/data-gathering/evaluation-process is inevitable.

Fluorescence microscopy has already been well-established since the end of the 1980s. Yet, the evolving number of multicolour fluorescent dyes, the introduction of high throughput plate readers, advances in digital imaging microscopy together with the emergence of high-performance computer hardware were prerequisites for the invention of the first automated microscopes in the mid-to-late 1990s.[35, 36, 37] Nowadays a number of commercial providers (Supporting Information, Table S1) and a growing number of open-source informatics tools for image analysis are available.[38, 39] Modern automated microscopes can read whole microtiterplates within minutes depending on the exposure time, the number of images acquired per well, the number of fluorescence channels used and the image resolution. This implies that imaging of one microtiterplate can easily produce tens of gigabytes of image data, which in turn requires proper storage systems.[38] As the whole acquisition process underlies automation, the data generated can be regarded as completely unbiased and statistically more robust than images taken by the experimenter from a manually operated microscope. On this account, automated microscopy has been regarded "as a technology to bridge the gap between depth and throughput of biological experiments"[39] and thus provides the basis for high content screening (HCS), high content analysis (HCA) and cellular profiling.

HCA has proven to be powerful for the generation of cellular profiles. Besides possessing the capability to analyse processes like protein phosphorylation, receptor/ligand interactions, cellular uptake, protein expression, cell cycle regulation, enzyme activation or cell proliferation, HCA excels at discerning cell-morphological changes from images of thousands of individual cells, which are generally not traceable by conventional biochemical methods. Morphological changes include intracellular protein translocation, organelle structure changes (e.g. changes in mitochondrial membrane potential, cytoskeletal remodelling, formation of micronuclei or quantification of internalization) and three dimensional structure modifications.[38] This approach then allows for

profiling of dose-dependent phenotypic effects induced by different compounds targeting distinct cellular processes, *e.g.*, cytostatic agents, transcription inhibitors, translation inhibitors or agents interfering with DNA replication.

The following sections will briefly outline how to conduct a HCA and how the data collected have been used to generate cellular profiles in order to classify orphan compounds.

Acquiring primary microscopy data from biological samples

The first step in the process of a HCA is the creation of arrays of biological samples in microtiterplates. Typically human-borne cell lines or primary cells are chosen as models for *in vivo* systems. One should always bear in mind that each cell type might respond in a different way to a given compound depending on its proteome, its membrane permeability or its physiological origin in general.[40]

Arrayed cells are treated with test and reference compounds at various concentrations in order to obtain reliable comparative phenotypic profiles. After treatment cells are usually fixed, washed and stained in an automated manner.[38, 41] This procedure implies that any data obtained reflect a single endpoint. Thus, the half time of cellular responses to a certain treatment needs to be estimated by the experimenter, and the time of fixation has to be set accordingly.

Hereafter, microscopic images are acquired by automated microscopes making use of laser- and/or image-based autofocus optics, so that microtiterplates are rapidly imaged.[23] By choosing lower magnifications (5× – 10×) higher cell numbers can be imaged at a time, which is generally desirable for cellular profiling approaches to obtain statistically more meaningful results.

From image data to numeric data

Each image set contains large amounts of imaging data. For example, if cells were imaged using three different channels and four sites of a well were visualized per sample, the 2D-readout of a whole 384-well will create 4608 single images. Such numbers cannot be inspected by eye and thus the huge amount of information contained in respective images has to be extracted bioinformatically, *i.e.* it has to be converted to numeric data. For example, the size of an object can be expressed by the constituting number of pixels. It should be noted that high quality images are fundamental for a reliable gain of information, as algorithms will generate numeric data even out of bad images (*e.g.*, blurry, out of focus, artefacts).

The process of image conversion into numeric data comprises three key steps that have recently been reviewed in detail: I) image pre-processing, II) object identification and segmentation and III) feature extraction.[38, 42] In brief, after image pre-processing (involving background corrections and other procedures), objects have to be identified and segmented. In the example given in Figure 6, cell nuclei were stained with the DNA-binding, blue-fluorescent dye Hoechst 33342 (Hoechst). The line drawn across the Hoechst-stained nucleus marks the section for which a fluorescence intensity profile was generated. This step exemplifies the direct

conversion of image information into numeric data, as the fluorescence intensity is now expressed as relative grey values. With the intensity profile at hand, minimal fluorescence intensity may be determined as a threshold value for the identification of this nucleus (red arrowhead). Each pixel with a higher intensity is then considered as part of the nucleus of that particular cell. Based on the intensity levels a mask for all Hoechst-stained objects in the image is calculated by the software.
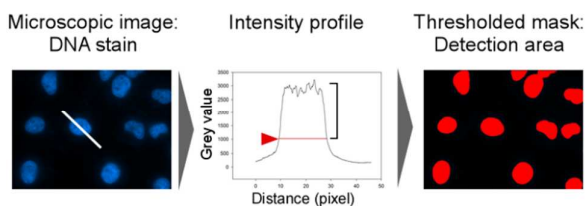

Figure 6. Conversion of image data to numeric data. The microscopic image on the left shows Hoechst stained nuclei of eukaryotic cells. The white line marks the detection region for the fluorescence intensity profile shown in the middle. A threshold value (red arrowhead) is set for defining a cellular object by its fluorescence intensity. The pixel area above this threshold is then considered as cellular object and marked by software generated masks as depicted on the right.

A subsequent segmentation step is essential for separating cells grown as a confluent layer in order to assign the features correctly to each individual cell.

Feature extraction refers to the act of measuring values from shapes or portions of objects in a given image. Such features, also termed descriptors, are extracted from the information covered by the respective thresholded masks, representing detection areas. In the case of DAPI-stained nuclei depicted in Figure 7, features like number, size, shape, fluorescence intensity or nuclear texture might be extracted, as they are all integral part of the thresholded mask defined by the object-identifying algorithm.[43] Such features are captured for each cell of an analysed image area; they serve for the generation of cellular profiles induced by different compound treatments.

Remarkably, a higher number of features does not necessarily generate more significant cellular profiles.[44, 45] It is more relevant to select features that reasonably represent the cellular reaction in response to a given compound. For instance, in a setup that aims at probing cytostatic agents, a feature like "shape of tubulin-staining" will add more informative content to a cellular profile than a feature like "average intensity of nuclear staining".

From numeric data to target identification

With the cellular profile in form of numerical features (descriptors) at hand, the mode of action of a compound of interest can be predicted by comparing its descriptor set to the sets of reference compounds with known modes of action. Figure 8 depicts exemplary cellular profiles of 28 different references and one compound assumed to have an unknown mode of action (denoted "X"). Each profile consists of 15 features derived from three fluorescence channels.
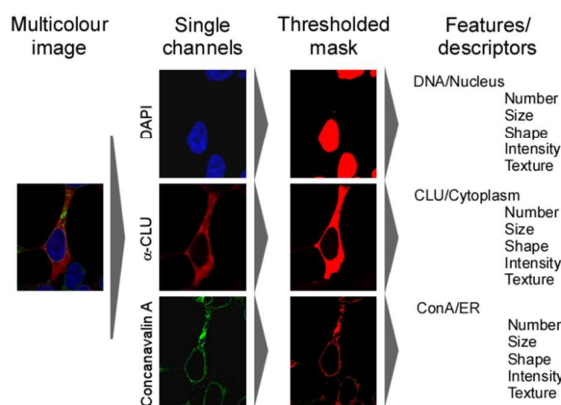
Figure 7. Conversion of image data into features. The multicolour image shows triply stained cells. The nucleus is stained with DAPI (blue), ER-membranous compartments are visualized with Alexa Fluor 488-conjugated concanavalin A (green) and the cytosolic isoform of clusterin (CLU) was immunodetected using a cyanin 3-labeled antibody. For the single channel images thresholds were determined. All pixels included for analysis are combined below the channel-specific thresholded mask. From the respective masks, distinct features can be extracted that are characteristic for every object.

For comprehensive visualization, the response of a given feature has to be normalized, so that the relative intensity range for all features is equal. Based on this, a colour intensity range can be applied expressing the relative change of a feature in response to a certain treatment. In a next step, hierarchical cluster analysis was performed to calculate the extent of similarity between cellular profiles of differentially treated samples.[46] The Euclidian distance as a similarity measure between the different profiles is then plotted as a dendrogram to indicate clustering of given compounds. In the example given in Figure 8, compound X is closely clustering together with reference compound 20, pointing towards a potential similar mode of action.

Several profiling approaches have been described where different strategies have been applied in order to classify reference compounds. In these broadly conceived studies, the number of cellular features extracted from microscopic images vastly exceeds the 15 features depicted in our example. A simple cluster analysis is not feasible anymore, as underlying mathematical algorithms cannot incorporate any number of descriptors. Hence, numeric data obtained from these HCA must further be converted bioinformatically by the use of different mathematical strategies (Table S2). In an important comparative study by Ljosa *et al.*, a single image dataset was used to create compound profiles following different bioinformatical approaches (Figure 9). Subsequently, the accuracy of the obtained mode of action predictions was determined; the authors concluded that "the profiles that best represented the phenotypes were obtained using factor analysis", with an accuracy of 94 % in correctly classifying compounds with different modes of action.[47]
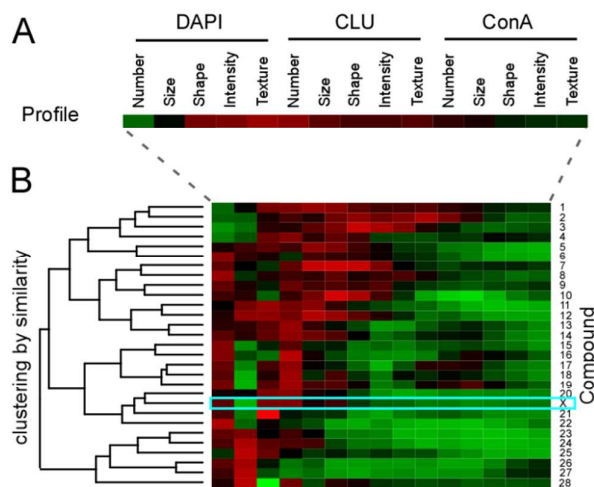

Figure 8. Profile clustering for detecting similarities. A. Exemplary compound profile consisting of measured features for the different fluorescence images and channels. B. Hierarchical cluster analysis of 29 compound profiles (28 reference compounds and one unknown X). The dendrogram on the left visualizes the relationship between different profiles. Highly similar profiles are clustered together whereas diverging profiles are located far away from each other.

Application examples for cell-based profiling

In a study by Young *et al.* arrayed cells were treated with a total of 6,547 different compounds, 58% of which were of natural origin.[48] HCA was performed with the stained cells and cellular profiling was achieved by mining the numerical data obtained with factor analysis (Figure 9 E). A total of 36 cytological features were extracted and reduced to 6 significant factors. For instance, 12 of the original features were combined to a single factor "nuclear size". From the relative change in the value of the different factors, cellular profiles were generated by cluster analysis. Eventually, the top 5% of the whole screening set (211 compounds) whose induced phenotypic responses were significantly different to the average control phenotypes (i.e. they show the highest Euclidean distance) were identified as hits. 96% of all hit-compounds with similar structure showed strikingly similar phenotypes, such as the cyclic depsipeptides aurantimycin A and diperamycin or the glucocorticoids clobetasol-17-propionate and dexamethasone. However, it became evident that a given phenotype must not necessarily point to a single structural class of compounds. Instead, structurally different classes of compounds may produce similar phenotypes, as shown for entobex. The latter clustered together with abovementioned glucocorticoids, but is structurally completely unrelated to these. In additional studies, the authors performed a computed target prediction of their hit-compounds by means of an annotated chemogenomics database (WOMBAT). Through combining the results obtained from the target-prediction algorithm together with the phenotype profiles it was confirmed, that phenotypes correlate well with the predicted compound targets.
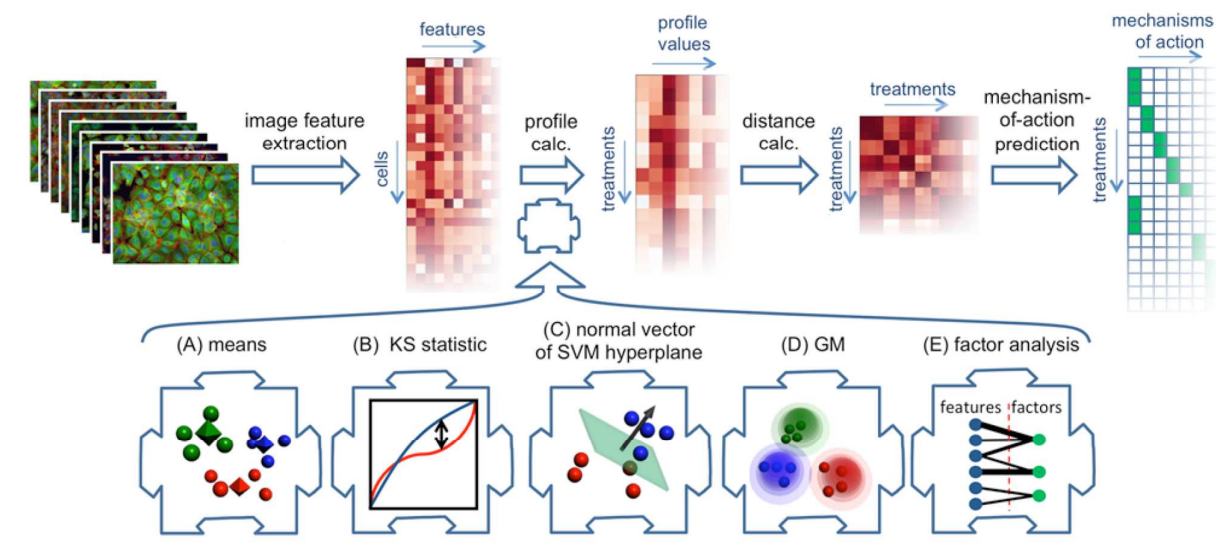
Figure 9. Overview of bioinformatic strategies for generating compound profiles. (Top) Experimental design. Cultured cells in microtiter plates are compound treated, labeled, fixed, and imaged. Then features are extracted from respective images. One of the profiling methods under investigation condensed these measurements into a profile (vector of numbers) that describes each sample. A sample with unknown mechanism of action (MOA) was predicted to have the same MOA as the sample whose profile is most similar to that of an unknown sample. (Bottom) Illustrations of five profiling methods. (A) Means of raw per-cell features. (B) Kolmogorov-Smirnov (KS) statistic relative to negative control. (C) Normal vector of decision plane of linear support-vector machine (SVM) versus negative control. (D) Proportion of cells in each component of a Gaussian mixture (GM). (E) Latent feature extraction using factor analysis. Figure modified from Ljosa *et al.*[47]

Even though there is no example of mode-of-action identification of an uncharacterized compound given in this study, it successfully merges complex imaging data with additional databases in order to predict mechanisms of action. In a study by Slack *et al.*, cellular profiles were generated based on cellular subpopulations.[45] A total of 35 different compounds were screened. Ten of these had either miscellaneous or undefined modes of action (e.g. green tea polyphenols, valproic acid). For each cell acquired, a 1,536-dimensional feature vector was computed and subsequently reduced to 25 dimensions by PCA. Based on the principal components, subpopulations of cells were identified by application of a Gaussian mixture model (see Table S2, Figure 9 D). The authors found that drugs of similar mechanism often yield similar subpopulation profiles. Interestingly, analysing a higher number of subpopulations (> 5) did not necessarily allow for a more accurate classification of compounds. Green-tea phenols and valproic acid both clustered with glucocorticoid (GC) receptor agonists. Additional biochemical experiments confirmed that a GFP-tagged GC-receptor is internalized by these compounds. However, two other compounds that clustered with GC-receptor agonists did not induce positive results in the biochemical assay, implying that the classification of compound treatments into mechanisms of action by the Gaussian Mixture Model is susceptible to false positives. In fact, the accuracy of this model was calculated to be 83% compared to the abovementioned factor analysis model with 94% accuracy.[47]

Caie *et al.* 2010 correlated phenotypic drug response with several cancer cell types of different genetic background.[49] A library of well-characterized drugs was investigated and HCA was run in a four-wavelength assay with four different cancer cell lines. After segmentation of imaged cells by identifying nuclear and cytoplasmic boundaries, 100 features were extracted. The multiparametric phenotypic response was then simplified by PCA. Compounds inducing distinct phenotypes compared to the control cells were classified by calculating the multidimensional Mahalanobis distance. To further compare the phenotypic responses across the different cancer cell types, a Kohonen neural network (self-organizing map) was calculated (Table S2). The resulting map visualized the phenotypic data for each compound across dose response and the four cell types used. It was found that some compounds, for example the microtubule stabilizer epothilone B, induced similar phenotypes across all cell lines tested. In case of the translation inhibitor emetine, phenotypic responses of the cell lines cluster differently, indicating highly sensitive, cell-specific responses against this particular drug. It was speculated that p53 is important for emetine activity, as the phenotypic profile of MCF7-p53 was significantly different to MCF7-wt, thus pointing to p53 as a mediator of emetine function. The authors then performed a k-nearest neighbour classification to make a prediction of a particular compound's mode of action. The analysis revealed that the different compound classes clustered well together in MCF7-wt, MCF7-p53 and MiaPaCa2 cells, providing proof of concept. In OvCar3 cells, however,

mechanistically unrelated compounds were ranked as nearest neighbours, *e.g.* proteasome inhibitor 1 was closely clustered with kinase inhibitor PP2 and protein synthesis inhibitors anisomycin and cycloheximide. This illustrates that the manifestation of a certain cellular profile can be dependent on the cell type analysed.

Perlman *et al.* developed a cytological profiling approach comprising multidimensional measurements of cells treated with a wide concentration range of a reference drug set, selected to cover common mechanisms of toxicity or therapeutic action.[9] One hundred compounds, including many natural products, were screened and a sum of 93 descriptors were extracted from stained cells. For generating compound profiles, the descriptors were plotted as a cumulative distribution and then reduced to a single number that represented the point of maximum difference between the control and treated population (Figure 9 B). Heat plots were generated for all reduced descriptors at different compound concentrations. For 61 of the 100 compounds, a strong response was obtained by this analysis. Structurally unrelated compounds sharing a common target showed similar response profiles. By applying a titration-invariant similarity score (TISS) the authors compared dose-response profiles obtained from analyses of different starting dosages. Unsupervised clustering of compound profiles by their TISS value revealed that compounds with similar targets can be successfully clustered together. For the subset of kinase inhibitors no clustering was observed even in case of overlapping targets, maybe due to a variable inhibition of other kinases. The authors also included three poorly characterized compounds in their profiling approach. One of these, austocystin, clustered together with transcription and translation inhibitors. According to unpublished preliminary data, the authors were able to verify inhibition of transcription of this compound *in vitro*. In this case, HCA correctly assigned a compound to a mode of action class. Of note, austocystin D was later on shown to be activated by CYP-enzymes and to induce DNA damage.[50]

The image set of Perlman *et al.* was reanalysed in a study by Loo *et al.*. They aimed at providing a multivariate method for classification of single cells in order to obtain better profiling accuracies.[44] Based on more than 300 descriptors the cells were displayed in the high-dimensional descriptor space. A support-vector machine (SVM) determined the optimum hyperplane that separated control from treated cells (Figure 9 C). By applying a SVM recursive feature elimination algorithm, 20 - 40 features were identified to be essential for the classification of most of the compounds. Similar normal vectors in a concentration series were clustered to yield a representative dosage (d)-profile. Significant d-profiles were then used for category prediction and the authors observed better classification accuracy, as also the kinase inhibitors grouped together in their analysis.

A profiling study by Tanaka *et al.* draws on simple comparison of means for each descriptor in order to construct cellular profiles (Table S2, Figure 9 A). They found that the compound hydroxy-PP induced a distinct phenotype that did not correlate with that of structurally related kinase inhibitors and micro-
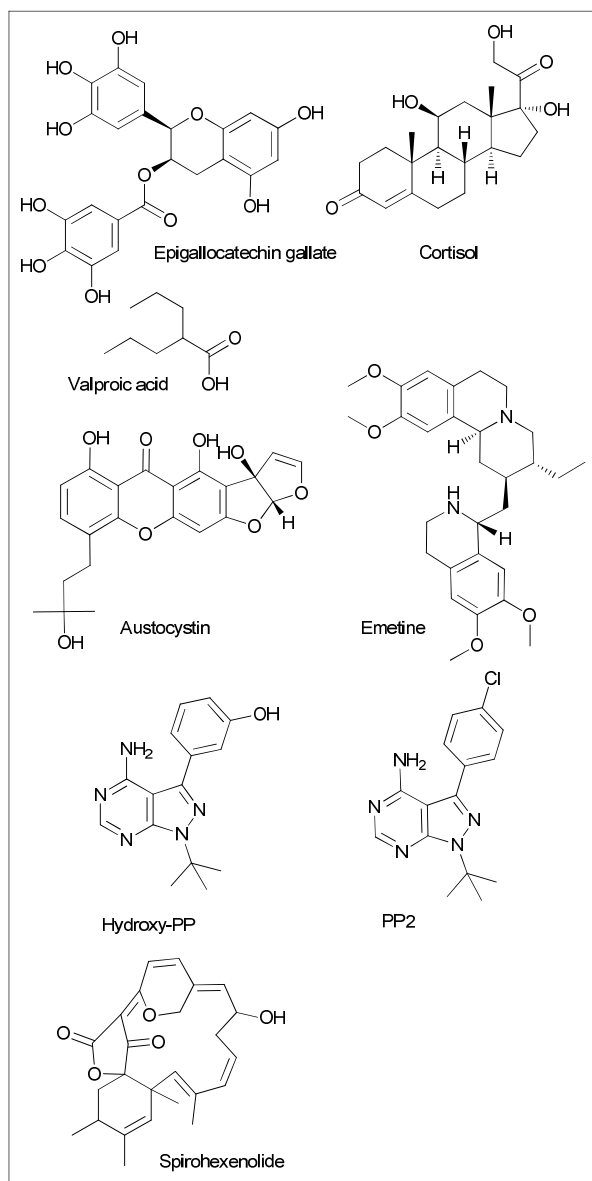
Scheme 2. Compounds for which image based profiling was applied to identify their MoA (left side). Respective reference compounds with similar phenotypic effect and cellular target are depicted on the right side.

tubule polymerization inhibitors.[51] It was not possible to assign hydroxy-PP a certain mode of action by comparison to reference compounds, thus it was postulated that hydroxy-PP must exert a different mode of action. The cellular target was then identified as carbonyl reductase 1 through a chemical pull down with immobilized hydroxy-PP. In this particular case, HCA gave the hint for a novel mode of action.

Phenotyping of prokaryotic cells

All of the above mentioned profiling approaches phenotyped eukaryotic cells. For microscopic imaging of prokaryotic cells a particular challenge is given by their comparably small size and potential movement due to flagella. In a first in field study

performed by Nonejuie et al., the phenotypic effects of inhibitors targeting the five major pathways namely translation, transcription, DNA replication, lipid synthesis and peptidoglycan synthesis, were evaluated.[52] Images of *E. coli* cells immobilized on agarose pads were manually acquired using an inverted microscope with a 100× objective. Hereafter images were modified and evaluated on single image basis by image editing and analysis programs. Eventually cellular profiles were generated by extracting 14 features from the edited images. Following PCA different inhibitor classes were successfully separated from each other and characterized correctly. As a proof of concept spirohexenolide A, a natural product with activity against Gram-positive bacteria and *E. coli* lptD4213 and formerly unknown MoA, was shown to possess a similar profile as the antibacterial peptide nisin. Further experiments validated that spirohexenolide A compromised membrane integrity and the proton motive force, as known for nisin. The study provides a proof-of-concept for the validity of phenotypic profiling of bacterial cells.

Perspectives

A major bottleneck in the exploration of natural products as tools for chemical biology research and as lead structures for therapeutic applications is the identification of their mode of action on a molecular and cellular level. As the initial bioactivity of natural products has been often discovered in phenotypic assays (like growth inhibition of bacterial or eukaryotic cells) that do not capture target information, a mode of action assignment is particularly relevant for natural products compared to other sources of bioactive compounds. Assignments based on image analysis have been successfully applied in multiple cases, as outlined in this review. Most application examples involve a visual inspection of images, while the use of automated HCA is still limited, maybe due to its technical complexity and the limited number of labs with fully established HCA workflows.

The central assumption of target identification by image analysis is that the modulation of a particular target results in a specific phenotype. However, not all phenotypes may become clearly visible with the applied set of antibodies and descriptors. Provided that no antibody is included that captures a given kind of phenotypic alteration, evaluation of HCA might lead to false negative results. Furthermore if a compound has an effect merely on cellular metabolism or a signalling pathway without inducing a visible morphological change, immunostainings may not applicable for target identification. In this case the use of physiological indicator dyes can be taken into consideration. Yet, the repertoire of these is limited and only few specific processes can be monitored, e.g., Ca^{2+}-distribution by Ca^{2+}-sensitive fura-dyes. Eventually, the applied concentration of compounds to be screened is a critical parameter that has to be carefully considered. Too high concentrations may induce apoptosis, thus masking any specific morphological change occurring at lower concentrations. This may lead to a false positive classification, e.g., with general apoptosis inducers. On the contrary, too low concentrations may give false negative results, as the characteristic phenotypic effects are not yet induced. Hence target prediction via cellular profiling can easily become ambiguous or lead into the wrong direction. Even so, it has to be pointed out that cellular image profiling is always preliminary and has to be proven by more specific biochemical or biophysical methods.

Nevertheless the potential of HCA itself can be further enhanced: Assaying more complex model systems, such as 3D cell cultures or even whole tissues by HCA will further increase the impact of cellular profiling in the course of drug discovery, as they better resemble the physiological state in a living system. Promising advances in this direction have been made in the quantification of tumor model phenotypes across whole tissues.[53] In terms of target identification it may be advantageous to use an 'easy to handle' model system for profiling and apply more complex model systems for final validation.

An emerging alternative method for cellular profiling by automated microscopes could be imaging flow cytometry (IFC). IFC represents a combination of flow cytometry with microscopic imaging and allows for analysis of 300-1000 events per second, albeit at lower spatial resolution compared to microscopic analysis.[54]

So far only data from fixed endpoints assays have been used for cellular profiling. Yet, modern instrumentation already permits kinetic measurements that capture the dynamics of cells and biological processes over time, thereby adding a significant dimension to cellular profiling analysis.[55] Hence, automated live cell imaging has recently been regarded as an important trend.[56] Indeed, live cell HCA combined with RNA-interference techniques has successfully been applied for profiling gene knockdowns on basis of the induced cellular phenotypes.[57]

Also for prokaryotic cells HCA-based profiling could be of interest in order to classify the mode of actions of novel antibiotics. One disadvantage to overcome so far is the high amount of manual work that is necessary for bacterial profiling. Until now microscopic analysis of prokaryotic cells in microtiter plates is not possible due to the minor size and movement of living bacteria. They can only be imaged at high resolution and by applying techniques that prevent motility.

Overall, we see a high potential in HCA for accelerating the target identification process, in particular when the method is combined with orthogonal target identification techniques: In such a workflow, HCA could be utilized to quickly identify a compound's cellular target if this belongs to a known mode of action class. The compound can then be directed to specific target-based biochemical and/or biophysical assays to verify the hypothesis generated by HCA. For compound profiles that cannot be matched, direct identification techniques like genetic screens or pulldown probes need to be applied. Such a systematic approach may render the assignment of a biological profile for natural products more efficient and increase their value for life sciences significantly.

References

1. G. C. Terstappen, C. Schlüpen, R. Raggiaschi and G. Gaviraghi, *Nat. Rev. Drug Discov.*, 2007, **6**, 891-903.
2. E. E. Carlson, *ACS Chem. Biol.*, 2010, **5**, 639-653.
3. D. C. Swinney and J. Anthony, *Nat. Rev. Drug Discov.*, 2011, **10**, 507-519.
4. S. Ziegler, V. Pries, C. Hedberg and H. Waldmann, *Angew. Chem. Int. Edit.*, 2013, **52**, 2744-2792.
5. B. Lomenick, R. W. Olsen and J. Huang, *ACS Chem. Biol.*, 2011, **6**, 34-46.
6. J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, E. S. Lander and T. R. Golub, *Science*, 2006, **313**, 1929-1935.
7. T. Böttcher, M. Pitscheider and S. A. Sieber, *Angew. Chem. Int. Edit.*, 2010, **49**, 2680-2698.
8. R. H. Shoemaker, *Nat. Rev. Cancer*, 2006, **6**, 813-823.
9. Z. E. Perlman, M. D. Slack, Y. Feng, T. J. Mitchison, L. F. Wu and S. J. Altschuler, *Science*, 2004, **306**, 1194-1198.
10. Y. A. Abassi, B. Xi, W. Zhang, P. Ye, S. L. Kirstein, M. R. Gaylord, S. C. Feinstein, X. Wang and X. Xu, *Chem. Biol.*, 2009, **16**, 712-723.
11. J.-M. Fritschy and W. Härtig, in *eLS*, John Wiley & Sons, Ltd, 2001, DOI: 10.1038/npg.els.0001174.
12. G. Thierbach and H. Reichenbach, *BBA - Bioenergetics*, 1981, **638**, 282-289.
13. V. V. Vintonyak, B. Kunze, F. Sasse and M. E. Maier, *Chem.-Eur. J.*, 2008, **14**, 11132-11140.
14. C. M. Woods, J. Zhu, P. A. McQueney, D. Bollag and E. Lazarides, *Mol. Med.*, 1995, **1**, 506-526.
15. S. Forli, *Curr. Top. Med. Chem.*, 2014, **14**, 2312-2321.
16. H. Irschik, R. Jansen, K. Gerth, G. Höfle and H. Reichenbach, *J. Antibiotics*, 1995, **48**, 31-35.
17. Y. A. Elnakady, F. Sasse, H. Lünsdorf and H. Reichenbach, *Biochem. Pharmacol.*, 2004, **67**, 927-935.
18. F. Sasse, B. Kunze, T. M. A. Gronewold and H. Reichenbach, *J. Nat. Cancer Inst.*, 1998, **90**, 1559-1563.
19. R. Diestel, H. Irschik, R. Jansen, M. W. Khalil, H. Reichenbach and F. Sasse, *ChemBioChem*, 2009, **10**, 2900-2903.
20. H. Brötz-Oesterhelt, D. Beyer, H. P. Kroll, R. Endermann, C. Ladel, W. Schroeder, B. Hinzen, S. Raddatz, H. Paulsen, K. Henninger, J. E. Bandow, H. G. Sahl and H. Labischinski, *Nat.Med.*, 2005, **11**, 1082-1087.
21. H. Noufflard-Guy-Loe and S. Berteaux, *Rev. Tuberc. Pneumol.*, 1965, **29**, 301-326.
22. A. Kling, P. Lukat, D. V. Almeida, A. Bauer, E. Fontaine, S. Sordello, N. Zaburannyi, J. Herrmann, S. C. Wenzel, C. Konig, N. C. Ammerman, M. B. Barrio, K. Borchers, F. Bordon-Pallier, M. Bronstrup, G. Courtemanche, M. Gerlitz, M. Geslin, P. Hammann, D. W. Heinz, H. Hoffmann, S. Klieber, M. Kohlmann, M. Kurz, C. Lair, H. Matter, E. Nuermberger, S. Tyagi, L. Fraisse, J. H. Grosset, S. Lagrange and R. Muller, *Science*, 2015, **348**, 1106-1112.
23. A. J. North, *J. Cell Biol.*, 2006, **172**, 9-18.
24. G. Lukinavicius, L. Reymond, E. D'Este, A. Masharina, F. Gottfert, H. Ta, A. Guther, M. Fournier, S. Rizzo, H. Waldmann, C. Blaukopf, C. Sommer, D. W. Gerlich, H. D. Arndt, S. W. Hell and K. Johnsson, *Nat. Methods*, 2014, **11**, 731-733.
25. M. Borowiak, W. Nahaboo, M. Reynders, K. Nekolla, P. Jalinot, J. Hasserodt, M. Rehberg, M. Delattre, S. Zahler, A. Vollmar, D. Trauner and O. Thorn-Seshold, *Cell*, 2015, **162**, 403-411.
26. F. Mollinedo and C. Gajate, *Apoptosis*, 2003, **8**, 413-450.
27. F. Sasse, H. Steinmetz, J. Heil, G. Höfle and H. Reichenbach, *J. Antibiotics*, 2000, **53**, 879-885.
28. M. W. Khalil, F. Sasse, H. Lünsdorf, Y. A. Elnakady and H. Reichenbach, *ChemBioChem*, 2006, **7**, 678-683.
29. D. M. Bollag, P. A. McQueney, J. Zhu, O. Hensens, L. Koupal, J. Liesch, M. Goetz, E. Lazarides and C. M. Woods, *Cancer Res.*, 1995, **55**, 2325-2333.
30. A. L. Risinger, S. M. Riffle, M. Lopus, M. A. Jordan, L. Wilson and S. L. Mooberry, *Mol. Cancer*, 2014, **13**, 41.
31. T. U. Mayer, T. M. Kapoor, S. J. Haggarty, R. W. King, S. L. Schreiber and T. J. Mitchison, *Science*, 1999, **286**, 971-974.
32. Y. Yang, B. Liu, J. Dai, P. K. Srivastava, D. J. Zammit, L. Lefrançois and Z. Li, *Immunity*, 2007, **26**, 215-226.
33. M. Huss, F. Sasse, B. Kunze, R. Jansen, H. Steinmetz, G. Ingenhorst, A. Zeeck and H. Wieczorek, *BMC Biochem.*, 2005, **6**.
34. B. Kunze, F. Sasse, H. Wieczorek and M. Huss, *FEBS Letters*, 2007, **581**, 3523-3527.
35. K. A. Giuliano, J. R. Haskins and D. L. Taylor, *Assay Drug Dev. Techn.*, 2003, **1**, 565-577.
36. D. L. Taylor, E. S. Woo and K. A. Giuliano, *Curr. Opin. Biotech.*, 2001, **12**, 75-81.
37. K. A. Giuliano, R. L. DeBiasio, R. T. Dunlay, A. Gough, J. M. Volosky, J. Zock, G. N. Pavlakis and D. L. Taylor, *J. Biomol. Screen.*, 1997, **2**, 249-259.
38. W. Buchser, M. Collins, T. Garyantes, R. Guha, S. Haney, V. Lemmon, Z. Li and O. J. Trask, in *Assay Guidance Manual* ed. C. M. Buchser W, Garyantes T, et al. , Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD), 2012.
39. F. Zanella, J. B. Lorens and W. Link, *Trends Biotechnol.*, 2010, **28**, 237-245.
40. D. L. Taylor, *Methods Mol. Biol. (Clifton, N.J.)*, 2007, **356**, 3-18.
41. C. L. Adams, V. Kutsyy, D. A. Coleman, G. Cong, A. M. Crompton, K. A. Elias, D. R. Oestreicher, J. K. Trautman and E. Vaisberg, *Methods Enzymol.*, 2006, **414**, 440-468.
42. A. Shariff, J. Kangas, L. P. Coelho, S. Quinn and R. F. Murphy, *J. Biomol. Screen.*, 2010, **15**, 726-734.
43. W. Buchser, M. Collins, T. Garyantes, R. Guha, S. Haney, V. Lemmon, Z. Li and O. J. Trask, in *Assay Guidance Manual*, eds. G. S. Sittampalam, N. P. Coussens, H. Nelson, M. Arkin, D. Auld, C. Austin, B. Bejcek, M. Glicksman, J. Inglese, P. W. Iversen, Z. Li, J. McGee, O. McManus, L. Minor, A. Napper, J. M. Peltier, T. Riss, O. J. Trask, Jr. and J. Weidner, Bethesda (MD), 2004.
44. L. H. Loo, L. F. Wu and S. J. Altschuler, *Nat. Methods*, 2007, **4**, 445-453.
45. M. D. Slack, E. D. Martinez, L. F. Wu and S. J. Altschuler, *P. Natl. Acad. Sci.USA*, 2008, **105**, 19306-19311.
46. G. J. Székely and M. L. Rizzo, *J. Classif.*, 2005, **22**, 151-183.
47. V. Ljosa, P. D. Caie, K. L. Sokolnicki, E. L. Jenkins, S. Daya, M. E. Roberts, T. R. Jones, S. Singh, A. Genovesio, P. A. Clemons, N. O. Carragher and A. E. Carpenter, *J. Biomol. Screen.*, 2013, **18**, 1321-1329.
48. D. W. Young, A. Bender, J. Hoyt, E. McWhinnie, G. W. Chirn, C. Y. Tao, J. A. Tallarico, M. Labow, J. L. Jenkins, T. J. Mitchison and Y. Feng, *Nat. Chem. Biol.*, 2008, **4**, 59-68.
49. P. D. Caie, R. E. Walls, A. Ingleston-Orme, S. Daya, T. Houslay, R. Eagle, M. E. Roberts and N. O. Carragher, *Mol. Cancer Ther.*, 2010, **9**, 1913-1926.
50. K. M. Marks, E. S. Park, A. Arefolov, K. Russo, K. Ishihara, J. E. Ring, J. Clardy, A. S. Clarke and H. E. Pelish, *J. Nat. Prod.*, 2011, **74**, 567-573.
51. M. Tanaka, R. Bateman, D. Rauh, E. Vaisberg, S. Ramachandani, C. Zhang, K. C. Hansen, A. L. Burlingame, J. K. Trautman, K. M. Shokat and C. L. Adams, *PLoS Biol.*, 2005, **3**, e128.
52. P. Nonejuie, M. Burkart, K. Pogliano and J. Pogliano, *P. Natl. Acad. Sci. USA*, 2013, **110**, 16169-16174.
53. B. L. Falcon, J. Stewart, S. Ezell, J. Hanson, J. Wijsman, X. Ye, E. Westin, G. Donoho, K. Credille and M. T. Uhlik, *Drug Discov. Today*, 2013, **18**, 510-522.
54. N. S. Barteneva, E. Fasler-Kan and I. A. Vorobjev, *J. Histochem. Cytochem.*, 2012, **60**, 723-733.
55. D. L. Taylor, *Methods Mol. Biol.*, 2007, **356**, 3-18.
56. R. Pepperkok and J. Ellenberg, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 690-696.
57. B. Neumann, M. Held, U. Liebel, H. Erfle, P. Rogers, R. Pepperkok and J. Ellenberg, *Nat. Methods*, 2006, **3**, 385-390.
58. B. Liu, S. Li and J. Hu, *Am. J. Pharmacogenomics*, 2004, **4**, 263-276.
59. Y. Feng, T. J. Mitchison, A. Bender, D. W. Young and J. A. Tallarico, *Nat. Rev. Drug Discov.*, 2009, **8**, 567-578.
60. J. V. Peppard, C. Rugg, M. Smicker, C. Dureuil, b. Ronan, O. Flamand, L. Durand and B. Pasquier, *Curr. Chem. Genom. Transl. Med.*, 2014, **8**, 3-15.