Discovering Antibiotic Efficacy Biomarkers

TOWARD MECHANISM-SPECIFIC HIGH CONTENT COMPOUND SCREENING

Christoph Freiberg‡§**, Nina Brunner‡§‡‡, Ludwig Macko¶, and Hans Peter Fischer¶

As current antibiotic therapy is increasingly challenged by emerging drug-resistant bacteria, new technologies are required to identify and develop novel classes of antibiotics. A major bottleneck in today's discovery efforts, however, is a lack of an efficient and standardized method for assaying the efficacy of a drug candidate. We propose a new high content screening approach for identifying efficacious molecules suitable for development of antibiotics. Key to our approach is a new microarray-based efficacy biomarker discovery strategy. We first produced a large dataset of transcriptional responses of Bacillus subtilis to numerous structurally diverse antibacterial drugs. Second we evaluated different protocols to optimize drug concentration and exposure time selection for profiling compounds of unknown mechanism. Finally we identified a surprisingly low number of gene transcripts (~130) that were sufficient for identifying the mechanism of novel substances with reasonable accuracy (~90%). We show that the statistics-based approach reveals a physiologically meaningful set of biomarkers that can be related to major bacterial defense mechanisms against antibiotics. We provide statistical evidence that a parallel measurement of the expression of the biomarkers guarantees optimal performance when using expression systems for screening libraries of novel substances. The general approach is also applicable to drug discovery for medical indications other than infectious diseases. Molecular & Cellular Proteomics 5:2326–2335, 2006.

In clinical practice, biomarkers are widely used for diagnosis, for disease prognosis, and for optimizing therapeutic strategies. Lately the use of biomarkers showed potential for guiding the discovery and development of novel therapeutics. For instance, the strategy sometimes referred to as toxicogenomics is a biomarker-based approach for evaluating the toxicity potential of a drug candidate by its characteristic biomarker expression profile (1, 2). It has been shown that the mRNA expression profiles of toxicity biomarkers triggered by a toxicant correlate well with the underlying toxicity mechanism of the toxicant (3). However, so far biomarkers are mostly used to predict adverse side effects of drug candidates. To date, there is little evidence that an analogous “efficacy biomarker approach” could help in evaluating the efficacy potential of a drug candidate (4). A biomarker-based compound efficacy evaluation would be very attractive for many applications in the pharmaceutical industry as it would enable a standardized and automated approach to the drug discovery process.

Here we propose a biomarker-based strategy for evaluating the efficacy of compounds with antibacterial activity. To systematically investigate the influence of currently known antibacterial compounds on bacterial gene expression, we produced a large antibiotic-response microarray dataset. We measured the whole-genome transcription response of Bacillus subtilis, a model organism for phylogenetically related major Gram-positive pathogens such as staphylococci, streptococci, and enterococci, to 112 different chemicals with antibacterial activity (see Supplemental Table S1). These antibiotic agents represent the majority of established and commercialized antibiotics such as β-lactams, macrolides, and quinolones. In addition, we included a number of developmental compounds as well as unspecifically acting substances with antibacterial activity such as DNA intercalators. For each of these compounds, we monitored the expression response of B. subtilis to different drug concentrations and exposure times. Here we present detailed results of the B. subtilis response to 46 different antibiotics.

To define objective decision rules applicable to the problem of predicting the mechanism of action (MOA) of novel substances in compound library screens, we used supervised classification algorithms. These algorithms require a training set of mRNA profiles to define a separation function (“to train the classifier”) to assign a compound to its antibiotic MOA category solely based on the mRNA profile it triggers in B. subtilis. In this study, we focused on support vector machines (SVMs) as this algorithm class has been shown to be well suited for categorizing microarray data (5). Using our mRNA profile reference database, we investigated several aspects of microarray-based MOA classifications. First we analyzed the effect of the experimental design on the MOA prediction performance. Second we identified gene sets whose expression carries most of the information that is indicative for characterizing the MOAs of the compound. In addition, we

© 2006 by The American Society for Biochemistry and Molecular Biology, Inc.

This paper is available on line at http://www.mcponline.org

From ‡Pharma Global Drug Discovery, European Research Center, Bayer HealthCare AG, D-42096 Wuppertal, Germany and ¶Genedata AG, Postfach 254, CH-4016 Basel, Switzerland

Received, April 10, 2006, and in revised form, August 29, 2006

Published, MCP Papers in Press, August 29, 2006, DOI 10.1074/mcp.M600127-MCP200

1 The abbreviations used are: MOA, mechanism of action; MIC, minimum inhibitory concentration; SVM, support vector machine; MCR, misclassification rate; ANOVA, analysis of variance.
present an in depth discussion of the biological context of these “antibiotic efficacy biomarkers.” Last we discuss the biology underlying the biomarker concept by exemplifying the response of *B. subtilis* to two agents that both block DNA replication but via different molecular mechanisms. We show that the analysis of pathway activity patterns can pinpoint mechanistic details of novel substances.

The central goal of our work was the identification of efficacy biomarkers whose compound-induced expression provides sufficient information about the underlying mechanisms of the compound. From a technical standpoint, this is critical information for developing robust assays to routinely characterize novel antibiotics. According to our knowledge, there are no published antibiotic efficacy biomarkers available to date that could be used for routinely assaying the mechanism of novel antibiotic substances. In addition, the implementation of MOA-specific screening processes is currently hampered by a lack of standard experiment protocols that could be used for compound library screens. Our work represents also in this regard a first step toward implementing novel expression-based high content screening strategies.

**EXPERIMENTAL PROCEDURES**

Transcription Profiling Experiments—*B. subtilis* 168 cultures were grown in Belitzky minimal medium (6) to $A_{600} = 0.4$ at 37°C. After removal of an aliquot ($t = 0$ min), the remaining untreated culture was treated with antibiotics and incubated further, $t_1 = 10$ and $t_2 = 40$ min. Each compound was examined in two concentrations, $C_{t1} = 2 \times$ MIC and $C_t = 1/2 \times$ MIC (see Fig. 1). In case that the applied concentration already caused more than 25% growth reduction compared with the control at $t = t_2$, the maximal concentration was reduced to 1-fold MIC or fractions of the MIC (see Supplemental Table S1). Details of sample preparation have been published previously (7). In brief, experimental procedures followed the protocol for Eurogentec microarrays based on a two-color fluorescence technology (Eurogentec, Seraing, Belgium). The fluorescence intensities of the two dyes were detected via the Axon GenePix 4000A confocal laser scanner using the image analysis software GenePix Pro (Axon Instruments, Foster City, CA). To minimize the influence of technical artifacts, we split each cDNA sample related to a drug concentration and exposure time and hybridized it separately to three different microarrays. For details regarding the technical and biological reproducibility, see Fischer et al. (7).

Microarray Analysis—The signal intensities obtained from the microarray experiments were further processed by Genedata Expressionist Pro 2.0 (Genedata, Basel, Switzerland) using the default settings. The signals of all features were used for a global microarray normalization, with the scaling factor being calculated, requiring the sum of all logarithmized values in the Cy3 channel to equal all logarithmized values in the Cy5 channel. Average-log fold factors of replicate features were calculated following the data processing strategy described previously (7).

Performance Evaluation of MOA Assignments Using Biomarker Expression—The performance of classifiers such as SVMs is typically assessed by so-called “leave-one-out” approaches or related strategies (5, 8). For this, the experiment dataset is split into the (trivial) set of one randomly chosen experiment to be classified and a training set of the remaining ($N - 1$) experiments. Here we applied the leave-one-out strategy by randomly separating the experiment set into $X$ experiments to be classified in a blinded manner and $N - X$ experiments remaining in the training set using a test set fraction $P = X/N = 25\%$. To calculate the average misclassification rates (MCRs), we compared the SVM-predicted results with the known compound MOAs and averaged the misclassification ratios over $r = 100$ independent runs. The MCR was calculated for different sets of marker genes defined by ANOVA $p$ value thresholds using the MOA classes listed in Table I. Figs. 2 and 3 show these dependences for various experimental conditions.

**Functional Annotation of the B. subtilis Gene Products**—The whole-genome annotation of *B. subtilis* was taken from the Genedata Phylosopher® system. Additionally we compiled and incorporated information from specialized publications. In this work, the following categories were considered: relA-dependent stringent response, FapR-dependent fatty acid and phospholipid biosynthesis genes, and the SOS-, CodY-, and alternative RNA polymerase sigma factor-dependent regulons such as sigB, sigH, sigF, sigE, sigG, sigK, sigW, sigM, sigX, and sigY (for details and references see Freiberg et al. (9)). The basis for pathway categorizations are the Genedata Phylosopher annotations. The full biomarker annotations are compiled in Supplemental Table S2.

**Statistical Significance of Pathway Over-representation**—We calculated the over-representation of biomarkers assigned to specific pathways or regimens using Fisher’s exact test. The frequency of each functional category in the biomarker set was compared with the frequency in the “background set” corresponding to the whole set of genes encoded by the *B. subtilis* genome. As a background, we used the annotation information for all ~4,100 *B. subtilis* genes. More specifically the probability of observing $i$ biomarkers that are predicted to be related to a specific functional class $Y$ in a biomarker set consisting of $n$ genes (in our example is $n = 181$) is given by $P_{F}(Y) = \frac{\binom{n}{i} / \binom{g}{f}}{\binom{G}{f}}$ where $f$ is the total number of genes being associated to the functional class $Y$ and $g$ is the total number of genes encoded by *B. subtilis*. The pathway over-representation calculations were performed using the Genedata Expressionist Pro 2.0 system.

**RESULTS**

**Generation of a Compendium of Antibiotics-induced Expression Profiles**—A whole-genome, two-channel microarray technology was used to monitor the transcriptional response of virtually all ~4,100 *B. subtilis* genes (10). In previous studies, we showed that this technology enables the parallel measurement of the transcriptional induction or repression of all genes when exposing bacterial cultures to substances with antibacterial potential (7, 9, 11). For the purpose of this study, the transcriptional response to a set of 36 well investigated antibiotics and 10 unspecifically acting substances were examined. The antibiotic compounds represent six different MOA classes (see Table I) broadly categorized as inhibitors of protein biosynthesis (“PROTEIN”), aminoacyl-tRNA synthesis (“ATS”), cell wall biosynthesis (“CELL WALL”), fatty acid biosynthesis (“FAB”), RNA synthesis (“RNA”), and replication inhibitors (“DNA”) (12).

Antibiotic dosage and treatment time are known to be of critical importance for gene expression (13), hence we suspected that these parameters will also play a role when using microarrays as highly parallel assay systems. Thus, in contrast to previous antibiotic microarray studies (14), it was necessary to systematically measure dosage and exposure time effects. We assayed each compound according to a
standard protocol based on objective growth curve characteristics. Each bacterial culture was split and treated with two different drug concentrations, a “high” (H) and a “low” (L) concentration. For each culture treated, mRNA was harvested after \( t_1 = 10 \) min and \( t_2 = 40 \) min exposure time, respectively (see Fig. 1 and “Experimental Procedures”). Exposure times \( t_1 \) and \( t_2 \) (“1” and “2”) were chosen with respect to the typical cell division time of \( B. subtilis \) in Belitzky minimal medium \( (t_2 = t_\text{replication} \approx 40 \text{ min}) \), to resolve antibiotic stress effects occurring on two biologically relevant time scales. The combination of different dosages and exposure times for a wide range of antibiotic compounds allowed us to obtain an unbiased view on a wide spectrum of antibiotics-induced stress states. In total, the results presented in this

| Code | Compound name | MOA: targeted pathway (protein) |
|------|---------------|---------------------------------|
| DNA replication inhibitors (DNA) | | |
| Az | Azaserine | Purine synthesis/DNA carboxymethylation |
| Fl | Fluorouracil | Pyrimidine synthesis |
| N | Novobiocin | DNA replication (gyrase-topoisomerase IV) |
| Na | Nalidixic acid | DNA replication (gyrase-topoisomerase IV) |
| P | Ciprofloxacin | DNA replication (gyrase-topoisomerase IV) |
| Su | Sulfamethoxazole | Nucleotide biosynthesis (folate biosynthesis) |
| T | Trimethoprim | Nucleotide biosynthesis (folate biosynthesis) |
| Tr | Trovafloxacin | DNA replication (gyrase-topoisomerase IV) |
| X | Moxifloxacin | DNA replication (gyrase-topoisomerase IV) |

| RNA synthesis inhibitors (RNA) | | |
| Co | Corallopyronin | RNA polymerase |
| Ri | Rifampicin | RNA polymerase |
| Sv | Streptavidin | RNA polymerase |

| Aminoacyl-tRNA synthesis inhibitors (AATS) | | |
| Phe | Phenyl-thiazolylurea derivative | Phenylalanyl-tRNA synthetase |
| U | Mupirocin | Isoleucyl-tRNA synthetase |

| Protein biosynthesis inhibitors (PROTEIN) | | |
| A | Actinonin | Peptide deformylase |
| Do | Doxycycline | Translation |
| E | Erythromycin | Translation |
| Fu | Fusidic acid | Translation (elongation factor G) |
| G | Gentamicin | Translation |
| H | Chloramphenicol | Translation |
| K | Kanamycin | Translation |
| Li | Linezolid | Translation |
| Pa | Pactamycin | Translation |
| Pu | Puromycin | Translation |
| Tan | TAN 1057 | Translation |
| Y | Azithromycin | Translation |

| Fatty acid biosynthesis inhibitors (FAB) | | |
| B | Moiramide B | Fatty acid biosynthesis (acetyl-CoA carboxylase) |
| C | Cerulenin | Fatty acid biosynthesis (\( \beta \)-ketoacyl-acyl carrier protein synthetase) |
| I | Triclosan | Fatty acid biosynthesis (enoyl-acyl carrier protein reductase) |

| Cell wall biosynthesis inhibitors (CELL WALL) | | |
| F | Cefoxitin | Cell wall biosynthesis |
| M | Methicillin | Cell wall biosynthesis |
| Me | Mersacidin | Cell wall biosynthesis |
| O | Oxacillin | Cell wall biosynthesis |
| R | Ristocetin | Cell wall biosynthesis |
| Ra | Ramoplanin | Cell wall biosynthesis |
| V | Vancomycin | Cell wall biosynthesis |

\( ^a \text{A priori MOA categorization. MOA categories: inhibition of protein biosynthesis (PROTEIN), aminocytot-IRNA synthesis (AATS), cell wall biosynthesis (CELL WALL), fatty acid biosynthesis (FAB), RNA synthesis (RNA), and DNA replication (DNA). Antibiotics acting via unspecific effects are DNA-intercalating or -binding agents actinomycin D, ethidium bromide, and netropsin; the pyrimidine synthesis-inhibiting and protein-alkylating antibiotic showdomycin; the membrane-perturbing agents monensin and polymyxin B; the protein-alkylating or -oxidizing compounds \( N \)-ethylmaleimide and nitrofurantoin; and the cell wall biosynthesis-inhibiting but also directly or indirectly membrane-permeabilizing antibiotics nisin and tunicamycin.} \)
study are based on 309 individual hybridization experiments. Efficacy Biomarker Identification—Obviously not all gene expression values contribute equally to the MOA-specific information of a compound-induced expression profile. Thus, identifying a biomarker set for MOA prediction with minimal false negative and false positive rates is a crucial step toward developing assays for routine compound MOA profiling. A primary goal of this study was to quantitatively evaluate the performance of supervised classification algorithms for predicting the MOA of novel antibiotics when using only the expression values of specific gene subsets as input. Here we focus on linear-kernel SVMs for classifying the MOA of a compound. The performance of the classifier was evaluated using a “leave-X-out” strategy (see “Experimental Procedures”). In brief, we randomly separated the experiment set into X experiments to be classified in a blinded manner and N − X experiments remaining in the training set. By comparing the results of the classifier with the known MOAs of the compounds, we were able to quantify the MOA prediction accuracy of the classifier. To investigate the dependence of the classifier accuracy on different gene sets, we ranked the genes according to their specific expression patterns, here quantified by their ANOVA p value based on the experiment groups defined by the six major a priori MOA categorizations (see Table I). The dependence of the averaged MCRs on the biomarker set size and various experimental conditions are displayed in Figs. 2 and 3.

When pooling all experimental conditions, the MCR is relatively high for smaller biomarker sets (Fig. 2, black solid line) even if the ANOVA p values of the biomarker are highly significant. For instance, for the 10 most discriminating ANOVA genes (all p values <10^-35) the misclassification rate still exceeds 50%. The expression profiles of the most discriminative genes are apparently not sufficient for distinguishing the underlying MOA classes. However, as more biomarkers are taken into account, the average misclassification rate is significantly reduced. The optimal biomarker set was identified for h = 2,896 genes with an MCR of 7.1%, which is only slightly below the MCR when taking into account all 4,100 B. subtilis gene expression profiles (MCR = 7.4%). Although the correct MOA classification rate is reasonably good in this setup (~93%), the number of efficacy biomarker signals that is required for accomplishing the optimal MOA prediction is very large. On the other hand, the experimental design might influence the antibiotic MOA prediction accuracy, a question that we addressed in the next step.

Optimizing the Experimental Design for MOA Screens—We sought to find experimental conditions that would allow us to...
identify the most informative biomarkers and concomitantly to reduce the number of necessary experiments by optimally tuning the experiment conditions. The latter is a critical step toward the definition of a standard protocol that can be used in a high throughput compound screening. Here the question is whether some experimental conditions, i.e. for instance a specific combination of drug concentration and exposure time (e.g. only the L-2 experiments), produce more MOA information-rich mRNA profiles than others.

To investigate the effect of experimental conditions, we first restricted our analysis to only the compound responses related to low antibiotic concentrations (L) and compared the misclassification rates with those calculated based on only the high concentration experiments (H). We applied the same analysis scheme to the experiments related to short and long drug exposure times (see Fig. 2). Finally we calculated MCR curves for specific antibiotic dosage-exposure time combinations (i.e. L-1, L-2, H-1, and H-2; see Fig. 3). As miniaturizing is critical for the development of MOA-diagnostic assays aiming at screening large quantities of drug candidate molecules, it is required to build assays using as few, but sufficiently informative, biomarkers as possible. Fig. 3 clearly shows that the high dosage, long exposure time conditions (H-2) were optimal for minimizing the MCR and the size of the optimal biomarker set at the same time. Apparently the H-2 experiment training set results in an optimal MCR of 13.6%, requiring only 181 biomarker genes (see Fig. 3, arrow, and Supplemental Table S2).

Remarkably our statistical approach for identifying antibiotic efficacy biomarkers reproduced a number of MOA-specific reporter genes that have been proposed earlier. For instance, the activation of the promoter of yorB, a gene that is known to be dependent on the DNA repair-SOS response regulator LexA, has been reported previously to be indicative of quinolone action, an important type of DNA synthesis-replication inhibition (15, 16). Reporter cells using promoters depending on the fatty acid-phospholipid synthesis regulator FabR such as the one of fabF have already been used successfully to screen large compound libraries for novel fatty acid biosynthesis inhibitors (7). Also the expression of the so far functionally uncharacterized gene yvgS, which is included in the efficacy biomarker set, has only recently been proposed to diagnose transcription inhibitors (16). The fact that established key MOA reporter genes are among the n = 181 biomarker set strongly supports the validity of our systematic approach for identifying efficacy biomarkers.

These findings suggest that by defining a specific protocol for the experimental procedure and cell culture preparation the expression of only relatively few genes has to be monitored to obtain a high MOA prediction accuracy. Besides being instrumental for miniaturizing the actual screening assays, this finding is critical for simplifying, standardizing, and automating microarray-based MOA screening processes of novel antibiotic substances.

**Biomarker Annotation: Consolidating Efficacy Biomarker Sets**—It is instructive to have a closer look into the biological meaning of the biomarkers identified solely by statistical means. Note that the statistical approach to MOA screens proposed in this study does not rely on any biological or physiological prior information. To get some insight into the biological context of the biomarkers, we sought to relate the biomarker genes with functional information. For this, we thoroughly annotated the whole *B. subtilis* genome by using automated annotation methods as well as by integrating literature-based information (see “Experimental Procedures”). The annotation comprised several biological levels, including broad physiological functions and pathways as well as transcription regulation processes and regulons, such as sigma factor dependences (see Supplemental Table S2 for a comprehensive list). In fact, approximately two-thirds of the 181 biomarkers could be assigned to known metabolic, signaling, or regulatory pathways, the majority of which have not yet been described as indicative for mechanism-specific expression. Approximately one-third of the biomarkers still await a functional characterization.
Annotating prokaryotic biomarker genes can help in further reducing the number of relevant biomarkers needed for building MOA-diagnostic assays. Several biomarker genes are co-transcribed from the same promoters, *i.e.* they are part of polycistronic operons. Based on our operon annotation, we were able to show that the 181 biomarkers are organized in only 136 operons (see Supplemental Table S2). For instance, each gene of the branched chain amino acid degradation *bkd* operon is a member of the biomarker set *(ptb-bcd-buk-lpd-kdA1-bkdA2-bkdB)* (17). Gene expression within the same operon is typically highly correlated (*e.g.* correlation coefficients for the *bkd* operon genes are in the range of *r* = 0.81–0.95). Thus, from an assay development perspective, additional genes organized in operons can be considered as redundant and may be removed from the biomarker list, resulting in only 136 relevant antibiotic efficacy biomarkers.

**Biological Context Analysis of Antibiotic Efficacy Biomarkers**—A hierarchical clustering of the expression profiles of the biomarker shows the structure of their transcription response to compounds belonging to different MOA categories. Fig. 4 provides an overview of all 181 biomarker expression -fold changes for the relevant H-2 experiments. To biologically...
interpret the biomarker expression clustering, we first investigated the statistical over-representation of distinct functional annotations of the biomarkers against the "background" of the overall distribution of functional categories in B. subtilis. Here, we use a standard technique, the Fisher’s exact test, to statistically quantify the over-representation of distinct annotation terms (see Methods).

The most over-represented regulons correspond to genes known to be involved in the stringent response of B. subtilis (Fisher’s exact test p value, \( P_{\text{FE}} = 4 \times 10^{-15} \)) and biomarkers belonging to the stationary phase regulon \( P_{\text{FE}} = 4 \times 10^{-25} \) for CodY-regulated genes and \( P_{\text{FE}} = 8 \times 10^{-5} \) for sigma factor H-regulated genes, see Fig. 4). It is known that these regulatory units are strongly influenced by two key metabolites of B. subtilis, GTP and branched chain amino acids (18, 19). Although GTP presumably reflects the energy status of the cell, branched chain amino acids are discussed as important factors for overall cellular physiology and are reflecting the stress state of B. subtilis being exposed to specific antibiotics. Their precursors, the branched chain keto acids, are also the precursors of the branched chain fatty acids, which are the major components of the B. subtilis membrane fatty acids (20). In fact, several genes involved in branched chain keto and amino acid metabolism are also part of the biomarker set. Recently the expression of stringent response and stationary phase regulon members has shown them to be typical markers for aminoacyl-tRNA and fatty acid synthesis inhibition (9).

Other pathways that are significantly over-represented in the efficacy biomarker set correspond to regulatory systems dependent on the extracytoplasmic function sigma factors W, X, and M \( P_{\text{FE}} = 1 \times 10^{-2} - 1 \times 10^{-5} \), which are known to respond in a characteristic manner to changes in the cell envelope. Their induction indicates direct inhibition of cell wall biosynthesis and indirect effects on the cell wall by membrane perturbations or fatty acid synthesis inhibition (9, 21, 22). Ribosomal and translation-association genes represent protein biosynthesis inhibition biomarkers, whereas inactivation of general stress response genes of the sigma factor B regulon \( P_{\text{FE}} = 7 \times 10^{-7} \) is maintained by transcription inhibitor treatment. Lastly two pathways connected with nucleotide metabolism, the purine and riboflavin synthesis pathways, are represented by biomarkers whose expression is indicative of DNA synthesis inhibitors.

The statistically over-represented pathways indicate that the major mechanisms of antibiotics can be associated to specific gene sets among the efficacy biomarkers. This annotation analysis represents another independent evidence of the validity of our approach; however, the large number of important although yet functionally characterized biomarkers shows that statistical approaches are indispensable for efficacy biomarker discovery.

**Toward Mechanistic Details: Compound-specific Pathway Activation Patterns—Analyzing biomarker expression patterns in conjunction with the functional context of the biomarker represents a starting point to reveal mechanistic details that can help to differentiate subtypes of antibiotic inhibition within broader MOA classes. Here we focus on DNA replication inhibitors to exemplify the effect of different inhibitors on the B. subtilis purine biosynthetic pathway, a replication inhibition-relevant pathway related to our efficacy biomarker set (see Fig. 5, arrows). Our data showed that some DNA replication inhibitors strongly induce the purine synthesis genes, whereas others repress the transcription of those genes (see Supplemental Table S3). For instance, trimethoprim is known to inhibit folate biosynthesis, which delivers essential building blocks for synthesizing the purines adenosine and guanosine. To investigate the antibiotic-induced effects on the transcription activation pattern of the pathway we projected the expression values resulting from trimethoprim treatment (H-2 condition) onto the purine pathway map (see Fig. 5A). Most striking is the strongly induced core of the purine biosynthesis pathway. This is consistent with the fact that trimethoprim treatment leads to a blocking of purine synthesis, which in turn leads to an enrichment of the precursor metabolite phosphoribosyl pyrophosphate. An excess of this metabolite is known to activate the PurR regulator-dependent purine synthesis genes (23), consistent with our experimental results. In contrast to the purine pathway, some of the connected pathways leading to nucleotide incorporation into DNA and RNA, especially the DNA and RNA polymerases, are down-regulated.

Novobiocin is an antibiotic compound representing another DNA replication inhibitor; however, it works via a very different molecular mechanism by inhibiting the DNA gyrase and thereby leading to an arrest of the replication machinery (24). Remarkably the novobiocin-induced pathway activation pattern is anticorrelated to the trimethoprim-induced pattern (see Fig. 5B). For novobiocin, the DNA and RNA polymerase genes are induced, whereas the purine synthesis pathway is down-regulated. Such differences in pathway activation patterns apparently enable the differentiation between nucleotide biosynthesis inhibitors and agents more directly blocking DNA replication. These results suggest that a combined two-step approach with a first biomarker-based broad screening approach followed by a second, focused pathway activity analysis step may help to systematically identify and characterize the most promising compounds for antibiotic development in large chemical libraries.

**DISCUSSION**

We propose a new microarray-based biomarker discovery approach for a high content screening strategy to predict the MOA of novel substances with antibacterial activity. We show that such an approach is indeed feasible from a conceptual and practical point of view. In addition, we provide evidence that standardization and upscaling of this strategy, requirements for an implementation on an industrial scale, are also
Fig. 5. Transcriptional activity of the purine biosynthesis pathway and connecting pathways when exposing B. subtilis to two chemically distinct DNA replication inhibitors. A shows the response to trimethoprim treatment, while B shows the effect of novobiocin treatment, both for condition H-2 (see text). Arrows indicate the genes that are members of the \( n = 181 \) efficacy biomarker set (see Fig. 4). The \(-\)fold changes of the transcript abundances are color-coded, see color scale (Genedata Phylosopher). Apparently inhibitors of the same MOA class but acting via different molecular mechanisms can induce different (in this case anticorrelated) expression patterns, which can reveal molecular details of the mechanisms of the compounds (see text).
practicable. The high content screening strategy proposed in this work enables a directed development of screening assays for identifying the MOA of uncharacterized compounds. As both natural as well as synthetic product libraries are still very difficult to characterize in a systematic manner, such a biomarker-based screening strategy would be extremely helpful for discovery and development of antibiotics.

To address the inherent complexities of gene expression-based MOA screens, we first produced a large experiment database of hundreds of whole-genome *B. subtilis* expression profiles induced by a comprehensive set of antibiotics. These compounds correspond to a diverse set of structurally different chemicals using different, but well investigated, molecular mechanisms. The diversity of the chemistry in our dataset is very important as only this guarantees that compound-specific effects can be filtered out and truly mechanism-specific responses can be identified. Then we independently investigated the impact of drug dosage and exposure time on the MOA prediction accuracy. To maintain statistical rigor, the experimental dataset was randomly divided into two groups, a training set that was used to develop an expression-based MOA classifier, and a test set on which we evaluated the accuracy of the classifier and its dependence on experimental parameters. We then sought to statistically identify defined sets of MOA biomarkers whose expression carries a maximum of information for compound MOA screens. Based on our results, we were able to suggest optimized experimental protocols that we expect to significantly facilitate the mechanistic interpretation of compound screening data.

Our approach can be viewed as a generalization of the recently proposed and increasingly used reporter strain screening in antibacterial drug discovery campaigns (7, 16, 25–27). However, instead of using only the readouts of one reporter gene, we followed an *ab initio* approach to identify the sets of antibiotic efficacy biomarkers with the highest predictive power. A major goal was to identify a minimal biomarker set for optimal MOA classification as this would enable the miniaturization of assays for large scale routine compound library screens. Ultimately we were able to identify a set of approximately ~130 antibiotic efficacy biomarkers that represent a first step toward the definition of a universal efficacy biomarker set for antibiotic MOA screens. Independent biomarker annotation efforts revealed that the statistical choice of antibiotic efficacy biomarkers resulted in a physiologically meaningful selection of biomarkers. However, most biomarkers identified are novel and have not been described as efficacy biomarkers so far.

Interestingly the majority of the functionally characterized biomarkers appear to be only indirectly related to the primary molecular targets of the drugs. These biomarkers were only identified based on their consistent expression behavior when being exposed to antibiotic stress. The biomarker set proposed in this study reproduced a number of already established MOA-sensitive transcriptional units. However, a very interesting result of our study is that the expression of any of those “reporter genes” alone is not sufficient for a reliable MOA classification. We show that taking into account the simultaneous expression of a whole set of suitable reporter genes guarantees optimal performance when predicting the MOA class of novel substances. Thus, the parallel measurement of the signals of multiple antibiotic efficacy biomarkers can significantly improve the results of compound library screens.

The surprisingly low number of required antibiotic efficacy biomarkers suggests that the construction of simple and robust assays is indeed feasible even for expression technologies other than microarrays. For instance, the actual biomarker-based compound screening can be based on parallel RT-PCR assays or on measuring the signals coming from different reporter strains. Such types of high content screening technologies are likely to boost a wide usage of MOA-diagnostic assays in compound library screening campaigns. Although we present only data from antibacterial research, the essential components of the proposed biomarker discovery strategy are not prokaryote-specific. We are therefore confident that an analogous biomarker-based, high content compound screening approach could be applied to eukaryotic systems relevant for other medical indications as well.

Acknowledgments—We thank U. Reimann for technical assistance, Driss Talibi for technical advice, and G. Höfle for providing corallopyronin. We are grateful to O. Pfannes for critically reading the manuscript.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** Present address: Niederrhein University of Applied Sciences, Dept. of Chemistry, Adlerstrasse 32, D-47798 Krefeld, Germany.

§ Both authors contributed equally to this work.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** Present address: AlCirus, Bayer HealthCare Pharma Center, Aprather Weg, D-42113 Wuppertal, Germany.

| To whom correspondence should be addressed. Tel.: 41-61-6976767; Fax: 41-61-6977244; E-mail: Hans-Peter.Fischer@genedata.com. |

** REFERENCES**

1. Ellinger-Ziegelbauer, H., Stuart, B., Wahle, B., Bomann, W., and Ahr, H. J. (2004) Characteristic expression profiles induced by genotoxic carcinogens in rat liver. *Toxicol. Sci.* 77, 19–34

2. Ellinger-Ziegelbauer, H., Stuart, B., Wahle, B., Bomann, W., and Ahr, H. J. (2005) Comparison of the expression profiles induced by genotoxic and nongenotoxic carcinogens in rat liver. *Mutat. Res.* 575, 61–84

3. Goodsaid, F. M. (2003) Genomic biomarkers of toxicity. *Curr. Opin. Drug Discov. Dev.* 6, 41–49

4. Gunther, E. C., Stone, D. J., Genwien, R. W., Bento, P., and Heyes, M. P. (2003) Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9608–9613

5. Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C. H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J. P., Poggio, T., Gerald, W., Loda, M., Lander, E. S. and Golub, T. R. (2001) Multiclass
cancer diagnosis using tumor gene expression signatures. Proc. Natl. Acad. Sci. U. S. A. 98, 15149–15154

6. Stulke, J., Hanschke, R., and Hecker, M. (1993) Temporal activation of beta-glucanase synthesis in Bacillus subtilis is mediated by the GTP pool. J. Gen. Microbiol. 139, 2041–2045

7. Fischer, H. P., Brunner, N. A., Wieland, B., Paquette, J., Macko, L., Ziegelbauer, K., and Freiberg, C. (2004) Identification of antibiotic stress-inducible promoters: a systematic approach to novel pathway-specific reporter assays for antibacterial drug discovery. Genome Res. 14, 90–98

8. Staunton, J. E., Slonim, D. K., Coller, H. A., Tamayo, P., Angelo, M. J., Park, J., Scherf, U., Lee, J. K., Reinhold, W. O., Weinstein, J. N., Mesirov, J. P., Lander, E. S. and Golub, T. R. (2001) Chemosensitivity prediction by transcriptional profiling. Proc. Natl. Acad. Sci. U. S. A. 98, 10787–10792

9. Freiberg, C., Fischer, H. P., and Brunner, N. A. (2005) Discovering the mechanism of action of novel antibacterial agents through transcriptional profiling of conditional mutants. Antimicrob. Agents Chemother. 49, 749–759

10. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borris, R., Bourri, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Danchin, A., et al. (1997) The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature 390, 249–256

11. Freiberg, C., and Brunner, N. A. (2002) Genome-wide mRNA profiling: impact on compound evaluation and target identification in anti-bacterial research. Targets 1, 20–29

12. Chopra, I., Hesse, L., and O’Neill, A. (2002) Exploiting current understanding of antibiotic action for discovery of new drugs. J. Appl. Microbiol. 92, (suppl.) 34S–15S

13. Freiberg, C., and Brotz-Oesterhelt, H. (2005) Functional genomics in antibacterial drug discovery. Drug Discov. Today 10, 927–935

14. Hutter, B., Schaab, C., Albrecht, S., Borgmann, M., Brunner, N. A., Freiberg, C., Ziegelbauer, K., Rock, C. O., Ivanov, I., and Loferer, H. (2004) Prediction of mechanisms of action of antibacterial compounds by gene expression profiling. Antimicrob. Agents Chemother. 48, 2838–2844

15. Lazarevic, V., Dusterhoff, A., Solido, B., Hilbert, H., Mael, C., and Karamata, D. (1999) Nucleotide sequence of the Bacillus subtilis temperate bacteriophage SP3&c2. Microbiology 145, 1055–1067

16. Hutter, B., Fischer, C., Jacob, A., Schaab, C., and Loferer, H. (2004) Panel of Bacillus subtilis reporter strains indicative of various modes of action. Antimicrob. Agents Chemother. 48, 2588–2594

17. Debarbouille, M., Gardan, R., Arnaud, M., and Rauport, G. (1999) Role of bkdR, a transcriptional activator of the sigL-dependent isoleucine and valine degradation pathway in Bacillus subtilis. J. Bacteriol. 181, 2059–2066

18. Sonenshein, A. L. (2005) CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. Curr. Opin. Microbiol. 8, 203–207

19. Shivers, R. P., and Sonenshein, A. L. (2005) Bacillus subtilis ilvB operon: an intersection of global regulons. Mol. Microbiol. 56, 1549–1559

20. de Mendoza, D., Schujman, G. E., and Aguilar, P. S. (2002) in Bacillus subtilis and its Closest Relatives: from Genes to Cells (Sonenshein, A. L., Hoch, J. A., and Losick, R., eds) pp. 45–56, American Society for Microbiology Press, Washington, D. C.

21. Cao, M., and Helmann, J. D. (2002) Regulation of the Bacillus subtilis bcrC bacitracin resistance gene by two extracytoplasmic function sigma factors. J. Bacteriol. 184, 6123–6129

22. Cao, M., Wang, T., Ye, R., and Helmann, J. D. (2002) Antibiotics that inhibit cell wall biosynthesis induce expression of the Bacillus subtilis sigmW and sigmM regulons. Mol. Microbiol. 45, 1267–1276

23. Johansen, L. E., Nygaard, P., Lassen, C., Agero, Y., and Saxild, H. H. (2003) Definition of a second Bacillus subtilis pur regulon comprising the pur and xpt-pbuX operons plus pbuG, nupG (yxjA), and pbuE (ydhL). J. Bacteriol. 185, 5200–5209

24. Gmuender, H., Kurati, K., Di Padova, K., Gray, C. P., Keck, W., and Evers, S. (2001) Gene expression changes triggered by exposure of Haemophilus influenzae to novobiocin or ciprofloxacin: combined transcription and translation analysis. Genome Res. 11, 28–42

25. Sun, D., Cohen, S., Mani, N., Murphy, C., and Rothstein, D. M. (2002) A pathway-specific cell based screening system to detect bacterial cell wall inhibitors. J. Antimot (Tokyo) 55, 279–287

26. Shapiro, E., and Baneyx, F. (2002) Stress-based identification and classification of antibacterial agents: second-generation Escherichia coli reporter strains and optimization of detection. Antimicrob. Agents Chemother. 46, 2490–2497

27. Mascher, T., Zimmer, S. L., Smith, T. A., and Helmann, J. D. (2004) Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of Bacillus subtilis. Antimicrob. Agents Chemother. 48, 2888–2896