Gaining Insight into Off-Target Mediated Effects of Drug Candidates with a Comprehensive Systems Chemical Biology Analysis

Josef Scheiber,*[,†,#] Bin Chen,[†] Mariusz Milik,[†] Sai Chetan K. Sukuru,[†] Andreas Bender,[†,⊥] Dmitri Mikhailov,[†] Steven Whitebread,[‡] Jacques Hamon,[§] Kamal Azzaoui,[∥] Laszlo Urban,[‡] Meir Glick,[†] John W. Davies,[†] and Jeremy L. Jenkins[†]

Lead Discovery Informatics and Preclinical Safety Profiling, CPC, Novartis Institutes for Biomedical Research, 250 Massachussetts Avenue, Cambridge, Massachusetts 02139, and Preclinical Safety Profiling and Molecular Libraries Informatics, CPC, Novartis Pharma AG, Forum 1, 4002 Basel, Switzerland

Received September 22, 2008

We present a workflow that leverages data from chemogenomics based target predictions with Systems Biology databases to better understand off-target related toxicities. By analyzing a set of compounds that share a common toxic phenotype and by comparing the pathways they affect with pathways modulated by nontoxic compounds we are able to establish links between pathways and particular adverse effects. We further link these predictive results with literature data in order to explain why a certain pathway is predicted. Specifically, relevant pathways are elucidated for the side effects rhabdomyolysis and hypotension. Prospectively, our approach is valuable not only to better understand toxicities of novel compounds early on but also for drug repurposing exercises to find novel uses for known drugs.

INTRODUCTION

The overall impact a drug makes on a human subject can be classified into a multitude of desired and undesired effects. The undesired effects — so-called adverse drug reactions (ADRs) — have gained broad public attention and rose to a major concern in recent years. Legitimate estimates suggest that every year about 2 million patients in the United States are affected by a serious adverse drug reaction. This results in approximately 100,000 fatalities, making ADRs the fourth leading cause of death in the U.S., not far behind cancer and heart diseases. Similar numbers have been estimated for other western countries.[1−3] ADRs of all levels of severity were estimated to account for 2.5% of the total cases of unintentional injury in 2005−2006, of which 16.7% were severe enough to require hospitalization. Serious ADRs have been estimated to account for 1.8−6.2% of all hospitalized cases. The consequence is a huge burden on national economies, which is estimated to be $136 billion annually in the U.S. alone — a cost higher than that spent on cardiovascular or diabetic care.[1−3]

As a consequence, over the past 10 years, 19 broadly used marketed drugs[1] were withdrawn after presenting unexpected severe side effects, with Rofecoxib (Vioxx) and Cerivastatin (Baycol) being the most prominent cases. Thus, it is not surprising that besides death and loss of quality of life for the patient, ADRs are also a big commercial concern for the pharmaceutical industry. Ideally, safety issues, such as in case of Vioxx,[1] should not arise once a drug is marketed and tested for the first time on the population.

Also, the increase in safety requirements has led to a decreasing number of drug approvals and an increase in late drug discovery stage attrition rate.[4] This has a significant financial impact, which is — among other factors — reflected in the steep rise of research investment costs in the pharmaceutical industry.[4] As a consequence of the risk to patients and the huge expense incurred before launching a drug on the market, it would of course be better to predict or assess potential ADRs earlier in the drug discovery pipeline rather than postmarket. Therefore, avoiding and addressing ADRs early on has become a key goal in the development of a drug. The standard method to address these problems early on is the application of preclinical *in vitro* safety profiling.[5] Preclinical safety pharmacology (PSP) attempts to predict ADRs during the early phases of drug discovery by testing compounds using *in vitro* biochemical and cellular assays. If a compound is found to bind to a certain target, its effect might translate into the possible occurrence of an ADR in humans. This approach works best for targets which have established links with a certain adverse reaction. The hERG-related K+ channel,[6−9] the 5-HT2B receptor,[10] or the PXR nuclear hormone receptor[11] are typical examples which fall into this category. Currently, the possible "ADR space" is only partly covered by this approach: This means that for many side effects that are most likely linked to the interaction with a target, the respective link has not yet been established. Another limitation of these approaches

* Corresponding author phone: +41 (61) 324 8407; e-mail: mail@josef-scheiber.de.
[#] Current address: Text Mining Services, NITAS, Novartis Pharma AG, Forum 1, 4002 Basel, Switzerland.
[†] Lead Discovery Informatics, CPC, Novartis Institutes for Biomedical Research.
[⊥] Current address: Leiden/Amsterdam Center for Drug Research, Medicinal Chemistry Division, Einsteinweg 55, 2333 CC Leiden, The Netherlands.
[‡] Preclinical Safety Profiling, CPC, Novartis Institutes for Biomedical Research.
[§] Preclinical Safety Profiling, CPC, Novartis Pharma AG.
[∥] Molecular Libraries Informatics, CPC, Novartis Pharma AG.

10.1021/ci800344p CCC: $40.75
Published on Web 01/13/2009

is that they cannot address ADRs which are linked to multiple off-target effects.

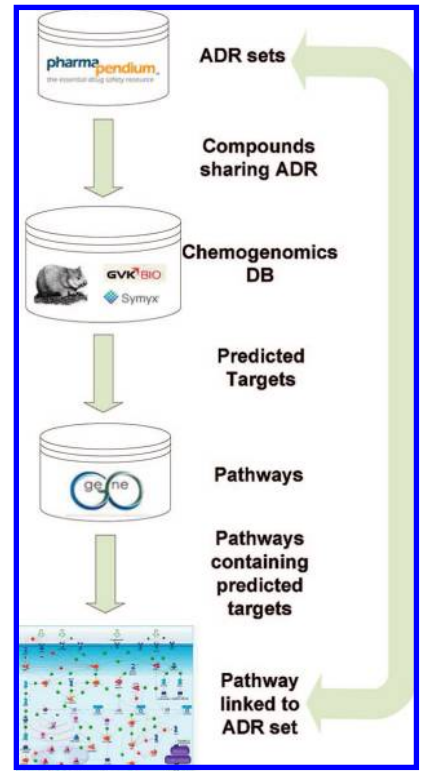
Extensive *in vitro* Safety Pharmacology Profiling can be quite expensive (in the order of about $3000−$10,000 for a panel of 50−100 targets).[5] To offset this there is increasing interest in *in silico* methods, which do not require any underlying functional knowledge, to predict possible ADRs. Bender et al.[12] established the proof-of-concept for predicting adverse events based on chemical structure alone. This approach aims to link targets to ADRs through correlation in chemical space. The work presented here, while related, does not focus on predictive information alone; rather, it combines knowledge-based annotation of ADRs to chemical structures and knowledge-based annotation of pathways. A similar concept has been pioneered by Ekins et al.[13] We extend their idea by using a much broader knowledge of possible targets and their active chemical space. Thereby we achieve a significantly broader coverage of biological space that possibly interacts with compounds sharing annotations. For an extensive overview of advances in predicting possible off-targets to better understand polypharmacology of compounds, the reader is referred to a recently published review by Jenkins et al.[14] Also, several successful applications of parallel pharmacophore-screening have been published recently.[15−17] All these methods are important and are used as a basis for the present study.

Not all of these approaches attempt to understand the interactions that happen between the different proteins and molecules in a human. Proteins function in networks and create complex cellular and larger-scale phenomena. Therefore if a drug is applied to such a system many different processes are initiated. First, there are the effects related to Absorption, Distribution, Metabolism and Excretion (ADME), which reflect the effect of complex body systems on the drug. As we focus on direct interaction with biological pathways metabolic fate is of secondary importance for us. This means that if we had annotated metabolite information available we would use it in our input, but we are not predicting metabolites as Ekins et al.[13] Second, there are effects that the drug exerts on the body that manifest themselves in physiological changes, which are usually simplistically considered as the therapeutic and toxic effects.

Generally, there is a large amount of data available related to genes and proteins and their interactions. Also there is a multitude of chemogenomics databases that describe modulation of proteins by small molecules. When used together, these data sources enable sophisticated systems biology and data analysis approaches. A few very promising methods in this area have been presented under the terms Systems pharmacology,[18] *in silico* pharmacology[19] and systems ADME/Tox.[20]

However, we feel that the term that best fits this area − due to its general applicability − is the one that Oprea et al. have coined in a seminal paper that calls for an in-depth integration of these sources in newly developed cheminformatics tools.[21] To describe this approach the term "Systems Chemical Biology" was established. Following their philosophy, in this contribution we present a novel workflow that integrates a well-established cheminformatics target-prediction method with systems biology data. More specifically, we aim to better understand the fact that often dissimilar chemotypes show the same side effects. Our analysis is based on the hypothesis that these compounds hit different targets in the same pathway and thereby cause the same phenotypic effect.

Scheme 1. Workflow To Identify Pathways Related to Defined Toxicities[a]



[a] A set of compounds sharing a common toxicity are extracted from PharmaPendium and the top 5 targets for each compound are predicted by using a Multiple Category Bayes model capable of predicting 2190 distinct targets. The predicted targets are then put into the context of pathways. Thereby toxicity can be linked to the pathways that have a high likelihood of being responsible for the undesired effect.

The outline of this work is represented in Scheme 1. First, a target set of compounds are collected that share the same ADR based on annotations from the PharmaPendium database (Elsevier). The subsequent and key step in the workflow is an *in silico* target prediction for the target set of compounds. This is based on a set of 2190 distinct protein target prediction models generated using chemical fingerprints for compounds from GVK Bio databases,[22] WOMBAT,[23] MDDR,[24] an in-house Kinase Knowledgebase, and Novartis Safety Profiling[5] databases. All targets therein were annotated with their corresponding NCBI GENESYMBOL as a common identifier where possible. After the targets are predicted for a set of compounds sharing the same ADR, the predictions are analyzed in the context of biological pathways the targets are known to be part of. Specifically, all predicted targets for all compounds in the ADR set are mapped into pathways in MetaBase (GeneGo) (for a data overview for Metabase see ref 25); all pathways are retrieved that contain at least one of the predicted targets. Then, the pathways are ranked (see Methods) to ultimately retrieve the pathway deemed most likely to be linked to the ADR under investigation.

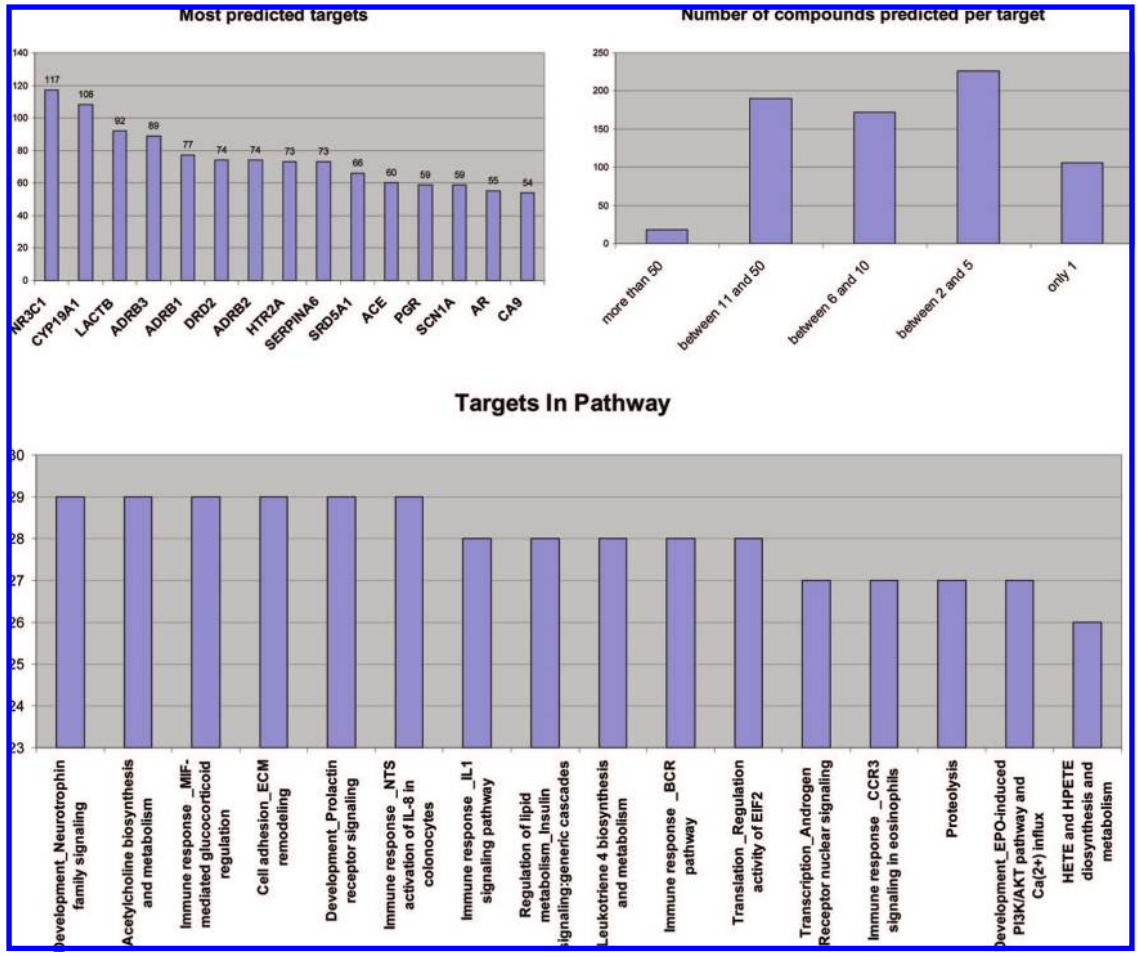


Figure 1. Top left: The frequencies of the most predicted targets and the number of associated compounds. Top right: Number of compounds that share predicted targets. Bottom: The most enriched pathways for the whole compound set (most targets predicted that are part of a pathway).

METHODS

Databases. The WOMBAT 2007.2,[23] GVK Bio,[22] and MDDR[24] Novartis *in vitro* Safety Profiling[5] data and Kinase knowledgebase were used to train a Multiple-Category Bayesian Model. In total these databases contain 2,902,930 data points, 2,381,785 of which are unique. This amounts to 1,458,680 unique molecules (as defined by unique Pipeline Pilot canonical smiles) with biological activity on 2190 unique targets (unique Entrez Gene GENESYMBOL Classification).

Elsevier's PharmaPendium database[26] was taken as input for retrieving compound sets that are known to share common toxicities. PharmaPendium makes drug safety data of U.S.-approved drugs available to researchers. This repository provides toxicity information annotated to molecules in the controlled vocabulary MedDRA,[27] which is the common terminology for ADR reporting. Preclinical, clinical, and postmarketing information is included and was used as input for our analysis. In total PharmaPendium contains 4405 different types of toxicities that are reported for at least two compounds. In total 1842 compounds are reported with ADR data. The median number of ADRs reported per compound is 46 on the lowest level of MedDRA. The compound with the most reported side effects is Aripiprazole (a schizophrenia treatment) with 692 distinct terms.

GeneGo's knowledge base MetaBase[25,28] was used as data source for biological pathway information.

MetaBase is, arguably, the most comprehensive manually curated database of mammalian biology and medicinal chemistry data available. Overall, it contains over 6 million experimental findings on protein−protein, protein−DNA, and protein−compound interactions. MetaBase also includes thousands of established signaling and metabolic pathways, ligand−receptor information for known drugs, drug targets and diseases, kinetic information on drug-metabolizing enzymes and relevant signaling proteins, ontologies for diseases, functional processes, toxicities, proteins, and drugs.

Training and Prediction of Targets Using Multiple-Category Bayesian Models. Laplacian-Modified Naive Bayesian Classification Models were built for each individual target and target family using components from Pipeline Pilot (SciTegic, Inc./Accelrys).[29] These models assume all variables are independent and use a Laplacian correction to reduce the bias caused through descriptors related to less prevalent features in the data set. The derivation of the Laplacian-modified Bayesian models for predicting biological targets based on chemical structures has been described previously.[30] The target prediction in our workflow uses exactly this method published by Nidhi et al.[30] All existing models contain multiple probabilities for multiple features. The overall probability ($P_{combined}$) for a compound having activity for a particular target is determined by summing the logs of the probabilities of activity for each feature.[31] Extended Connectivity Fingerprints with a radius of 4

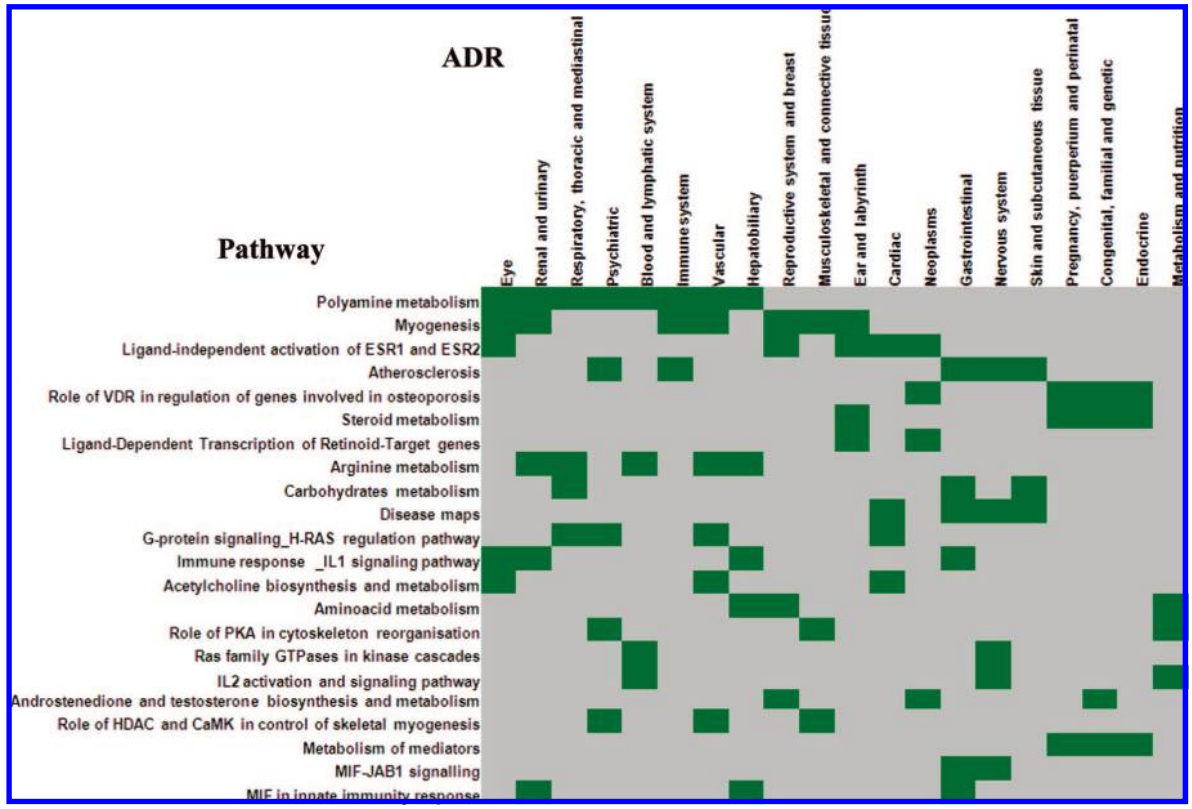


Figure 2. A global overview of pathways linked to ADRs in at least 3 different System Organ Classes. A green box means that there are chemical substructures that link the pathway on the left to the System Organ Classes on the top.

(ECFP_4, ECFPs) were used as chemical descriptors,[32,33] since it has been established that Bayesian models using circular fingerprints work well in virtual screening tasks[34,35] and also with primary screening results.[36,37] Other recent work on target prediction for small molecules employed 3D descriptors[38] which were shown to perform better if structures with high 2D similarity are not known. Current target prediction methods were also the subject of a recent review.[14]

Identifying Pathways. The workflow presented here was incorporated in its entirety in a single Pipeline Pilot protocol.

For each compound sharing the same toxicity, targets were predicted using the Laplacian-Modified Naive Bayesian Classification Model and ranked according to the Bayes Scores. Bayes Scores were normalized to account for the inherent molecular weight bias. Therefore the scores were divided by the molecular weight of the compound (as determined by the Pipeline Pilot component) under scrutiny. The five targets per compound that retrieved the highest Bayes Scores were used in the remaining part of the analysis; no further cutoffs were applied. We are aware that some compounds may associate with even more targets,[39] but this limit was necessary to reduce the complexity of the analysis and keep the number of adjustable parameters as low as possible. For every selected target that is linked to an ADR under scrutiny all associated pathways were retrieved from MetaBase.

To ultimately identify pathways that have the highest correlation with the ADR that is investigated it is also important to identify pathways that are consistently linked to many different drugs. The rationale behind this is that one wants to exclude pathways that contain targets that are predicted very often or that are also predicted to be hit by compounds not having the ADR under scrutiny. Therefore it is necessary to de-enrich the final result set for these pathways. All pathways are identified that are linked to drugs not showing the investigated ADR. To generate this "negative set" for every example, the same process as described before was performed for the remaining compounds in PharmaPendium. Again, for every compound the top 5 targets and their respective pathways were retrieved. This is then used to generate the final result.

MetaBase pathway IDs were used to compute how many different compounds have predicted targets in a certain pathway and per target in the pathway. To further focus the analysis, only pathways were considered that contained predicted targets for at least 10% of the compounds under investigation. Next, the sum of all scores for each identified pathway was calculated by taking all the Bayes Scores for the positive set and summing them up. The same addition is performed for the "negative control set". This negative score comprises the information on how often a certain pathway is predicted for compounds not showing the side effect under investigation. To compute the enrichment for every pathway the score for the positively correlated set is divided by the score for the negative set. All pathways with a quotient smaller than 1 (no enrichment toward analyzed toxicity) are excluded. The pathways are subsequently ranked by their enrichment score. Pathways that retrieve no negative score are ranked by the sum of the scores only. In a standard example two different lists result: one with pathways that only have a positive correlation and one which is enriched for compounds showing the specific toxicity.

RESULTS AND DISCUSSION

Large-Scale Analysis Predicts Modulation of Pathways for Marketed Drugs. As a first step, targets were predicted for all compounds from PharmaPendium, using a method described previously.[30] In total, 712 distinct targets

Table 1. Top Pathways Linked to Rhabdomyolysis through the Predicted Targets

pathway/rank/ enrichment factor	targets predicted in pathway (only showing targets predicted >5x)	number of compds per target	supporting literature evidence
cAMP signaling/1/∞	OPRM1	44	OPRM1 agonists induce hyperthermia[48,49]
	OPRK1	35	
	OPRD1	25	somatostatin induces hypoglycemia[50]
	C5AR1	16	
	CCKBR	13	skeletal muscle hyperthermia and hypoglycemia are causes for rhabdomyolysis[51]
	SSTR1	13	
	NTSR1	10	
	SSTR3	10	
	AGTR2	9	
	SSTR5	8	
	LIPE	6	
steroid metabolism/2/∞	CYP19A1	106	
	SRD5A1	64	HMGCR is the primary target for statins, a drug class known to share rhabdomyolysis as a side effect.[52−56]
	HMGCR	35	
	SOAT1	23	
	EBP	23	
	CYP17A1	23	
	ALDH1A1	17	
	COMT	15	
	CYP2C19	12	
	LSS	7	
transcription_CREB pathway/3/∞	ADRB3	89	calcium channels → hypercalcinaemia
	ADRB1	77	
	ADRB2	74	alpha1- plus beta3-adrenoreceptor antagonists attenuate rhabdomyolysis[57]
	HRH1	38	
	DRD1	28	The ADRB3 gene product, beta-3-adrenergic receptor, is located mainly in adipose tissue and is involved in the regulation of lipolysis and thermogenesis. **Beta adrenergic receptors** are involved in the epenephrine and norepinephrine-induced **activation of adenylate cyclase** through the action of G proteins.
	HTR7	25	
	CACNA1D	25	
	CACNA1A	23	
	MAP2K1	20	
	DRD5	16	
	CACNB1	16	
	CACNA1G	11	
	LDHA	7	
metabolism of mediators/4/∞	CYP19A1	106	linked to cAMP signaling in MetaBase (GeneGo)
	CHRM1	54	
	CHRM3	51	
	CHRM2	49	
	ACHE	35	
	MAOA	28	
	ODC1	22	
	ALDH1A1	17	
	CHRM5	16	
	COMT	15	
	CYP2C19	12	
	MAOB	12	
	CHRNB2	7	
lipid metabolism/5/∞	CYP19A1	106	Statins are used to lower blood lipids.
	CERK	39	
	TBXAS1	33	
	PTGS2	32	
	PLA2G1B	20	
	DPEP1	19	
	ALDH1A1	17	
	ALOX15	15	
	SPHK1	14	
	ALOX5	14	
	EPHX2	13	
	CYP2C19	12	
	PTGS1	11	
	PTGIS	10	
	EPHX1	9	
	LTA4H	9	
	CPT1A	8	
	ALOX12	7	
	LIPE	6	

were predicted at least once for at least one compound. For these 712 targets, there are 408 pathways in MetaBase that have at least one of the predicted targets as a member.

Figure 1 shows an overview for all predictions in aggregate. The most frequently predicted target for compounds in the data set is NR3C1, the glucocorticoid receptor, which

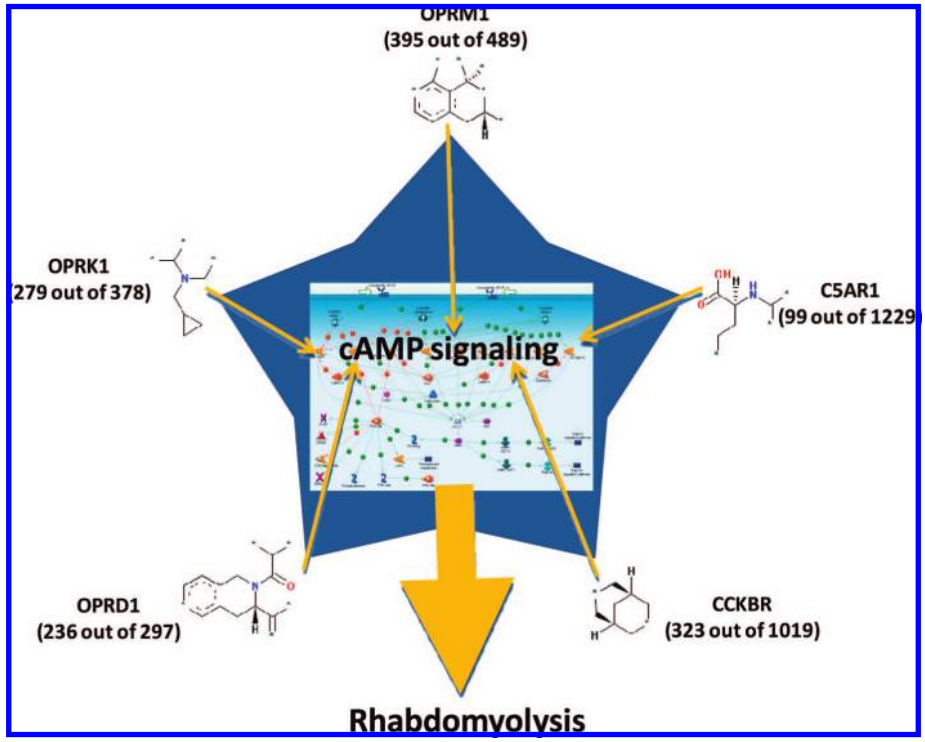


Figure 3. An illustration of how chemistry relates to biology in our approach. Shown are the top 5 predicted targets found in the cAMP signaling pathway, which was linked to rhabdomyolysis by our approach (see Table 1). For each of them we extracted the chemical feature with the highest correlation to the respective target activity. Also shown is the number of times the feature was found in the positive data points (with respect to activity against each particular target) in relation to the total number of ligands in the respective target class. It can be seen that chemically totally unrelated compounds that modulate different targets can still cause the same phenotypic effect downstream, i.e., rhabdomyolysis.

is predicted as a target for 117 drugs out of PharmaPendium. The high number of compounds predicted to modulate NR3C1 is in part due to the data set size in GVK: 7490 molecules are reported as being active against this receptor, so that the target prediction model for NR3C1 covers a large chemical space. On the other hand, it is known that many marketed drugs interact with this receptor. Indeed, DrugBank[40−42] reports 15 drugs as binders of NR3C1.

The vast majority of targets are predicted to be hit by a range of 2 to 50 small molecules, which reflects the results of Yildirim et al.[39] In stark contrast, the number of targets specifically predicted for 1 compound only is very low. The graph at the bottom of Figure 1 shows the pathways that have the highest coverage in terms of targets predicted for at least one compound in the PharmaPendium data set. It is important to mention that the targets in the pathways often overlap to a great extent, as human biology is not a compendium of distinct routes but rather a complex combination of interacting components. As such, certain pathways are linked to several MedDRA System Organ Classes (SOCs), i.e., compound interference with one target may affect several organs. Figure 2 shows all the pathways that can be linked to at least three different ADR SOCs by our approach. The pathway-ADR links range from intuitive (myogenesis pathway and musculoskeletal and connective tissue SOC) to the less obvious (myogenesis pathway and ear SOC). One limitation with our approach is that in certain cases correlations are not present between pathways and ADRs that might seem obvious to a pharmacologist (e.g., myogenesis pathway and cardiac disorders). This is not entirely unexpected in that our input information is limited by the diversity of chemical space around ADR-drug sets, the breadth of ligand-target knowledge in our chemogenomics databases, and the extent to which curated pathway gene sets match our intuitive understanding of pharmacology.

Specific Example - Rhabdomyolysis. Drug-induced rhabdomyolysis caused the withdrawal of Cerivastatin in 2001. During postmarketing surveillance, 52 deaths were reported in patients using Cerivastatin, mainly from rhabdomyolysis and its resultant renal failure.[43,44] Rhabdomyolysis is the rapid breakdown or lysis of skeletal muscle tissue due to injury to the muscle tissue. The muscle damage may be caused by physical (e.g., crush injury), chemical, or biological factors. The destruction of the muscle leads to the release of breakdown products of damaged muscle cells into the bloodstream; some of these, such as myoglobin, are harmful to the kidney and may lead to acute kidney failure. Risks were considerably higher in patients using fibrates, − mainly gemfibrozil (Lopid), and in patients using the highest approved (0.8 mg/day) dose of Cerivastatin.[45] Frequency of deadly incidents of rhabdomyolysis with Cerivastatin was 16 to 80 times higher than with other statins.[46] Another 385 nonfatal cases of rhabdomyolysis were reported. This put the risk of this (rare) complication at 5−10 times that of the other statins. Cerivastatin also induced myopathy (weakness of muscles) in a dose-dependent manner when administrated as monotherapy.[47]

PharmaPendium lists rhabdomyolysis as a side effect for 165 drugs. These drugs were used as the positive set, and the remaining compounds of the database were used as a negative set. The number of drugs appears to be very high, but the influence on the analysis outcome is not that high. This is because drug families where all family members cause rhabdomyolysis, such as the statins or fibrates, will

Table 2. Top Pathways Linked to Hypotension through the Predicted Targets[a]

pathway/rank/ enrichment factor	targets predicted in pathway (only showing targets predicted >5x)	number of compds per target	supporting literature evidence
ligand-independent activation of ESR1 and ESR2/1/∞	ADRB3	89	Estrogen receptor variants are clearly linked to hypertension.[59]
	ADRB1	77	
	ADRB2	74	The following links are to other targets in this pathway:
	HRH1	38	The prevalence of orthostatic hypotension was significantly different among GNAS1 genotypes and G-protein β3 subunit (GNB3) genotypes (members of this pathway). Multiple logistic regression analysis showed that both GNAS1 CC genotype and GNB3 C allele were independent risks for orthostatic hypotension. [60]
	ESR1	32	
	DRD1	28	ADR and DRD **agonists** are approved **to treat hypertension**.
	HTR7	25	
	MAP2K1	20	
	ESR2	18	
	DRD5	16	
	HRH2	15	
	HTR4	12	
	HTR6	8	
	MAPK3	7	
	AKT3	7	
	EGFR	6	
	ERBB2	6	
steroid metabolism/2/∞	**CYP19A1**	106	Simvastatin, an inhibitor of human HMGCR protein, is in phase II clinical trials to treat pulmonary hypertension.
	SRD5A1	64	
	HMGCR	35	In a conscious female mouse, the mutant aromatase [CYP19A1] gene (homozygous knockout) decreases diastolic blood pressure.[61]
	SOAT1	23	
	EBP	23	
	CYP17A1	23	
	STS	19	
	ALDH1A1	17	
	COMT	15	
	LIPA	13	
	CYP2C19	12	
	SQLE	12	
	FDFT1	11	
	HSD17B3	10	
	FDPS	8	
	ACOX1	6	
myogenesis/3/∞	CACNA1D	25	Mibefradil, a blocker of human CACNA1G protein, was approved to treat essential hypertension (withdrawn due to drug−drug interactions).
	CACNA1A	23	
	ACLY	22	Isradipine, Nicardipine, and Nisoldipine, blockers of human CACNA1D protein, have been approved to treat essential hypertension.
	CACNB1	16	
	PTPN1	11	The mutant mkp-1 [DUSP1] gene (homozygous knockout) in mice further decreases blood pressure that is already decreased by LPS in mouse peritoneum.
	CACNA1G	11	In mice, the mutant CACNB3 gene (homozygous knockout) increases blood pressure that involves high salt diets.[62,63]
	BAD	11	
	CACNB3	10	
	HDAC7A	9	
	AKT3	7	
	RPS6KB1	7	
PKA signaling/4/3.00	CACNA1D	25	Amrinone, an inhibitor of human PDE4A protein, has been approved to treat congestive heart failure (which is directly linked to hypotension).
	CACNA1A	23	
	CACNB1	16	
	CACNA1G	11	
	CACNB3	10	
	ADCY1	7	
	PDE4A	7	
role of HDAC and calcium/calmodulin-dependent kinase (CaMK) in control of skeletal myogenesis/5/2.78	CACNA1A	25	The mutant mouse CACNB3 gene (homozygous knockout) increases hypertrophy of smooth muscle layer in aorta related to high salt diets.
	HDAC7A	23	
	AKT3	16	Mouse CACNA1S is involved in the contraction of striated muscle.[64]
	CACNA1D	11	

Table 2. Continued

pathway/rank/ enrichment factor	targets predicted in pathway (only showing targets predicted >5x)	number of compds per target	supporting literature evidence
	RPS6KB1	10	In a 3 day-old mouse, the mutant Akt1 gene (homozygous knockout) and the mutant Akt3 gene (heterozygous knockout) affect heart morphology.[65]
	CACNA1S	9	
	CACNA1G	7	
	CACNB3	7	

[a] The targets in bold italics in the steroid metabolism pathway are also predicted for compounds causing rhabdomyolysis.

be overrepresented. This improves the signal-to-noise-ratio to a great extent and thereby enables us to better extract the pathways linked to rhabdomyolysis.

Table 1 shows the overview of the top 5 pathways linked to rhabdomyolysis in our analysis. It becomes apparent that in such cases also the primary target pathways are found if the input data set contains a family of drugs that share a common pathway as in this case the statins are HMG-CoA-Reductase-inhibitors and all of them are known to cause rhabdomyolysis. However, a literature search around the data for the identified pathways quickly reveals the fact that all of them can be associated with various events that are known to induce rhabdomyolysis. The pathway showing the highest enrichment is cAMP-signaling. Figure 3 shows an overview on how different chemical structures can ultimately affect the same biological pathway. Importantly, the top 5 targets predicted that relate to cAMP signaling all have different chemical features as their most correlated link to rhabdomyolysis. The literature evidence for this pathway's link can be found through factors inducing hyperthermia or hypoglycemia. Both are known reasons for rhabdomyolysis.[51] A key component of this pathway are the opioid receptors. For agonists of opioid receptors, it is well established that they induce hyperthermia.[48,51] This now clearly provides a link on how compounds interfering with cAMP signaling can cause rhabdomyolysis. However, this is further supported by the fact that Somatostatin induces hypoglycemia.[50] Thereby a second link into the pathway is established. The appropriate links are shown in Table 1 for all pathways being linked to rhabdomyolysis by our approach.

Specific Example — Hypotension (Low Blood Pressure). In theory the same approach can also be used to find new uses for old drugs.[58] If a compound hits a pathway that is known to also have a beneficial effect for certain diseases, this minor effect can be further investigated, and the drug can also be used for the secondary mode of action. Recent years have seen several successes with this so-called drug repurposing. A hypothetic example might be a cholesterol-lowering drug that also reduces blood pressure. To investigate such an example we applied our methodology to link pathways to the side effect "hypotension". PharmaPendium lists hypotension as a side effect for 729 drugs, which were used as positive set, and the remainder of the database was used as a negative set.

Table 1 shows an overview to the pathways linked to hypotension in our analysis. For all of them literature data are available to corroborate the link to lower blood pressure if a compound interacts with the mentioned pathway. This again supports the fact that noisy data sets can be analyzed by systems chemical biology approaches in a way that provides very interesting and useful insights in the underlying biology. The pathway retrieving the highest score for its link to hypotension is "ligand independent activation of ESR1 and ESR2". For the members of this pathway it is only a short distance to a hypotension link. Sufficient genetic evidence is available that links estrogen receptor variants to changes in blood pressure.[58] The evidence becomes even clearer when one takes into account the fact that adrenergic receptor and DRD antagonists are approved medications for the treatment of hypertension (high blood pressure). Consequently, if a drug hits the same pathway in a patient taking the drug for different reasons, the side effect is explainable.

Also germane to this discussion is that steroid metabolism is linked to both rhabdomyolysis and hypotension. Nine targets are predicted for both compounds having either rhabdomyolysis or hypotension as a side effect (marked in bold in Table 2). This can be explained by the data set composition: there are 102 drugs that have both side effects reported in Pharmapendium, and therefore this relationship is reflected in the predictions and consecutive analysis.

CONCLUSIONS

Naïve Bayesian modeling used in conjunction with Extended Connectivity Fingerprints has previously been shown to perform well in identifying actives in comparison to other substructure-based searching methods. It has been shown that its performance is also excellent for predicting targets for compounds. We have extended this approach by integrating with this well-established methodology the vast amount of knowledge available through curated molecular pathway databases. Preselecting the data set that will be investigated toward compounds that share a common phenotypic effect and analyzing the pathways that are common for the predicted targets typically produces results that agree with and are supported by published literature. Going forward, one can gain insight into less well-understood toxicities and link them to appropriate pathways, if such a link is detectable, given the strength of the pathway−toxicity link as well as the available data to support this connection. Also, side effects of compounds can be exploited to propose novel indications to repurpose drugs that are already known to be safe. The advantage of the systems chemical biology analysis is that identical targets are not only considered but rather that compounds may cause the same phenotype by hitting different targets in the same pathway.

Besides the application of this method to ADRs one can also envisage a use as a compound triaging tool to engineer desired properties into compounds. By hitting several targets in the target-pathway or other beneficial pathways one can come closer to the design of promiscuous drugs.[66]

SUMMARY AND OUTLOOK

We see systems chemical biology approaches as the key area for future development in cheminformatics. The vision would be to integrate all predictive approaches with data from biological databases. Importantly, the increasing public availability of data on drugs and druglike molecules may make analyses similar to that described above possible for scientists outside the private sector. For example chemical repositories such as DrugBank (http://redpoll.pharmacy.ualberta.ca/drugbank/),[40−42] PubChem,[67] PDSP,[68−70] ChemSpider (www.chemspider.com), and others consist of target and small molecule data that could be used to extend systems chemical biology approaches. In the future we envisage the full integration of pathway analysis tools with other informatics tools in order to fully leverage their content. As described in the examples above there have been other early efforts in this direction. A key issue will be (beside an integration of tools) the integration and federation of various data sources. One can imagine a data environment where all data from toxicogenomics are streamlined with results from all biochemical assays for a compound and easily accessible for every researcher. Having this kind of global molecular profile will also make it possible to describe the differences between compounds in detail and finally lead to novel ideas on how to get better drugs by enabling simulation of a "virtual human".[71]

ACKNOWLEDGMENT

J.S., S.C.K.S., and A.B. gratefully thank the NIBR Education Office for postdoctoral fellowships. B.C. thanks the NIBR Education Office for an intern fellowship.

REFERENCES AND NOTES

(1) Giacomini, K. M.; Krauss, R. M.; Roden, D. M.; Eichelbaum, M.; Hayden, M. R.; Nakamura, Y. When good drugs go bad. *Nature* **2007**, *446* (7139), 975–977.
(2) Leone, R.; Sottosanti, L.; Luisa Iorio, M.; Santuccio, C.; Conforti, A.; Sabatini, V.; Moretti, U.; Venegoni, M. Drug-related deaths: an analysis of the italian spontaneous reporting database. *Drug Saf.* **2008**, *31* (8), 703–713.
(3) van der Hooft, C. S.; Sturkenboom, M. C. J. M.; van Grootheest, K.; Kingma, H. J.; Stricker, B. H. C. Adverse drug reaction-related hospitalisations: a nationwide study in The Netherlands. *Drug Saf.* **2006**, *29* (2), 161–168.
(4) Hughes, B. Vaccine partnerships to tackle neglected diseases. *Nat. Rev. Drug Discovery* **2008**, *7* (4), 277–278.
(5) Whitebread, S.; Hamon, J.; Bojanic, D.; Urban, L. Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. *Drug Discovery Today* **2005**, *10* (21), 1421–1433.
(6) Hoffmann, P.; Warner, B. Are hERG channel inhibition and QT interval prolongation all there is in drug-induced torsadogenesis? A review of emerging trends. *J. Pharmacol. Toxicol. Methods* **2006**, *53* (2), 87–8105.
(7) Redfern, W. S.; Carlsson, L.; Davis, A. S.; Lynch, W. G.; MacKenzie, I.; Palethorpe, S.; Siegl, P. K. S.; Strang, I.; Sullivan, A. T.; Wallis, R.; Camm, A. J.; Hammond, T. G. Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development. *Cardiovasc. Res.* **2003**, *58* (1), 32–45.
(8) Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E. M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjogren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G.; Schneider, G. Development of a virtual screening method for identification of "frequent hitters" in compound libraries. *J. Med. Chem.* **2002**, *45* (1), 137–142.
(9) Sanguinetti, M. C.; Tristani-Firouzi, M. hERG potassium channels and cardiac arrhythmia. *Nature* **2006**, *440* (7083), 463–469.
(10) Ekins, S. Predicting undesirable drug interactions with promiscuous proteins in silico. *Drug Discovery Today* **2008**, *21*, 1304–1314.
(11) Ekins, S.; Erickson, J. A. A pharmacophore for human pregnane X receptor ligands. *Drug Metab. Dispos.* **2002**, *30* (1), 96–99.
(12) Bender, A.; Scheiber, J.; Glick, M.; Davies, J. W.; Azzaoui, K.; Hamon, J.; Urban, L.; Whitebread, S.; Jenkins, J. L. Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem* **2007**, *2* (6), 861–873.
(13) Ekins, S.; Andreyev, S.; Ryabov, A.; Kirillov, E.; Rakhmatulin, E. A.; Sorokina, S.; Bugrim, A.; Nikolskaya, T. A combined approach to drug metabolism and toxicity assessment. *Drug Metab. Dispos.* **2006**, *34* (3), 495–503.
(14) Jenkins, J. L.; Bender, A.; Davies, J. W. In silico target fishing: Predicting biological targets from chemical structure. *Drug Discovery Today: Technol.* **2006**, *3* (4), 413–421.
(15) Steindl, T. M.; Schuster, D.; Wolber, G.; Laggner, C.; Langer, T. High-throughput structure-based pharmacophore modelling as a basis for successful parallel virtual screening. *J. Comput.-Aided Mol. Des.* **2006**, *20* (12), 703–715.
(16) Steindl, T. M.; Schuster, D.; Laggner, C.; Langer, T. Parallel screening: a novel concept in pharmacophore modeling and virtual screening. *J. Chem. Inf. Model.* **2006**, *46* (5), 2146–2157.
(17) Steindl, T. M.; Schuster, D.; Laggner, C.; Chuang, K.; Hoffmann, R. D.; Langer, T. Parallel screening and activity profiling with HIV protease inhibitor pharmacophore models. *J. Chem. Inf. Model.* **2007**, *47* (2), 563–571.
(18) Ekins, S.; Mestres, J.; Testa, B. In silico pharmacology for drug discovery: applications to targets and beyond. *Br. J. Pharmacol.* **2007**, *152* (1), 21–37.
(19) Ekins, S.; Nikolsky, Y.; Nikolskaya, T. Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharmacol. Sci.* **2005**, *26* (4), 202–9.
(20) Ekins, S. Systems-ADME/Tox: resources and network approaches. *J. Pharmacol. Toxicol. Methods* **2006**, *53* (1), 38–66.
(21) Oprea, T. I.; Tropsha, A.; Faulon, J. L.; Rintoul, M. D. Systems chemical biology. *Nat. Chem. Biol.* **2007**, *3* (8), 447–450.
(22) GVK-BIO. *GVK Bio database suite*. http://www.gvkbio.com (accessed January 15, 2008).
(23) Sunset Molecular Discovery LLC. *WOMBAT: World of Molecular Bioactivity. [2007.2]*. http://www.sunsetmolecular.com (accessed January 15, 2008).
(24) Symyx Technologies. *MDL Drug Data Report*. http://www.symyx.com (accessed January 15, 2008).
(25) Ekins, S.; Bugrim, A.; Brovold, L.; Kirillov, E.; Nikolsky, Y.; Rakhmatulin, E.; Sorokina, S.; Ryabov, A.; Serebryiskaya, T.; Melnikov, A.; Metz, J.; Nikolskaya, T. Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms. *Xenobiotica* **2006**, *36* (10−11), 877–901.
(26) Elsevier. *PharmaPendium*. http://www.pharmapendium.com (accessed March 15, 2008).
(27) MedDRA MSSO. *MedDRA - the Medical Dictionary for Regulatory Activities*; 2008.
(28) GeneGo. *MetaBase for Systems Biology and Bioinformatics*. http://www.genego.com (accessed January 15, 2008).
(29) *Pipeline Pilot. [6.1]*; Scitegic, A. S. D. C.: 2008.
(30) Nidhi;Glick, M.; Davies, J. W.; Jenkins, J. L. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model.* **2006**, *46* (3), 1124–1133.
(31) Xia, X.; Maliski, E. G.; Gallant, P.; Rogers, D. Classification of Kinase Inhibitors Using a Bayesian Model. *J. Med. Chem.* **2004**, *47* (18), 4463–4470.
(32) Klon, A. E.; Glick, M.; Davies, J. W. Combination of a Naive Bayes Classifier with Consensus Scoring Improves Enrichment of High-Throughput Docking Results. *J. Med. Chem.* **2004**, *47* (18), 4356–4359.
(33) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures *Org. Biomol. Chem.* **2004**, *2* (22), 3256−3266.
(34) Glen, R. C.; Bender, A.; Arnby, C. H.; Carlsson, L.; Boyer, S.; Smith, J. Circular fingerprints: flexible molecular descriptors with applications from physical chemistry to ADME. *IDrugs* **2006**, *9* (3), 199–204.
(35) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (5), 1708–1718.
(36) Klon, A. E.; Glick, M.; Davies, J. W. Application of Machine Learning To Improve the Results of High-Throughput Docking Against the HIV-1 Protease. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (6), 2216–2224.

(37) Crisman, T. J.; Parker, C. N.; Jenkins, J. L.; Scheiber, J.; Thoma, M.; Kang, Z. B.; Kim, R.; Bender, A.; Nettles, J. H.; Davies, J. W.; Glick, M. Understanding False Positives in Reporter Gene Assays: in silico Chemogenomics Approaches To Prioritize Cell-Based HTS Data. *J. Chem. Inf. Model.* **2007**, *47* (4), 1319–1327.

(38) Nettles, J. H.; Jenkins, J. L.; Bender, A.; Deng, Z.; Davies, J. W.; Glick, M. Bridging Chemical and Biological Space: "Target Fishing" Using 2D and 3D Molecular Descriptors. *J. Med. Chem.* **2006**, *49* (23), 6802–6810.

(39) Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabasi, A. L.; Vidal, M. Drug-target network. *Nat. Biotechnol.* **2007**, *25* (10), 1119–1126.

(40) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36* (Database issue), 901–906.

(41) Wishart, D. S. In silico drug exploration and discovery using DrugBank *Current Protocols in Bioinformatics*; 2007; Chapter 14.

(42) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34* (Database issue), 668–672.

(43) Evans, M.; Rees, A. Effects of HMG-CoA reductase inhibitors on skeletal muscle: are all statins the same. *Drug Saf.* **2002**, *25* (9), 649–663.

(44) Furberg, C.; Pitt, B. Withdrawal of cerivastatin from the world market. *Curr. Control Trials Cardiovasc. Med.* **2001**, *2* (5), 205–207.

(45) Graham, D. J.; Staffa, J. A.; Shatin, D.; Andrade, S. E.; Schech, S. D.; La Grenade, L.; Gurwitz, J. H.; Chan, K. A.; Goodman, M. J.; Platt, R. Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs. *JAMA* **2004**, *292* (21), 2585–2590.

(46) Zeitlinger, M.; Muller, M. [Clinico-pharmacologic explanation models of cerivastatin associated rhabdomyolysis]. *Wien. Med. Wochenschr.* **2003**, *153* (11−12), 250–254.

(47) Saito, M.; Hirata-Koizumi, M.; Miyake, S.; Hasegawa, R. [Withdrawal of cerivastatin revealed a flaw of post-marketing surveillance system in the United States]. *Kokuritsu Iyakuhin Shokuhin Eisei Kenkyusho Hokoku* **2005**, (123), 41–45.

(48) Chen, X. H.; Geller, E. B.; DeRiel, J. K.; Liu-Chen, L. Y.; Adler, M. W. Antisense confirmation of mu- and kappa-opioid receptor mediation of morphine's effects on body temperature in rats. *Drug Alcohol Depend.* **1996**, *43* (3), 119–124.

(49) Wang, Y. X.; Xu, W. G.; Sun, X. J.; Chen, Y. Z.; Liu, X. Y.; Tang, H.; Jiang, C. L. Fever of recombinant human interferon-alpha is mediated by opioid domain interaction with opioid receptor inducing prostaglandin E2. *J. Neuroimmunol.* **2004**, *156* (1−2), 107–112.

(50) Vezzosi, D.; Bennet, A.; Courbon, F.; Caron, P. Short- and long-term somatostatin analogue treatment in patients with hypoglycaemia related to endogenous hyperinsulinism. *Clin. Endocrinol. (Oxford, U.K.)* **2008**, *68* (6), 904–911.

(51) Huerta-Alardin, A. L.; Varon, J.; Marik, P. E. Bench-to-bedside review: Rhabdomyolysis -- an overview for clinicians. *Crit. Care* **2005**, *9* (2), 158–169.

(52) Armitage, J. The safety of statins in clinical practice. *Lancet* **2007**, *370* (9601), 1781–1790.

(53) Harper, C. R.; Jacobson, T. A. The broad spectrum of statin myopathy: from myalgia to rhabdomyolysis. *Curr. Opin. Lipidol.* **2007**, *18* (4), 401–408.

(54) Zarowitz, B. J. Fixed-dose combinations for improving medication adherence in assisted living environments. *Geriatr. Nurs.* **2007**, *28* (6), 341–345.

(55) Radcliffe, K. A.; Campbell, W. W. Statin myopathy. *Curr. Neurol. Neurosci. Rep.* **2008**, *8* (1), 66–72.

(56) Ueda, S. Drug-induced rhabdomyolysis]. *Nippon Rinsho* **2007**, *65 Suppl 8*, 567–571.

(57) Sprague, J. E.; Brutcher, R. E.; Mills, E. M.; Caden, D.; Rusyniak, D. E. Attenuation of 3,4-methylenedioxymethamphetamine (MDMA, Ecstasy)-induced rhabdomyolysis with alpha1- plus beta3-adrenoreceptor antagonists. *Br. J. Pharmacol.* **2004**, *142* (4), 667–670.

(58) Chong, C. R.; Sullivan, D. J. New uses for old drugs. *Nature* **2007**, *448* (7154), 645–646.

(59) Lehrer, S.; Rabin, J.; Kalir, T.; Schachter, B. S. Estrogen receptor variant and hypertension in women. *Hypertension* **1993**, *21* (4), 439–441.

(60) Tabara, Y.; Tachibana-Iimori, R.; Yamamoto, M.; Abe, M.; Kondo, I.; Miki, T.; Kohara, K. Hypotension associated with prone body position: a possible overlooked postural hypotension. *Hypertens. Res.* **2005**, *28* (9), 741–746.

(61) Head, G. A.; Obeyesekere, V. R.; Jones, M. E.; Simpson, E. R.; Krozowski, Z. S. Aromatase-deficient (ArKO) mice have reduced blood pressure and baroreflex sensitivity. *Endocrinology* **2004**, *145* (9), 4286–4291.

(62) Murakami, M.; Yamamura, H.; Suzuki, T.; Kang, M. G.; Ohya, S.; Murakami, A.; Miyoshi, I.; Sasano, H.; Muraki, K.; Hano, T.; Kasai, N.; Nakayama, S.; Campbell, K. P.; Flockerzi, V.; Imaizumi, Y.; Yanagisawa, T.; Iijima, T. Modified cardiovascular L-type channels in mice lacking the voltage-dependent Ca2+ channel beta3 subunit. *J. Biol. Chem.* **2003**, *278* (44), 43261–43267.

(63) Zhao, Q.; Wang, X.; Nelin, L. D.; Yao, Y.; Matta, R.; Manson, M. E.; Baliga, R. S.; Meng, X.; Smith, C. V.; Bauer, J. A.; Chang, C. H.; Liu, Y. MAP kinase phosphatase 1 controls innate immune responses and suppresses endotoxic shock. *J. Exp. Med.* **2006**, *203* (1), 131–140.

(64) Chaudhari, N. A single nucleotide deletion in the skeletal muscle-specific calcium channel transcript of muscular dysgenesis (mdg) mice. *J. Biol. Chem.* **1992**, *267* (36), 25636–25639.

(65) Yang, Z. Z.; Tschopp, O.; Di-Poi, N.; Bruder, E.; Baudry, A.; Dummler, B.; Wahli, W.; Hemmings, B. A. Dosage-dependent effects of Akt1/protein kinase Balpha (PKBalpha) and Akt3/PKBgamma on thymus, skin, and cardiovascular and nervous system development in mice. *Mol. Cell. Biol.* **2005**, *25* (23), 10407–10418.

(66) Hopkins, A. L.; Mason, J. S.; Overington, J. P. Can we rationally design promiscuous drugs. *Curr. Opin. Struct. Biol.* **2006**, *16* (1), 127–136.

(67) PubChem. http://pubchem.ncbi.nlm.nih.gov/ (accessed January 15, 2008).

(68) NIMH Psychoactive Drug Screening Program. The NIMH PDSP is Directed by Bryan L. Roth, 2008.

(69) Roth, B. L.; Sheffler, D. J.; Kroeze, W. K. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat. Rev. Drug Discovery* **2004**, *3* (4), 353–359.

(70) Strachan, R. T.; Ferrara, G.; Roth, B. L. Screening the receptorome: an efficient approach for drug discovery and target validation. *Drug Discovery Today* **2006**, *11* (15−16), 708–716.

(71) Editorial. All systems go. *Nat. Rev. Drug Discovery* **2008**, *7* (4), 278–279.

CI800344P