JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Assessing Relative Bioactivity of Chemical Substances Using Quantitative Molecular Network Topology Analysis

Anna Edberg,[†] Daniel Soeria-Atmadja,[‡] Jonas Bergman Laurila,[§] Fredrik Johansson,[§] Mats G. Gustafsson,*[,∥] and Ulf Hammerling*[,⊥]

[†]Division of Food Data, National Food Agency, SE-75126 Uppsala, Sweden

[‡]Division of R&D Information, AstraZeneca Research and Development, SE-15185, Södertälje, Sweden

[§]Division of Information Technology, National Food Agency, SE-75126 Uppsala, Sweden

[∥]Division of Cancer Pharmacology and Computational Medicine, Department of Medical Sciences, Uppsala University and Uppsala Academic Hospital, SE-75185 Uppsala, Sweden

[⊥]Department of Risk Benefit Assessment, National Food Agency, SE-75126 Uppsala, Sweden

Supporting Information

ABSTRACT: Structurally different chemical substances may cause similar systemic effects in mammalian cells. It is therefore necessary to go beyond structural comparisons to quantify similarity in terms of their bioactivities. In this work, we introduce a generic methodology to achieve this on the basis of Network Biology principles and using publicly available molecular network topology information. An implementation of this method, denoted QuantMap, is outlined and applied to antidiabetic drugs, NSAIDs, 17β-estradiol, and 12 substances known to disrupt estrogenic pathways. The similarity of any pair of compounds is derived from topological comparison of intracellular protein networks, directly and indirectly associated with the respective query chemicals, via a straightforward pairwise comparison of ranked proteins. Although output derived from straightforward chemical/structural similarity analysis provided some guidance on bioactivity, QuantMap produced substance interrelationships that align well with reports on their respective perturbation properties. We believe that QuantMap has potential to provide substantial assistance to drug repositioning, pharmacology evaluation, and toxicology risk assessment.

INTRODUCTION

Data derived from a wealth of novel techniques and methodologies originally developed within exploratory-type fundamental research such as cellular gene expression, protein mapping, and metabolic pathway delineation have become available for the compilation of information to highly aggregated resources.[1−3] An extension of this concept and involving the action of small molecules on biological systems—named Systems Chemical Biology—has leveraged closer cooperation between scientific fields with little traditional communication, notably bioinformatics and cheminformatics.[4] Thus, a variety of novel molecular and informatics avenues have been exploited to gain better insight in drug efficacy and toxicity.[5] For example, drugs may be interconnected because of target promiscuity, also referred to as polypharmacology.[6,7] A variety of correlations can be accordingly identified between, for example, drugs and therapeutic indications or side effects, drugs and diseases, as well as gene expression and pathological reactions.[8−11] Another intriguing direction involves mapping of interrelationships across large sets of drugs by means of gene expression profiling in a standardized in vitro experimental setting, typified by the Connectivity Map.[12,13] Although much useful information has been gleaned from this approach, microarray experiments are inherently noisy, which often impose difficulties to making correct classifications without additional prior knowledge.[14−16]

Of particular interest to informatics-oriented pharmacology and toxicology is the incorporation of in silico techniques that appreciate both target(s) and functionally related intracellular proteins. The latter part can be perceived as interconnected intracellular biochemical pathways, i.e., signal transduction.[17] Systematically carried out comparisons between different chemical involvement in target binding and signal paths have opened up a new dimension to the understanding of similarities and differences in biological effect of various chemicals: a theoretical and practical applicable view assigned network pharmacology.[18−20] Much seminal work has, however, been conducted within fundamental studies in other areas, such as disease—including tumor—classification and cell biology.[21,22] Although each among several stratifications of this broad area may apply, two partly overlapping routes have emerged with notable clarity: interpretation of cellular perturbation (or tumor malignancy grade or tissue developmental stage) in the context of (i) prior knowledge on systematically delineated signal transduction pathways, on the

Received: September 12, 2011
Published: April 8, 2012

one hand, and (ii) protein−protein interaction (PPI) patterns and thereof derived modules, on the other. Both of these directions are highly dependent on publicly available data generated by yeast two hybrid and mass spectrometer experimentation.[22,23]

In regard to route one, various international initiatives systematize intracellular signal transduction in species-specific pathway maps, which are available in an array of public and commercial databases.[24] Some among the most significant repositories in the public domain are KEGG, Pathway Commons, HPD, and Nature Pathway Interaction Database (PID).[25,26] Although based on far from comprehensive knowledge, this direction has nonetheless proven useful to applied research in, for example, pharmacology and toxicology.[27−29] The second route involves a more strict focus on PPIs, either largely devoid of further processing or tied to various levels of operational module (or subnetwork) definition, in some cases being derived from quite sophisticated algorithmic elaboration.[30−33] This track is perceived as enabling an even better inroad to classification and interpretation of phenotypic state, relative to a pathway-bound view, because of considerable lag time from PPI knowledge to pathway definition and a considerable overlap between many pathways.[32]

Apart from fundamental resources within the PPI and pathway domains, several specialized and amalgamating databases have been built and made publicly available to support scientific progress in the area.[26,34] Notably, the STRING and HAPPI servers mine and integrate data from a diverse set of sources, including those appearing in published reports.[35,36] There are also resources devoted to protein binding to chemicals, STITCH 3 being a widely recognized source of such information.[34,37] Such interrelations are typically depicted as network graphs, which are amenable to various sorts of analyses, including identification of key network proteins.[38] A range of Web servers and standalone packages are available for this purpose, typified by refs 39−42.

In this paper, we introduce a multistep procedure that— based on publicly available computational tools—enables quantitative comparison across a data set of chemicals, most of which are functionally related. Referred to as quantitative molecular network topology-guided distance mapping of chemicals (QuantMap), our procedure first involves identification and characterization of relevant chemical-associated network topologies in a mammalian physiological environment, as accomplished by sequential operations in STITCH 3, STRING, and HUBBA.[35,37,40,43] These exercises ultimately result in chemical-associated protein listings used to create aggregated scores for each substance, in relation to a query compound, by means of a special ranking principle. Thus, a score was developed from ranked protein listings, for each pair of chemicals, that provides an advanced indication of overall biochemical relatedness across a composite set of substances. To demonstrate its potential, QuantMap was applied to a rather broad class of chemicals, including those of the endocrine disrupting chemicals (EDCs) category that interfere with mammalian steroidal hormonal homeostasis,[44] as well as two pharmaceutical classes outside the EDC domain. High network topology resemblance was found across several EDCs, xenoestrogens, and estradiol. Clustering analysis of QuantMap output data revealed several intriguing relationships. For example, the largest (and most proximal) clusters housed all substances known to directly act on the estrogen pathways, whereas an antiandrogen chemical appeared as a single cluster. Moreover, antidiabetic and NSAID drugs appeared as highly segregated groups. The QuantMap technique as well as findings on the chemical set are outlined and discussed below.

METHODS

List of the 18 Compounds Studied. The study encompassed the principal human endogenous estrogen and 12 EDCs of either natural or anthropogenic origin as well three NSAIDs and two antidiabetic drugs. They are specified as follows: 17β-estradiol (E2), bisphenol A (BPA), coumestrol (COUM), dibuthylphtalate (DBP), diethylstilbestrol (DES), endosulfan (ENDO), fulvestrant (FULV), genistein (GEN), p-nonylphenole (4-NP), raloxifene (RALOX), resveratrol (RES), tamoxifen (TAMOX), zearalenone (ZEA), glimepiride (GLI), rosiglitazone (ROS), ibuprofen (IBU), aspirin (ASP), and dicloflenac (DICLO). Chemical structure and estrogenic potency or other hormone-disrupting activity varies considerably across the set of EDC chemicals. The last five substances are, however, devoid of documented EDC activity, but were included for method development and validation purposes. For details, see Table 1.

Similarity Based on Chemical/Structural information. Structural similarity analyses of all chemicals selected for this study were conducted by means of PubChem's Chemical Structure Clustering Tool, wherein pairwise Tanimoto scores were computed from 2D structure fingerprints.[45,46] A data matrix of similarity values was imported from PubChem and subsequently processed equivalently with results from the QuantMap procedure, i.e., using a preprocessing step based on multidimensional scaling (MDS) prior to hierarchical clustering and visualization. Here, $D = 6$ dimensions were found to meet the adopted criteria (see also below under Visualization of Results). The PubChem compound identifiers (CIDs) corresponding to the 18 chemicals are in Table 1.

Outline of QuantMap Procedure for Quantification of Similarity Based on Molecular Network Topology. QuantMap is a composite procedural/algorithmic methodology, that in its current design incorporates information from two public servers—STITCH3 and STRING—both housing aggregated data from various other sources. A third publicly accessible resource—HUBBA—was consulted for network topology processing. Briefly, QuantMap involves the following two major steps:

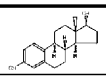
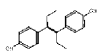
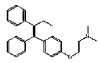
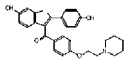
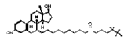
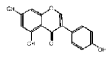
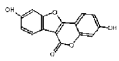
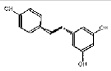
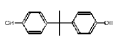
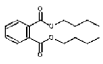
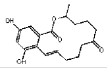
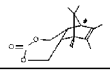
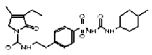
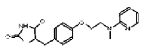
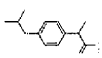
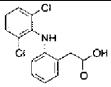
1. Retrieve/generate a protein network topology signature for each substance (see Procedure I below.)
2. Quantify similarities between each compound pair on the basis of protein network topology signatures (See Procedure II below).

The two procedures are described below in greater detail, including the consulted publicly available resources.

Resources Used in the Implemented Version of QuantMap. STRING (search tool for retrieval of interacting genes/proteins) is a frequently updated protein−protein interaction (PPI) meta-database, which compiles and renders available such associations. Thus, interactions shown in STRING are aggregated from various sources and represent an amalgamation of experimentally verified and predicted connections, the latter part derived by various indirect associative mapping methods. Each association is assigned an individual and a combined confidence score, thereby reflecting the relevance of each source and the total weight, respectively.[47]

STITCH3 (search tool for interaction of chemicals, version 3) is a sister database of STRING that, in addition to PPIs, includes protein−chemical relations. It integrates data dispersed over the literature and various repositories on biological pathways, drug−target relationships, and binding affinities. Networks elaborated in STITCH3 are created by mapping interactions onto a consolidated

Table 1. Molecular Structures, CAS RNs, Trivial Names, Acronymes (used in this paper), and CIDs (PubChem compound identifiers) for Substances Addressed in This Study[a]

Molecular structure	CAS RN	Trivial name	Acronym in this paper	CID (PubChem)
	82115-62-6	17β-estradiol	E2	5757
	56-53-1	Diethyl stilbestrol	DES	448537
	10540-29-1	Tamoxifen	TAMOX	2733526
	82640-04-8	Raloxifene	RALOX	5035
	129453-61-8	Fulvestrant	FULV	104741
	446-72-0	Genistein	GEN	5280961
	479-13-0	Coumestrol	COUM	5281707
	501-36-0	Resveratrol	RES	445154
	80-05-7	Bisphenol A	BPA	6623
	104-40-5	4-nonyl phenol	4-NP	1752
	84-74-2	Di-butyl phthalate	DBP	3026
	17924-92-4	Zearalenone	ZEA	5281576
	115-29-7	Endosulfan	ENDO	3224
	93479-97-1	Glimepiride	GLI	3476
	122320-73-4	Rosiglitazone	ROS	77999
	88566-80-7	Aspirin	ASP	2244
	9261-49-7	Ibuprofen	IBU	3672
	15307-86-5	Dicloflenac	DICO	3033

[a]E2 and DES are an endogenous sex steroid and a discontinued estrogenic drug, respectively. TAMOX and RALOX are selective estrogen receptor modifiers (SERMs), whereas FULV is a potent antiestrogen, all being market-approved drugs. GEN, COUM, and RES are natural phytoestrogens. BPA, 4-NP, and DBP have wide application as industrial plasticizers. ZEA and ENDO are a mycotoxin and a pesticide, respectively. GLI and ROS are approved drugs for clinical therapy of type 2 diabetes mellitus. ASP, IBU, and DICLO are NSAID drugs.

set of chemicals, as derived from PubChem. The Web interface and scoring system are identical to those of STRING.[37]

HUBBA (hub objects analyzer) is a framework focused on the computation of network centrality measures.[40] A centrality measure (CM) is a mathematical quantification based on graph theory and aimed at identifying important nodes within a network. All CMs assign numerical values to every node and thus enable identification and ranking according to functional

importance. Various CMs are available to identifying key nodes of biological networks. Here, four CMs perceived to carry relevance to toxicology topics were chosen: *betweenness*, *node degree*, *edge percolation component*, and *bottleneck betweenness* reflects the role of a node in global network architecture and is expressed as the fraction of shortest paths between node pairs that pass through a given node.[30] *Node degree* specifies the number of edges connected to a given node. *Edge percolation component* estimates correlation of nodes by removing randomly chosen fractions of links in the network, followed by inspection of whether they remain connected. This operation can provide a measure of node connection strength to the rest of the network, and thus its importance.[48] It takes, in contrast to *node degree*, the topology of the entire network into account.[40] *Bottleneck* nodes may be hubs with central or peripheral positions in the network or nonhubs linking functional protein clusters together.[30]

QuantMap Design. *Procedure I: Generation of Protein Network Topology Signature of Substance X.* For a query substance X of interest:

1. Invoke STICH3 and enable all available experimental and predictive associations, thereby rendering verified and presumed interactions involving chemicals and cellular proteins open to network delineation. This setting allows for the identification of an equivalent number of target proteins across the set of chemicals of this study.
2. Select the $M_{STITCH3}$ closest proteins (i.e., with high confidence scores, which mirrors the number and quality of sources supporting the interaction) from the STITCH3 network. This set is subsequently used as a query in STRING (or an analogous PPI database), which suggests a network of PPIs.
3. Query for the top M_{STRING} interacting components in STRING, using restrictions R_{STRING}, which ensure retrieval of a protein−protein network of acceptable confidence (see details in QuantMap Parameter Settings). The union of $M_{STITCH3}$ and M_{STRING} proteins, including all interactions between them conforming to restrictions R_{STRING}, constitutes a network N, being prepared for topology analysis.
4. Use HUBBA (or an equivalent tool for network analysis) to create M_{node} different importance scores for proteins of the protein−protein network N.
5. For each protein, compress M_{node} importance scores into a single aggregated score. Example (used in this article): Perform a transformation of all the importance scores into ranks and use the median rank as the final importance score.
6. Listings of proteins, ranked according to their combined/overall network weight, represent signatures of the chemicals.

QuantMap Parameter Settings. While keeping all other parameters fixed, various combinations of seed protein numbers, as retrieved from STITCH3 ($M_{STITCH3}$), and network-integrated proteins identified in STRING (M_{STRING}), were assessed by clustering appearance of a subset of size eight from the total set of 18 chemicals addressed in this paper. This training/tuning set incorporated eight substances of three distinct classes (TAMOX, RALOX, ZEA, ASP, IBU, DICLO, GLI, and ROS). Four different network sizes (50, 100, 150, and 200 proteins) and three different lengths of seed protein listings (10, 15, and 20 proteins) were accordingly evaluated in eight different (among the 12 possible) combinations; see Figure 1 of the Supporting Information). On the basis of outputs of this exercise $M_{STITCH3}$ was set to 10 and M_{STRING} to 150. M_{node} was set to four, thus all the four centrality measures outlined above were used. This setting enables robust identification of key network proteins, i.e., attached to essential biochemical/signal transduction functions. The chosen restrictions of R_{STRING} keeps only interactions with confidence scores above 0.7 (STRING confidence measure in the interval [0,1]) and support from either of the experiments, databases, or textmining types (according to naming conventions of STRING).

Procedure II: Rank Based Similarity of Signatures. Spearman's foot rule is a well-established method to compute similarities among ranked listings, with reported application in molecular profiling.[49,50] Thus, the mutual similarity of two drugs X_A and X_B, with protein network signatures S_A and S_B, respectively, and each attached to an individual $M_{protsign}$ signature, was quantified by means of this principle.[51] Initially, proteins devoid of common appearance in all signatures of a given $M_{protsign}$ pair are to the bottom of the respective listing (rank = $M_{protsign}$ + 1). In this context, Spearman's foot rule distance is expressed as

$$D_{protein}(X_A, X_B) = D_{protein}(S_A, S_B) = \sum_{i=1}^{P} |\tau(i) - \sigma(i)|$$

where $\tau(i)$ is the rank of protein i in the signature S_A, $\tau(i)$ is the rank of the same protein in the signature S_B, and P is the total number of proteins.

Processing and Visualization of Results. *MDS as a Preparatory Step To Visualize QuantMap Results by Hierarchical Clustering.* To prepare the derived relationships across all substances in the data set for clustering analysis and visualization, data from all score values in the 18 × 18 distance/similarity matrix D_{global} was subjected to nonmetric multidimensional scaling (MDS), using the Matlab function mdscale (default settings according to Matlab version 7.11.0) on data from all the score values in the 18 × 18 distance/similarity matrix D_{global}. The function mdscale D_{global} uses as input and returns an $N_S \times P$ matrix, here denoted Y, which is a configuration of N_S points (18 substances) in P dimensions. Distances between N_S rows in Y are selected by midscale to enable the best possible preservation of distances across substances described in D_{global}. In addition to coordinates specifying the position of each drug in the reduced space, a stress value is reported (reflecting the percentage of unexplained variation in D_{global}). This stress value was used to select the number of dimensions P. More specifically, P was increased until the relative stress value increments were below τ_{MDS} = 1%. This thresholding step to select a suitable value of P was performed to avoid data over fitting. For data matrices produced in this study, P = 6 dimensions were found to meet the above-mentioned criterion. The MDS data processing step was used partly to achieve noise suppression but chiefly to obtain coordinate vectors for each observation, which are needed prior to data submission to a multibranching divisive hierarchical clustering algorithm (see next section). Contrary to most hierarchical clustering algorithms that only use distance matrices, this algorithm needs access to coordinate vectors for each individual observation.

Hierarchical Top Down Clustering of MDS Outputs. The hierarchical top down cluster analysis of MDS output data was conducted by means of a multibranching algorithm, which in many respects is an implementation of a procedure outlined by Lamrous and Taileb.[52] This algorithm employs classical clustering algorithm k-means to achieve cluster representations at each branching point and the likewise classical average silhouette width for automatic selection of the most suitable number of subclusters k.

The algorithm starts with all substances as one cluster and at each branching point an optimization procedure is employed to select the most favorable subcluster configuration. The branch lengths in the resulting dendrogram are directly proportional to dissimilarity of subcluster centers to their respective parent cluster centers.[53] The maximum number of clusters allowed at each branching point was set to 18 (equaling the number of substances) and the number of k-mean replicates to 150 (at each branching, k-mean is run 150 times and the best cluster configuration is used). A Euclidian distance measure was used.

RESULTS

QuantMap: Summary of Key Procedures. The starting point involved consultation of the STITCH3 server, using any among the 18 selected chemicals as a query submission (Table 1). A typical output from such an inquiry appears as a rather condensed network representation of the query substance's interaction with other chemicals and several pertinent cellular proteins.[37,43] The identified proteins were subsequently placed in a more comprehensive biological activity context by adding protein−protein interactions from STRING.[35] An accordingly derived final network comprises a markedly expanded number of proteins. The ensuing processing involved network analysis, incorporating four distinct centrality measures, as conducted in HUBBA[40] and applied to every substance-specific network. Ranked protein listings were thus obtained, here referred to as each chemical's signature, and used to create distance matrices by means of Spearman's foot rule. Finally, such matrices were, subsequent to compression of dimensionality by MDS, fed to an in-house implemented clustering algorithm, which allows divisive multibranching clustering analysis (for details, see Methods).

The settings for the network analysis were derived from a comparison of eight different combinations of the parameters $M_{STITCH3}$ and M_{STRING} using a subset of eight chemical substances (see Methods). The procedure proved quite robust against parameter changes but output dissimilarities were nonetheless observed (Figure 1 of the Supporting Information). On the basis of these results, $M_{STITCH3} = 10$ and $M_{STRING} = 150$ were chosen. The value $M_{STITCH3} = 10$ was found small enough to accommodate a rather wide range of chemicals, including those attached to modest documentation on primary and secondary target proteins, but sufficiently large to providing a distinguishing and altogether useful input to the ensuing process (not shown). Key parts of this scheme are depicted in Figure 1 and further detailed in Methods.

Chemical Similarity among the Test Substances. Despite sharing a hormone-disrupting action— largely of the estrogenic type—the set of EDC substances as well as those encompassing chemicals of two non-EDC drug classes are rather diverse with regard to chemical structure. To acquire a reasonably accurate view on this matter, a straightforward structural similarity analysis was conducted. First, fingerprint representations were computed by means of algorithms implemented in PubChem BioAssay, and molecular similarity was subsequently derived by the Tanimoto distance metric.[45,46,54] On the basis of a resulting distance matrix, then MDS was employed to represent each substance as a point in a six-dimensional coordinate system (see Methods). Finally, the accordingly derived relationships between substances allocated to the study were visualized by means of divisive hierarchical clustering (Figure 2).

The substances fall in two major clusters, one housing ENDO only, whereas the remaining chemicals appear in the other, which segregates in two main subclusters. Clearly, ENDO is the structurally most remote chemical, relative to the remaining test compounds. Moreover, E2 and FULV—the only steroids among the substances addressed here—aggregate tightly, as do TAMOX and RALOX (also being mutually similar in a structural context). Two additional phytoestrogens (COUM and GEN), representing the coumestan and isoflavone structural class, respectively, cocluster with a mycotoxin (ZEA) and a NSAID (ASP). Another assembly houses mono- or biphenols, incorporating environmental pollutants (BPA and 4-NP), a stilbene-type phytoestrogen (RES), and a discontinued pharmaceutical xenoestrogen (DES). The antidiabetic drugs (GLI and ROS) appear in a common cluster, whereas NSAIDs (ASP, IBU and DICLO) are dispersed in the chart (Figure 2). On a general basis, the EDCs—based on structural similarity— largely coappear in a broad common subcluster with several constellations at lower hierarchical levels. Two of those, holding either selective estrogen receptor modifiers (SERMs) or antidiabetic drugs, are pharmacologically coherent (except for a NSAID in the latter), whereas most such subclusters display functional heterogeneity within the EDC range of actions (and potency).

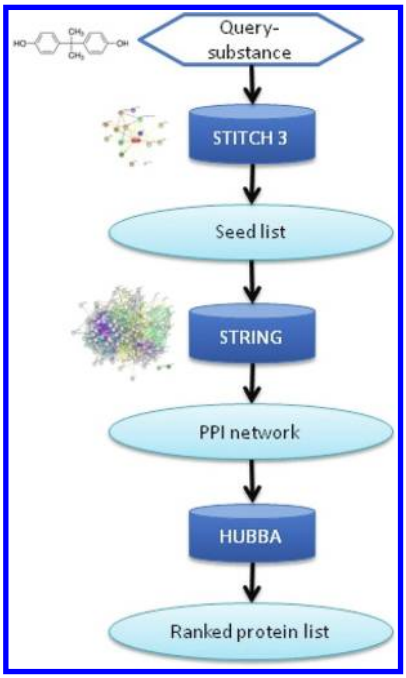


Figure 1. Flow-chart summarizing key steps of the tiered QuantMap procedure. Consultation of STITCH3 and STRING servers in sequence allows identifying proteins directly connected with the query substances (seed list) and functionally associated proteins (PPI network), respectively. The HUBBA server provides online assistance to computational analysis of biological networks (see also Methods).

Global Bioactivity Similarity Map. Analysis of the chemicals set by QuantMap (see Methods for details) produced a distance matrix, which eventually was processed and thereby translated into a hierarchical dendrogram (Figure 3). Relationships across the chemicals are strikingly distinct from an analogous display in Figure 2, on the basis of chemical similarity analysis using fingerprint encoding and the Tanimoto similarity measure. QuantMap revealed five main clusters, with estrogenic-type EDCs appearing on two proximal branches, and DBP—being an antiandrogen—as a singleton cluster. Moreover, non-EDC substances emerge as two well-separated groups, ASP and IBU, appearing as a subcluster (Figure 3). Notably, phytoestrogens and

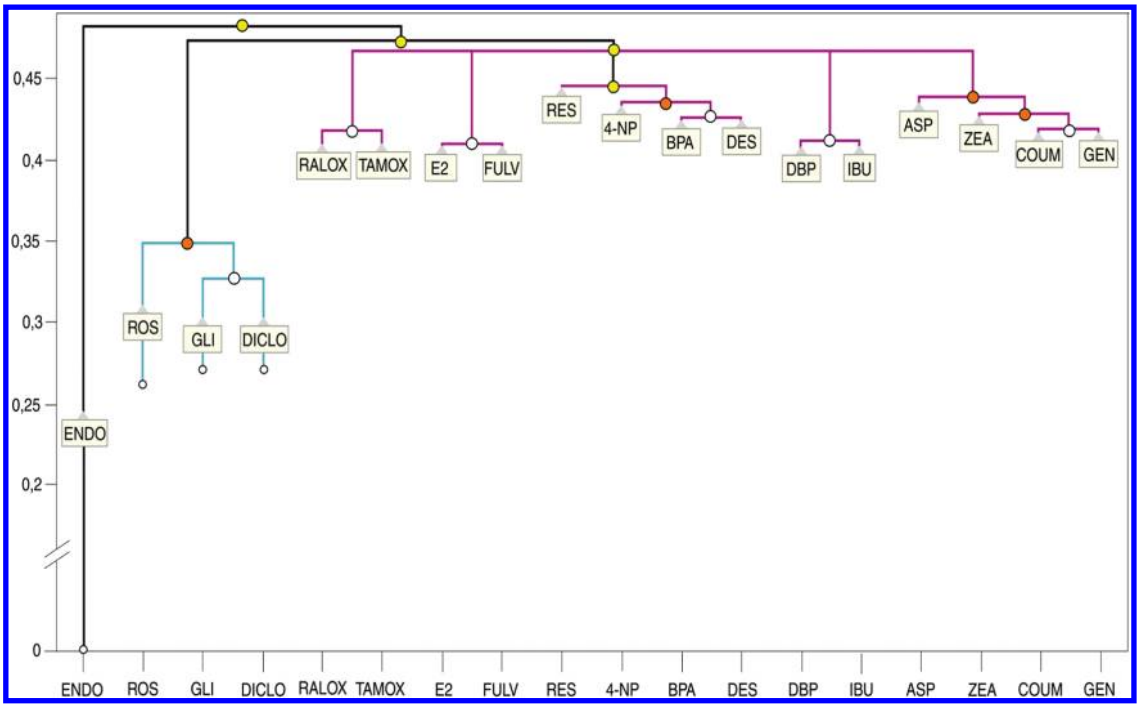


Figure 2. Multibranching dendrogram displaying structural similarity within the set (18) of molecules addressed in this study. The data matrix was created by means of PubChem's Chemical Structure Clustering Tool, wherein pairwise Tanimoto scores were computed from 2D structure fingerprints. A preprocessing step, involving multidimensional scaling, preceded hierarchical clustering and visualization. Color-coded nodes at each branching point indicate (sub-) cluster separation degree, ranging (in this figure) from yellow (modest) to white (good). Notably, lengths of each vertical edge in the dendrogram are inversely proportional to degree of structural relatedness, at each clustering level. Thus, RALOX and TAMOX are much closer at a molecular structural level than GLI and DICLO (see also Methods).

SERMs are grouped together, whereas the most potent estrogenic compounds (E2, DES, and ZEA) coaggregate in another cluster that also holds the environmental pollutants BPA and 4-NP. As shown in the chart, TAMOX and FULV as well as GEN and RALOX are mutually very similar, whereas RES—appearing in the same cluster—attains a relatively remote position.

Robustness. To probe for QuantMap robustness against changes in the input data set, an iterative analysis was performed where in each iteration two chemicals were deleted prior to processing. This operation did not cause disruption of the overall clustering patterns (not shown).

■ DISCUSSION

QuantMap in a Systems Biology and Systems Medicine Perspective. In the emerging field of systems biology,[55,56] different levels of information from various parts of an organism are integrated to develop an understandable model of how biological systems function. A new branch in systems biology, specifically involving the action of small molecules on biological systems, has been named systems chemical biology.[4] This field has leveraged closer cooperation between bioinformatics and cheminformatics, two scientific fields that have had relatively little communication in the past.[57,58] In this context, QuantMap is a natural and timely development.

Recent advancements in systems biology have generated an impressive accumulated amount of in vitro data across a very broad range of assay settings, as typified by achievements within the ToxCast and Tox21programmes.[59,60] Other significant advancements within this broad area of chemicals activity mapping, but even more focused on nontargeting profiling technology, are the NCI-60 and the Connectivity Map projects as well as the BioMAP concept.[12,61−64] Accumulated data from such initiatives, jointly with relevant multivariate computational data analysis, has opened new avenues to the comprehensive understanding of cellular responses to chemical exposure and other external stimuli.[65] As much of the important seminal work was conducted broadly within the molecular pharmacology domain, network pharmacology has become a scientifically well-recognized concept.[4,18,19] One example from this nonreductionist course has become known as pathway network mapping, largely built on the systematic compilation of accumulated data on protein−protein interactions, small molecule−protein associations, and network graph theory.[30,34,66] Thus, QuantMap may be perceived as a new contribution to the subfield of pathway network mapping.

Enzyme signaling pathways in cells adopt a rather interwoven structure. Thus, signal transmission routes to perturbation caused by chemical substances are also complex.[19] Such networks often consist of modules, each containing a variable number of quite strongly connected nodes, which in turn are connected with other locally cohesive clusters. This structure is, by virtue of its modular structure, quite tolerant of changes.[67] Nonetheless, so-called bottleneck nodes, typically individual proteins or protein complexes, are often key regulators of gene expression with a potential to incur significant impact on cellular functions.[68,69] QuantMap is designed to focus on this kind of key regulators. By means of HUBBA, QuantMap attempts identifying this sort of central network nodes and employing them as a compressed network signature suitable for comparison with other networks. In other words, QuantMap is based on the idea to compare extracted signatures of this kind for any substance pair of interest.

Chemical Similarity Map versus QuantMap Results. A straightforward chemical similarity analysis revealed an overview of molecular relatedness within the set of EDCs, NSAIDs,

dx.doi.org/10.1021/ci200429f | J. Chem. Inf. Model. 2012, 52, 1238−1249

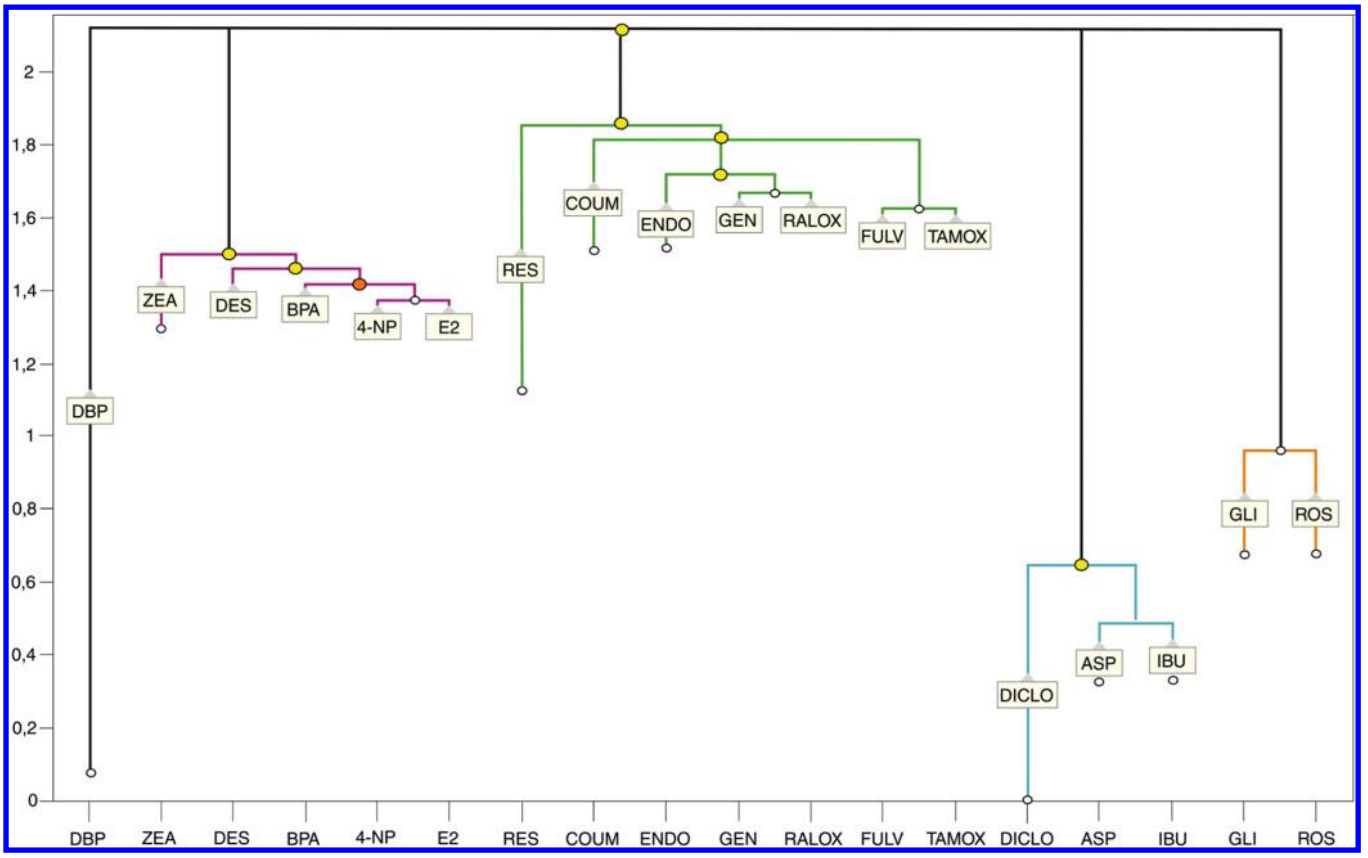


Figure 3. Multibranching cluster dendrogram of molecules, addressed in this study, based on results from QuantMap. Downstream processing, i .e., nonclassical MDS, clustering analysis, and visualization, were analogous to that of Figure 2 (see also Methods). Small circles, appearing at the end of each branch, represent molecules. Color-coded nodes at each branching point indicate (sub-) cluster separation degree, ranging from red (poor) to white (good). The lengths of each vertical edge of the dendrogram are inversely proportional to degree of QuantMap-assessed relatedness, at each clustering level. EDCs appear in the three left clusters (including a singleton), whereas NSAIDs and drugs used in therapy against diabetes type 2 fall in the remaining two clusters.

and antidiabetic drugs addressed in this paper. This exercise was in part conducted to provide a basis for comparison with results generated by network topology mapping. As obvious from the clustering charts (Figures 2 and 3), however, there is modest—at best—resemblance between clustering patterns derived from structural similarity, on the one hand, and QuantMap, on the other. One exception is that all EDCs—based on structural similarity— appear in a wide subcluster and several among them even aggregate in functionally relevant constellations. Nonetheless, NSAIDs appear scattered among the EDCs, and several substances of the latter category do not seem to form functionally homogeneous groups. Notably, ZEA, DES, and E2 are all members of different subclusters (Figure 2), which is altogether in disagreement with any intuitive bioactivity geography and also distinct from QuantMap output. Thus, structural similarity analysis of this rather small set of substances shows information on EDCs and antidiabetic substances with some degree of functional association, but with rather low resolution. In this context, it is important to underline that chemical similarity analysis, especially in highly advanced settings and with large numbers (not exactly used here), has rightly found wide application, not least in virtual screening campaigns and provides a foundation in essential parts of drug development.[70,71] Much of this recognition stems from an unparalleled speediness and consistency of molecular similarity analysis in large data sets, relative to that of various composite bioactivity mapping protocols. Nonetheless, relevant delineation of substance-mediated cellular perturbation, especially for sets of chemicals with low to modest structural similarity, needs either data on aggregated bioactivity profiles or protein target and pathway mapping, the latter area being increasingly acknowledged. The Connectivity Map concept, an outstanding example of achievements within the former area and focused on gene expression in a standardized cellular experimental setting, has already proven versatile for such endeavors.[12,13] Actually, various extensions and further development of this fundamental concept have recently appeared in the literature.[14,72,73] Future algorithms for search and characterization procedures, based on bioactivity and/or pathway involvement, may thus eventually provide better complementary tools to (modest sized) virtual screening programmes or even alternatives in some cases.[5,74] It is our perception that QuantMap can be developed to higher sophistication and automation, thereby contributing further to this evolution.

Parameter Settings in QuantMap. As outlined in Figure 1, QuantMap involves a tier of sequential procedures and is thus a composite build. Indeed, certain parameters are clearly open to refined optimization, but nonetheless underwent a testing program prior to setting. A relevant clustering pattern of an excerpt (see Methods) of the entire set of chemicals addressed in this study served as a main criterion to set substance target and closely associated proteins (primary set; STITCH3) and PPI-size (secondary set; STRING) and network signature analysis (in HUBBA). This set encompasses three very distinct substance

classes, two of those also enabling assessment by inspection of subcluster formation. These exercises ultimately arrived at 10 and 150 as suitable parameter values of the aforementioned protein numbers, respectively. Lower or higher numbers (except for the highest number combination) resulted in either aberrant (lowest numbers only) or less marked segregation between all three main substance classes or worse subcluster output. Because several parameter combinations resulted in similar cluster configurations (appearing as anticipated), we chose to proceed with the simplest (lowest value on both protein number parameters) parameter setting that still showed good overall performance to QuantMap. Notably, low numbers, especially in the primary set of proteins, may support QuantMap applicability to modestly documented substances.

Refined Interpretations of QuantMap Patterns. Proper interpretation of the QuantMap-derived perturbation similarity map is compounded by many factors, such as EDC's general display of a spectrum of agonist and antagonist properties as well as bimodal activity profiles in several cases.[75,76] Actually, most among these compounds have either estrogenic or antiestrogenic activity or both, i.e. interact with estrogen response elements (ER)α, ERβ, or GPER, or all with variable action.[77−80] Another complication relates to an array of interferences, among many of the EDCs including xenoestrogens of natural origin, with mammalian biochemical pathways apart from those directly tied to sex steroid hormone biosynthesis, signaling, and metabolism. Commonly reported EDC mechanisms involve those linked to the thyroid receptor, the peroxisome proliferator-activated receptor-γ as well as retinoid neurotransmittor, and aryl hydrocarbon receptor functions.[44,79,81,82] Perturbations outside of this range are also documented. For example, RES is reported to interfere with survivin, AP-1, and p53, as well as gap junctions and the inflammatory response; ZEA has immunomodulatory action on lymphocytes. GEN is reported to inhibit advanced glycation end product formation.[83−85] This overall view receives support from our own global distance matrix, which consistently houses proteins implicated in various intracellular actions, such as broad target transcription regulatory factors, alongside with proteins involved in sex steroid hormone homeostasis. Clearly, many cellular actions imposed by estrogen-perturbating EDCs commonly lie outside steroidal hormone regulation, which for a particular substance may be poorly documented from experimental settings or even unknown. However, it seems reasonable to apply an estrogen signaling EDC view as a major inroad to the interpretation of QuantMap findings on those test substances, which fall within this broad interaction−perturbation domain.

Apart from DBP appearing as a singleton component in the QuantMap chart, all EDCs segregate into either of two relatively closely situated major clusters (Figure 3). The smaller one, encompassing five members, brings together the principal endogenous estrogen (E2), a discontinued pharmaceutical (DES), and a mycotoxin (ZEA), as well as two anthropogenic xenoestrogens (BPA and 4-NP). The first three are all potent ERα and ERβ agonists, with a preference for the former receptor type, thereby sharing a key general feature.[76,77,86] The latter two chemicals, widely known as environmental pollutants, have relatively weak activity in nuclear ER transcription assays, but on the other hand show prompt stimulation through the GPER pathway.[79,87] Actually, binding affinities of BPA and 4-NP to GPER is comparable to that of ZEA.[88] Moreover, BPA features high similarity to DES and ZEA with regard to the global gene expression pattern in Ishikawa plus cells.[89] Seemingly, these cluster members have closely related functional properties in common in the context of estrogen signaling/perturbation. The second major EDC cluster holds two pharmaceutical SERMs (TAMOX and RALOX), a pure ER antagonist (FULV), three phytoestrogens (GEN, COUM, and RES), and an insecticide/acaricide-type agrichemical (ENDO) already banned in many countries. This assembly is seemingly more heterogeneous with regard to estrogen-associated bioactivity, in relation to that outlined above, but certain shared intracluster key properties are also evident. Notably, the SERM pharmaceuticals have well-documented ER antagonistic action in the mammary gland, jointly with agonistic interaction with estrogen signaling in various other mammalian tissues (TAMOX is agonistic in both uterus and ovary glands, whereas RALOX is agonistic in uterus only).[90] Juxtaposition of FULV, reported as a pure ER antagonist,[91] with TAMOX may seem puzzling, but there are clearly shared interferences with estrogenic signaling pathways. Actually, the gene expression profile of FULV is reported to display high correlation to several xenoestrogens.[89] Intriguingly, GEN (juxtaposed with RALOX in the QuantMap chart) has been reported to demonstrate typical SERM properties, i.e., it stimulates estrogen receptors in bone but acts antagonistically in uterus and bone marrow.[92] Perhaps the remaining phytoestrogens of this cluster, as well as ENDO, also share this feature, as suggested by the QuantMap chart. RES appears as a quite distal member of this constellation. Actually, among reported phytoestrogens, it is unusually weak with regard to ERα/ERβ activation and binding affinity as well as connected with a range of other properties, such as antiinflamatory and antithrombotic effects.[78,83] Moreover, phytoestrogens are generally perceived as being able to activate ERβ stronger than ERα and to exert pleiotropic effects on cellular signaling.[93] They have been suggested as selective estrogen enzyme modulators (SEEMs) because of their interference with estrogen signaling not only through interaction with ERs, but also by modulating steroidogenesis enzymes.[94] Concisely, the two major EDC clusters (Figure 3) can be perceived as largely representing (penchant) ERα/ERβ/GPER stimulators and (penchant) ERβ activators/SERMs/SEEMs, respectively. The appearance of DBP as a singleton cluster in QuantMap results conforms with the general perception of a quite specific pathognomonic, known as the phthalate syndrome, associated with intrauterine exposure to phthalate esters.[80,95]

The remaining five substances of this study, being outside the EDC frame, were—jointly with TAMOX, RALOX, and ZEA—used to define QuantMap parameters. Thus, their (anticipated) appearance in the chart of Figure 3 should be appreciated in this context. GLI and ROS show remote structural resemblance only and bind mostly to distinct targets, but are nonetheless perceived as largely equivalent in the clinical setting to treat type 2 diabetes mellitus, and neither is documented on EDC bioactivity.[96,97] Inferred protein−protein interaction networks of these antidiabetic drugs have been reported as being almost identical, which was construed in the context of clinical equivalence.[98] In the QuantMap chart, GLI and ROS appear as a clearly distinct cluster at the highest segregation level, a result in conformity with those of the aforementioned publication. Lastly, three NSAIDs also segregate into a separate assembly at the highest clustering level, with ASP and IBU appearing in a common subcluster, thereby all aligning with their functional distinction from both EDC and antidiabetic substances as well as intragroup relationships (Figure 3).

Further Developments. The various steps of QuantMap are clearly amenable to further improvement, including the various

parameter settings, which can be perceived as penultimate at this stage. Another direction of advancement pertains to automation of the QuantMap procedure; the current scheme involves several manual operations. Such work is, however, already underway.

CONCLUSIONS

The work described in this study has exploited already established network pharmacology principles and publicly available tools to further extend small-molecule global perturbation profiling. The overarching aim was to delineate a generic computational procedure—QuantMap—that can help predicting which compounds are most similar with respect to their systemic effects in mammalian cells. Thus, QuantMap yields quantitative perturbation scores for each pair in a predefined set of substances, eventually translating into a global distance matrix. The particular implementation of QuantMap presented and employed to 18 test substances used three publicly available Web servers. The unique feature of the QuantMap, however, lies in its tiered build on information extracted from publicly available Web servers fused with in-house based computational processing and display, providing relative bioactivity mapping of chemical substances.

QuantMap incorporates aggregated and publicly available systems biology-type data on perturbation effect(s) of a query substance into a composite framework, thereby enhancing the efficient utilization of such information for various purposes, including risk assessment of chemicals. In contrast to certain other methods, QuantMap appreciates not only protein targets but also intracellular signaling proteins, which enables a broader view to similar actions among substances.[99,100] By considering also implicit relationships between proteins, it is possible to create a rich network utilizing a great proportion of pre-existing but often neglected biomedical knowledge.

ASSOCIATED CONTENT

Supporting Information

An excerpt of results from parameter testing, involving various protein numbers derived from STITCH3 and STRING, appears as Figure 1. This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Ulf.Hammerling@slv.se (U.H.), Mats.Gustafsson@medsci.uu.se (M.G.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Swedish Civil Contingencies Agency. The authors are grateful to Dr. Ola Spjuth for valuable instructions on chemical similarity mapping. We wish to thank Mrs. Merethe Andersen for excellent assistance on artwork and Associate Professor Anders Glynn for critically reading the manuscript.

REFERENCES

(1) Lee, S.; Park, K.; Kim, D. Building a drug-target network and its applications. *Expert Opin. Drug Discovery* **2009**, *4*, 1177−1189.

(2) Yang, Y.; Adelstein, S. J.; Kassis, A. I. Target discovery from data mining approaches. *Drug Discovery Today* **2009**, *14*, 147−154.

(3) Sayers, E. W.; Barrett, T.; Benson, D. A.; Bolton, E.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; Dicuccio, M.; Federhen, S.; Feolo, M.; Geer, L. Y.; Helmberg, W.; Kapustin, Y.; Landsman, D.; Lipman, D. J.; Lu, Z.; Madden, T. L.; Madej, T.; Maglott, D. R.; Marchler-Bauer, A.; Miller, V.; Mizrachi, I.; Ostell, J.; Panchenko, A.; Pruitt, K. D.; Schuler, G. D.; Sequeira, E.; Sherry, S. T.; Shumway, M.; Sirotkin, K.; Slotta, D.; Souvorov, A.; Starchenko, G.; Tatusova, T. A.; Wagner, L.; Wang, Y.; John Wilbur, W.; Yaschenko, E.; Ye, J. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2010**, *38* (Databaseissue), D5−16.

(4) Oprea, T. I.; Tropsha, A.; Faulon, J. L.; Rintoul, M. D. Systems chemical biology. *Nat. Chem. Biol.* **2007**, *3*, 447−450.

(5) Dudley, J. T.; Deshpande, T.; Butte, A. J. Exploiting drug-disease relationships for computational drug repositioning. *Briefings Bioinf.* **2011**, *12*, 303−311.

(6) Paolini, G. V.; Shapland, R. H.; van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L. Global mapping of pharmacological space. *Nat. Biotechnol.* **2006**, *24*, 805−815.

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

(8) Nacher, J. C.; Schwartz, J. M. A global view of drug-therapy interactions. *BMC Pharmacol.* **2008**, *8*, 5.

(9) Campillos, M.; Kuhn, M.; Gavin, A. C.; Jensen, L. J.; Bork, P. Drug target identification using side-effect similarity. *Science* **2008**, *321*, 263−266.

(10) Scheiber, J.; Jenkins, J. L.; Sukuru, S. C.; Bender, A.; Mikhailov, D.; Milik, M.; Azzaoui, K.; Whitebread, S.; Hamon, J.; Urban, L.; Glick, M.; Davies, J. W. Mapping adverse drug reactions in chemical space. *J. Med. Chem.* **2009**, *52*, 3103−3107.

(11) Keiser, M. J.; Irwin, J. J.; Shoichet, B. K. The chemical basis of pharmacology. *Biochemistry* **2010**, *49*, 10267−10276.

(12) Lamb, J.; Crawford, E. D.; Peck, D.; Modell, J. W.; Blat, I. C.; Wrobel, M. J.; Lerner, J.; Brunet, J. P.; Subramanian, A.; Ross, K. N.; Reich, M.; Hieronymus, H.; Wei, G.; Armstrong, S. A.; Haggarty, S. J.; Clemons, P. A.; Wei, R.; Carr, S. A.; Lander, E. S.; Golub, T. R. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006**, *313*, 1929−1935.

(13) Lamb, J. The Connectivity Map: A new tool for biomedical research. *Nat. Rev. Cancer* **2007**, *7*, 54−60.

(14) Iorio, F.; Bosotti, R.; Scacheri, E.; Belcastro, V.; Mithbaokar, P.; Ferriero, R.; Murino, L.; Tagliaferri, R.; Brunetti-Pierri, N.; Isacchi, A.; di Bernardo, D. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. U S A* **2010**, *107*, 14621−14626.

(15) Dudley, J. T.; Sirota, M.; Shenoy, M.; Pai, R. K.; Roedder, S.; Chiang, A. P.; Morgan, A. A.; Sarwal, M. M.; Pasricha, P. J.; Butte, A. J. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med.* **2011**, *3*, 96ra76.

(16) Ideker, T.; Dutkowski, J.; Hood, L. Boosting signal-to-noise in complex biology: Prior knowledge is power. *Cell* **2011**, *144*, 860−863.

(17) Hyduke, D. R.; Palsson, B. O. Towards genome-scale signalling network reconstructions. *Nat. Rev. Genet.* **2010**, *11*, 297−307.

(18) Hopkins, A. L. Network pharmacology: The next paradigm in drug discovery. *Nat. Chem. Biol.* **2008**, *4*, 682−690.

(19) Berger, S. I.; Iyengar, R. Network analyses in systems pharmacology. *Bioinformatics* **2009**, *25*, 2466−2472.

(20) Fliri, A. F.; Loging, W. T.; Volkmann, R. A. Analysis of information flows in interaction networks: Implication for drug discovery and pharmacological research. *Discovery Med.* **2011**, *11*, 133−143.

(21) Loscalzo, J.; Kohane, I.; Barabasi, A. L. Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. *Mol. Syst. Biol.* **2007**, *3*, 124.

(22) Vidal, M.; Cusick, M. E.; Barabasi, A. L. Interactome networks and human disease. *Cell* **2011**, *144*, 986−998.

(23) Jessulat, M.; Pitre, S.; Gui, Y.; Hooshyar, M.; Omidi, K.; Samanfar, B.; Tan, L. H.; Alamgir, M.; Green, J.; Dehne, F.; Golshani, A. Recent advances in protein-protein interaction prediction: Experimental and computational methods. *Expert Opin. Drug Discov.* **2011**, *6*, 921−935.

(24) Kandasamy, K.; Mohan, S. S.; Raju, R.; Keerthikumar, S.; Kumar, G. S.; Venugopal, A. K.; Telikicherla, D.; Navarro, J. D.;

Mathivanan, S.; Pecquet, C.; Gollapudi, S. K.; Tattikota, S. G.; Mohan, S.; Padhukasahasram, H.; Subbannayya, Y.; Goel, R.; Jacob, H. K.; Zhong, J.; Sekhar, R.; Nanjappa, V.; Balakrishnan, L.; Subbaiah, R.; Ramachandra, Y. L.; Rahiman, B. A.; Prasad, T. S.; Lin, J. X.; Houtman, J. C.; Desiderio, S.; Renauld, J. C.; Constantinescu, S. N.; Ohara, O.; Hirano, T.; Kubo, M.; Singh, S.; Khatri, P.; Draghici, S.; Bader, G. D.; Sander, C.; Leonard, W. J.; Pandey, A. NetPath: A public resource of curated signal transduction pathways. *Genome Biol.* **2010**, *11*, R3.

(25) Bauer-Mehren, A.; Furlong, L. I.; Sanz, F. Pathway databases and tools for their exploitation: Benefits, current limitations and challenges. *Mol. Syst. Biol.* **2009**, *5*, 290.

(26) Klingstrom, T.; Plewczynski, D. Protein−protein interaction and pathway databases, a graphical review. *Briefings Bioinf.* **2011**, *12*, 702−713.

(27) Wallach, I.; Jaitly, N.; Lilien, R. A structure-based approach for mapping adverse drug reactions to the perturbation of underlying biological pathways. *PLoS ONE* **2010**, *5*, e12063.

(28) Kleemann, R.; Bureeva, S.; Perlina, A.; Kaput, J.; Verschuren, L.; Wielinga, P. Y.; Hurt-Camejo, E.; Nikolsky, Y.; van Ommen, B.; Kooistra, T. A systems biology strategy for predicting similarities and differences of drug effects: Evidence for drug-specific modulation of inflammation in atherosclerosis. *BMC Syst. Biol.* **2011**, *5*, 125.

(29) Judson, R. S.; Kavlock, R. J.; Setzer, R. W.; Cohen Hubal, E. A.; Martin, M. T.; Knudsen, T. B.; Houck, K. A.; Thomas, R. S.; Wetmore, B. A.; Dix, D. J. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. *Chem. Res. Toxicol.* **2011**, *24*, 451−462.

(30) Raman, K. Construction and analysis of protein−protein interaction networks. *Autom. Exp.* **2010**, *2*, 2.

(31) Audouze, K.; Juncker, A. S.; Roque, F. J.; Krysiak-Baltyn, K.; Weinhold, N.; Taboureau, O.; Jensen, T. S.; Brunak, S. Deciphering diseases and biological targets for environmental chemicals using toxicogenomics networks. *PLoS Comput. Biol.* **2010**, *6*, e1000788.

(32) Su, J.; Yoon, B. J.; Dougherty, E. R. Identification of diagnostic subnetwork markers for cancer in human protein−protein interaction network. *BMC Bioinf.* **2010**, *11* (Suppl 6), S8.

(33) Suthram, S.; Dudley, J. T.; Chiang, A. P.; Chen, R.; Hastie, T. J.; Butte, A. J. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Comput. Biol.* **2010**, *6*, e1000662.

(34) Klipp, E.; Wade, R. C.; Kummer, U. Biochemical network-based drug-target prediction. *Curr. Opin. Biotechnol.* **2010**, *21*, 511−516.

(35) Szklarczyk, D.; Franceschini, A.; Kuhn, M.; Simonovic, M.; Roth, A.; Minguez, P.; Doerks, T.; Stark, M.; Muller, J.; Bork, P.; Jensen, L. J.; von Mering, C. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **2011**, *39* (Database issue), D561−568.

(36) Chen, J. Y.; Mamidipalli, S.; Huan, T. HAPPI: An online database of comprehensive human annotated and predicted protein interactions. *BMC Genomics* **2009**, *10* (Suppl 1), S16.

(37) Kuhn, M.; Szklarczyk, D.; Franceschini, A.; von Mering, C.; Jensen, L. J.; Bork, P. STITCH 3: Zooming in on protein−chemical interactions. *Nucleic Acids Res.* **2012**, *40* (Database issue), D876−880.

(38) Pavlopoulos, G. A.; Secrier, M.; Moschopoulos, C. N.; Soldatos, T. G.; Kossida, S.; Aerts, J.; Schneider, R.; Bagos, P. G. Using graph theory to analyze biological networks. *BioData Min.* **2011**, *4*, 10.

(39) Brohee, S.; Faust, K.; Lima-Mendez, G.; Sand, O.; Janky, R.; Vanderstocken, G.; Deville, Y.; van Helden, J. NeAT: a toolbox for the analysis of biological networks, clusters, classes and pathways. *Nucleic Acids Res.* **2008**, *36* (Web server issue), W444−451.

(40) Lin, C. Y.; Chin, C. H.; Wu, H. H.; Chen, S. H.; Ho, C. W.; Ko, M. T. Hubba: Hub objects analyzer—a framework of interactome hubs identification for network biology. *Nucleic Acids Res.* **2008**, *36* (Web Server issue), W438−443.

(41) Wu, J.; Vallenius, T.; Ovaska, K.; Westermarck, J.; Makela, T. P.; Hautaniemi, S. Integrated network analysis platform for protein−protein interactions. *Nat. Methods* **2009**, *6*, 75−77.

(42) Mueller, L. A.; Kugler, K. G.; Dander, A.; Graber, A.; Dehmer, M. QuACN: An R package for analyzing complex biological networks quantitatively. *Bioinformatics* **2011**, *27*, 140−141.

(43) Kuhn, M.; von Mering, C.; Campillos, M.; Jensen, L. J.; Bork, P. STITCH: Interaction networks of chemicals and proteins. *Nucleic Acids Res.* **2008**, *36* (Databaseissue), D684−688.

(44) Diamanti-Kandarakis, E.; Bourguignon, J. P.; Giudice, L. C.; Hauser, R.; Prins, G. S.; Soto, A. M.; Zoeller, R. T.; Gore, A. C. Endocrine-disrupting chemicals: An Endocrine Society scientific statement. *Endocr. Rev.* **2009**, *30*, 293−342.

(45) Cheng, T.; Wang, Y.; Bryant, S. H. Investigating the correlations among the chemical structures, bioactivity profiles and molecular targets of small molecules. *Bioinformatics* **2010**, *26*, 2881−2888.

(46) PubChem. http://pubchem.ncbi.nlm.nih.gov/assay/?p=clustering/ (accessed February 8, 2012).

(47) Jensen, L. J.; Kuhn, M.; Stark, M.; Chaffron, S.; Creevey, C.; Muller, J.; Doerks, T.; Julien, P.; Roth, A.; Simonovic, M.; Bork, P.; von Mering, C. STRING 8: A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **2009**, *37* (Database issue), D412−416.

(48) Chin, C. S.; Samanta, M. P. Global snapshot of a protein interaction network-a percolation based approach. *Bioinformatics* **2003**, *19*, 2413−2419.

(49) Kendall, M. *Rank Correlation Methods*; Griffin: London, 1970.

(50) Jurman, G.; Merler, S.; Barla, A.; Paoli, S.; Galea, A.; Furlanello, C. Algebraic stability indicators for ranked lists in molecular profiling. *Bioinformatics* **2008**, *24*, 258−64.

(51) Fagin, R.; Kumar, R.; Sivakumar, R. Comparing top k lists. *SIAM J. Discrete Math.* **2003**, *17*, 134−160.

(52) Lamrous, S.; Taileb, M. Divisive Hierarchical K-Means. In International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), Sidney, Australia, Nov. 28−Dec. 06, 2006; IEEE Computer Society, 2007, *c1*, p 18.

(53) Maddah, F.; Soeria-Atmadja, D.; Malm, P.; Gustafsson, M. G.; Hammerling, U. Interrogating health-related public databases from a food toxicology perspective: Computational analysis of scoring data. *Food Chem. Toxicol.* **2011**, *49*, 2830−2840.

(54) Willett, P. Similarity-based data mining in files of two-dimensional chemical structures using fingerprint measures of molecular resemblance. *WIREs Comput. Mol. Sci.* **2011**, *1*, 241−251.

(55) Edwards, S. W.; Preston, R. J. Systems biology and mode of action based risk assessment. *Toxicol. Sci.* **2008**, *106*, 312−318.

(56) Yildirim, M. A.; Vidal, M. Systems engineering to systems biology. *Mol. Syst. Biol.* **2008**, *4*, 185.

(57) Wishart, D. S. Bioinformatics in drug development and assessment. *Drug Metab. Rev.* **2005**, *37*, 279−310.

(58) Chen, B.; Dong, X.; Jiao, D.; Wang, H.; Zhu, Q.; Ding, Y.; Wild, D. J. Chem2Bio2RDF: A semantic framework for linking and data mining chemogenomic and systems chemical biology data. *BMC Bioinf.* **2010**, *11*, 255.

(59) Reif, D. M.; Martin, M. T.; Tan, S. W.; Houck, K. A.; Judson, R. S.; Richard, A. M.; Knudsen, T. B.; Dix, D. J.; Kavlock, R. J. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. *Environ. Health Perspect.* **2010**, *118*, 1714−1720.

(60) Judson, R. S.; Houck, K. A.; Kavlock, R. J.; Knudsen, T. B.; Martin, M. T.; Mortensen, H. M.; Reif, D. M.; Rotroff, D. M.; Shah, I.; Richard, A. M.; Dix, D. J. In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project. *Environ. Health Perspect.* **2010**, *118*, 485−492.

(61) Weinstein, J. N.; Myers, T. G.; O'Connor, P. M.; Friend, S. H.; Fornace, A. J., Jr; Kohn, K. W.; Fojo, T.; Bates, S. E.; Rubinstein, L. V.; Anderson, N. L.; Buolamwini, J. K.; van Osdol, W. W.; Monks, A. P.; Scudiero, D. A.; Sausville, E. A.; Zaharevitz, D. W.; Bunow, B.; Viswanadhan, V. N.; Johnson, G. S.; Wittes, R. E.; Paull, K. D. An information-intensive approach to the molecular pharmacology of cancer. *Science* **1997**, *275*, 343−349.

(62) Holbeck, S. L.; Collins, J. M.; Doroshow, J. H. Analysis of Food and Drug Administration-approved anticancer agents in the NCI60 panel of human tumor cell lines. *Mol. Cancer Ther.* **2010**, *9*, 1451−1460.
(63) Rhodes, D. R.; Kalyana-Sundaram, S.; Tomlins, S. A.; Mahavisno, V.; Kasper, N.; Varambally, R.; Barrette, T. R.; Ghosh, D.; Varambally, S.; Chinnaiyan, A. M. Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia* **2007**, *9*, 443−454.
(64) Berg, E. L.; Kunkel, E. J.; Hytopoulos, E. Biological complexity and drug discovery: A practical systems biology approach. *Syst. Biol. (Stevenage)* **2005**, *152*, 201−206.
(65) Wagner, B. K.; Clemons, P. A. Connecting synthetic chemistry decisions to cell and genome biology using small-molecule phenotypic profiling. *Curr. Opin. Chem. Biol.* **2009**, *13*, 539−548.
(66) Venkatesan, K.; Rual, J. F.; Vazquez, A.; Stelzl, U.; Lemmens, I.; Hirozane-Kishikawa, T.; Hao, T.; Zenkner, M.; Xin, X.; Goh, K. I.; Yildirim, M. A.; Simonis, N.; Heinzmann, K.; Gebreab, F.; Sahalie, J. M.; Cevik, S.; Simon, C.; de Smet, A. S.; Dann, E.; Smolyar, A.; Vinayagam, A.; Yu, H.; Szeto, D.; Borick, H.; Dricot, A.; Klitgord, N.; Murray, R. R.; Lin, C.; Lalowski, M.; Timm, J.; Rau, K.; Boone, C.; Braun, P.; Cusick, M. E.; Roth, F. P.; Hill, D. E.; Tavernier, J.; Wanker, E. E.; Barabasi, A. L.; Vidal, M. An empirical framework for binary interactome mapping. *Nat. Methods* **2009**, *6*, 83−90.
(67) Hase, T.; Tanaka, H.; Suzuki, Y.; Nakagawa, S.; Kitano, H. Structure of protein interaction networks and their implications on drug design. *PLoS Comput. Biol.* **2009**, *5*, e1000550.
(68) Yu, H.; Kim, P. M.; Sprecher, E.; Trifonov, V.; Gerstein, M. The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS Comput. Biol.* **2007**, *3*, e59.
(69) Hwang, W. C.; Zhang, A.; Ramanathan, M. Identification of information flow-modulating drug targets: A novel bridging paradigm for drug discovery. *Clin. Pharmacol. Ther.* **2008**, *84*, 563−572.
(70) Eckert, H.; Bajorath, J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discovery Today* **2007**, *12*, 225−233.
(71) Willett, P. Similarity searching using 2D structural fingerprints. *Methods Mol. Biol.* **2011**, *672*, 133−158.
(72) Zhang, S. D.; Gant, T. W. A simple and robust method for connecting small-molecule drugs using gene-expression signatures. *BMC Bioinf* **2008**, *9*, 258.
(73) Sirota, M.; Dudley, J. T.; Kim, J.; Chiang, A. P.; Morgan, A. A.; Sweet-Cordero, A.; Sage, J.; Butte, A. J. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* **2011**, *3*, 96ra77.
(74) Lussier, Y. A.; Chen, J. L. The emergence of genome-based drug repositioning. *Sci. Transl. Med.* **2011**, *3*, 96ps35.
(75) Paris, F.; Balaguer, P.; Terouanne, B.; Servant, N.; Lacoste, C.; Cravedi, J. P.; Nicolas, J. C.; Sultan, C. Phenylphenols, biphenols, bisphenol-A and 4-tert-octylphenol exhibit alpha and beta estrogen activities and antiandrogen activity in reporter cell lines. *Mol. Cell. Endocrinol.* **2002**, *193*, 43−49.
(76) Mueller, S. O.; Simon, S.; Chae, K.; Metzler, M.; Korach, K. S. Phytoestrogens and their human metabolites show distinct agonistic and antagonistic properties on estrogen receptor alpha (ERalpha) and ERbeta in human cells. *Toxicol. Sci.* **2004**, *80*, 14−25.
(77) Leffers, H.; Naesby, M.; Vendelbo, B.; Skakkebaek, N. E.; Jorgensen, M. Oestrogenic potencies of Zeranol, oestradiol, diethylstilboestrol, Bisphenol-A and genistein: Implications for exposure assessment of potential endocrine disrupters. *Hum. Reprod.* **2001**, *16*, 1037−1045.
(78) Matsumura, A.; Ghosh, A.; Pope, G. S.; Darbre, P. D. Comparative study of oestrogenic properties of eight phytoestrogens in MCF7 human breast cancer cells. *J. Steroid Biochem. Mol. Biol.* **2005**, *94*, 431−443.
(79) Bonefeld-Jorgensen, E. C.; Long, M.; Hofmeister, M. V.; Vinggaard, A. M. Endocrine-disrupting potential of bisphenol A, bisphenol A dimethacrylate, 4-n-nonylphenol, and 4-n-octylphenol in vitro: New data and a brief review. *Environ. Health Perspect.* **2007**, *115* (Suppl 1), 69−76.
(80) Wilson, V. S.; Blystone, C. R.; Hotchkiss, A. K.; Rider, C. V.; Gray, L. E., Jr. Diverse mechanisms of anti-androgen action: impact on male rat reproductive tract development. *Int. J. Androl.* **2008**, *31*, 178−187.
(81) Martin, M. T.; Dix, D. J.; Judson, R. S.; Kavlock, R. J.; Reif, D. M.; Richard, A. M.; Rotroff, D. M.; Romanov, S.; Medvedev, A.; Poltoratskaya, N.; Gambarian, M.; Moeser, M.; Makarov, S. S.; Houck, K. A. Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. *Chem. Res. Toxicol.* **2010**, *23*, 578−590.
(82) Shanle, E. K.; Xu, W. Endocrine disrupting chemicals targeting estrogen receptor signaling: Identification and mechanisms of action. *Chem. Res. Toxicol.* **2011**, *24*, 6−19.
(83) Shakibaei, M.; Harikumar, K. B.; Aggarwal, B. B. Resveratrol addiction: To die or not to die. *Mol. Nutr. Food Res.* **2009**, *53*, 115−128.
(84) Severino, L.; Russo, R.; Luongo, D.; De Luna, R.; Ciarcia, R.; Rossi, M. Immune effects of four Fusarium-toxins (FB1, ZEA, NIV, DON) on the proliferation of Jurkat cells and porcine lymphocytes: In vitro study. *Vet. Res. Commun.* **2008**, *32* (Suppl 1), S311−313.
(85) Lv, L.; Shao, X.; Chen, H.; Ho, C. T.; Sang, S. Genistein inhibits advanced glycation end product formation by trapping methylglyoxal. *Chem. Res. Toxicol.* **2011**, *24*, 579−586.
(86) le Maire, A.; Bourguet, W.; Balaguer, P. A structural view of nuclear hormone receptor: Endocrine disruptor interactions. *Cell. Mol. Life Sci.* **2010**, *67*, 1219−1237.
(87) Watson, C. S.; Jeng, Y. J.; Kochukov, M. Y. Nongenomic signaling pathways of estrogen toxicity. *Toxicol. Sci.* **2010**, *115*, 1−11.
(88) Thomas, P.; Dong, J. Binding and activation of the seven-transmembrane estrogen receptor GPR30 by environmental estrogens: A potential novel mechanism of endocrine disruption. *J. Steroid Biochem. Mol. Biol.* **2006**, *102*, 175−179.
(89) Boehme, K.; Simon, S.; Mueller, S. O. Gene expression profiling in Ishikawa cells: A fingerprint for estrogen active compounds. *Toxicol. Appl. Pharmacol.* **2009**, *236*, 85−96.
(90) Swaby, R. F.; Sharma, C. G.; Jordan, V. C. SERMs for the treatment and prevention of breast cancer. *Rev. Endocr. Metab. Disord.* **2007**, *8*, 229−239.
(91) Martin, L. A.; Pancholi, S.; Chan, C. M.; Farmer, I.; Kimberley, C.; Dowsett, M.; Johnston, S. R. The anti-oestrogen ICI 182,780, but not tamoxifen, inhibits the growth of MCF-7 breast cancer cells refractory to long-term oestrogen deprivation through down-regulation of oestrogen receptor and IGF signalling. *Endocr. Relat. Cancer* **2005**, *12*, 1017−1036.
(92) Erlandsson, M. C.; Islander, U.; Moverare, S.; Ohlsson, C.; Carlsten, H. Estrogenic agonism and antagonism of the soy isoflavone genistein in uterus, bone and lymphopoiesis in mice. *Apmis* **2005**, *113*, 317−323.
(93) Lorand, T.; Vigh, E.; Garai, J. Hormonal action of plant derived and anthropogenic non-steroidal estrogenic compounds: phytoestrogens and xenoestrogens. *Curr. Med. Chem.* **2010**, *17*, 3542−3574.
(94) Basly, J. P.; Lavier, M. C. Dietary phytoestrogens: Potential selective estrogen enzyme modulators? *Planta Med.* **2005**, *71*, 287−294.
(95) Foster, P. M. Disruption of reproductive development in male rat offspring following in utero exposure to phthalate esters. *Int. J. Androl.* **2006**, *29*, 140−7 ; discussion, p 181−185.
(96) Pfutzner, A.; Wilhelm, B.; Forst, T. Rosiglitazone and glimeperide: Review of clinical results supporting a fixed dose combination. *Vasc. Health Risk. Manag.* **2007**, *3*, 211−220.
(97) Yoon, K. H.; Shin, J. A.; Kwon, H. S.; Lee, S. H.; Min, K. W.; Ahn, Y. B.; Yoo, S. J.; Ahn, K. J.; Park, S. W.; Lee, K. W.; Sung, Y. A.; Park, T. S.; Kim, M. S.; Kim, Y. K.; Nam, M. S.; Kim, H. S.; Park, Ie, B.; Park, J. S.; Woo, J. T.; Son, H. Y. Comparison of the efficacy of glimepiride, metformin, and rosiglitazone monotherapy in korean drug-naive type 2 diabetic patients: the practical evidence of antidiabetic monotherapy study. *Diabetes Metab. J.* **2011**, *35*, 26−33.

(98) Fliri, A. F.; Loging, W. T.; Volkmann, R. A. Drug effects viewed from a signal transduction network perspective. *J. Med. Chem.* **2009**, *52*, 8038−8046.
(99) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* **2007**, *25*, 197−206.
(100) Nigsch, F.; Mitchell, J. B. Toxicological relationships between proteins obtained from protein target predictions of large toxicity databases. *Toxicol. Appl. Pharmacol.* **2008**, *231*, 225−234.