Analysis and Development of Database Resources for the Integration of Transporter Protein Research Across Biological Disciplines with an Emphasis on Vacuolar Transporters in Plants

By

MATTHEW E MICHEL
THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Agricultural and Environmental Chemistry

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Sascha CT Nicklisch, Chair

Steven M Theg

Oliver Fiehn

Committee in Charge

2022

Abstract

Transmembrane transporter proteins facilitate an essential biological role in all living organisms by regulating the movement of substances across lipid membranes. The ABC (ATPase-binding cassette) and SLC (solute carrier) transporter superfamilies are a large, ubiquitous, and highly conserved subset of active transporters across all domains of life. These superfamilies are known to play a key role in xenobiotic defense mechanisms protecting cells from potentially harmful chemicals, known as multixenobiotic resistance (MXR), multidrug resistance (MDR) in cancer cells, and the transport of numerous endogenous substances.

Due to their key influence on the absorption and distribution of chemicals, ABC and SLC transporters are of critical interest in pharmacology, with many resources devoted to characterizing drug-transporter interactions including databases like UCSF-FDA TransPortal. However, they are also of critical importance in other fields including aquatic toxicology, plant biology, nutrition, and microbiology. Information on transporter-chemical interactions from across such disciplines is currently disorganized and not contained in any publicly accessible database. Consequently, transporters are studied as a niche subtopic within other fields of study instead of as a self-sufficient discipline with relevance to many others. This thesis aims to alter that dynamic.

First in this work, an analysis of the newly updated UCSF-FDA TransPortal + UCSD/UCD-NIEHS TICBase is given, elucidating trends in drug transporter research and making recommendations for the field. Second, we present a new database, UC Transportal, an expansive data resource meant to integrate transporter research from across biological disciplines and become a central repository for all ABC and SLC transporter interaction data. Third, a method development project to isolate intact vacuoles from *Arabidopsis thaliana* leaves is described to support plant vacuolar transport studies, with glycosylated monolignol transport as a topic of primary interest.

Table of Contents

Introduction

Transmembrane transport proteins play a critical role in one of the most basic cellular processes in all living organisms: they facilitate the transport of substances across lipid membrane barriers which form the boundaries of cells and the compartments within them (Almén et al. 2009; Sahoo et al. 2014). One subcategory among this extremely broad class of proteins is primary and secondary active transporters, defined by their ability to transport molecules against a concentration gradient using chemical energy (from a molecule such as ATP) or energy from an electrochemical gradient respectively (Higgins 2001; Cho et al. 2014). As integrated constituents of the cell membrane, these proteins interact directly with substances to move them from one side of the membrane to the other. The nature of such interactions can vary greatly, from rapid transport of a substrate across the membrane to strong binding interactions which inhibit or modify transporter activity. Active transporters are responsible for the net partitioning of substances between compartments in a manner that is beneficial to the organism at the cellular and histological levels. dedicated transporters for endogenous and exogenous compounds. Among those, most transporters are selective for a one or a few specific substrates. For example, there are multiple known transporters which recognize only glucose and are responsible for its transport (Thorens and Mueckler 2010). A certain subset of transporters termed “multidrug resistance” or MDR transporters respond to a host of exogenous molecules to remove them from the cell (Gottesman et al. 2002; Nigam 2015; Giacomini and Sugiyama 2017; Nigam 2018). MDR-type transporters are members of the solute carrier (SLC) and ATP-binding Cassette (ABC) family of membrane transporters and are key determinants of xenobiotic uptake in all organisms. The molecular mechanisms of how these cellular gatekeepers can recognize, bind and eliminate a broad spectrum of structurally unrelated xenobiotics is still unknown and under continuous investigation in

pharmacology and toxicology research fields. Regardless of their intended substrates, the sheer number of compounds that transporters encounter is enormous.

Researchers therefore have a monumental task when faced with characterizing and cataloguing these interactions. Considering just a single transporter-chemical interaction, researchers may choose to examine whether the chemical is a substrate, whether it inhibits or stimulates the transport of other accepted substrates, whether orthosteric or allosteric binding occurs, or the kinetic parameters of any observed phenomena. Due to this immense scope, researchers have necessarily limited the breadth of their studies to those interactions which are deemed most critical to the public interest, especially regarding human health outcomes. As a result, a substantial proportion of transporter-chemical interaction studies focus squarely on the pharmacological effects of drugs on human transporters to determine the efficacy and safety of medical treatments. Due to the upregulation of drug efflux transporters in many cancers, extensive resources have also been devoted to studying the modulation of their activity to develop targeted chemotherapeutic treatments for cancer patients (Gottesman et al. 2002; Xu et al. 2007; Robey et al. 2018). The pharmaceutical focus in MDR transporter research can be clearly seen in the literature, while data on interactions concerning all other organisms and non-pharmaceutical compounds are sparse and disorganized by comparison.

In recent years, researchers have begun to broaden the scope of transporter protein research to include interactions with food and herb compounds, natural chemicals produced by plants and fungi, nanoparticles, and anthropogenic environmental chemicals such as flame retardants and pesticides (Hu et al. 2005; Marchetti et al. 2007; Li et al. 2012; Nicklisch and Hamdoun 2020) There is also a desire to understand the relationship between chemical exposures and transporter gene expression, and at a structural level the differences between transporter orthologs and

polymorphic variants (Giacomini et al. 2013; Zou et al. 2018; Cheung et al. 2019) and their effects on interactions with xenobiotic chemicals. Due to the key cellular function they mediate and their widespread conservation across all domains of life, transporter-chemical interactions are relevant in a wide range of fields, including nutritional science, plant biology, veterinary medicine, microbiology, and environmental toxicology. In contrast to pharmacological studies however, studies involving transporters in these fields are generally treated as particular to niche topics within those fields and not as part of a larger compendium of research in which transporter proteins are considered as the main subject. This may cause active transporters to be overlooked in many areas of study.

Due to their fundamental role in biology, there is an enormous amount of potential utility in expanding our knowledge of transporter proteins in nonhuman organisms and of transporter interactions with non-pharmaceutical chemicals. The ability to compare interaction data in model mammalian organisms with interactions in humans can directly improve our ability to predict pharmacological phenomena affected by active transporters. Understanding how transporters in organisms living in environments vulnerable to chemical contamination (such as aquatic environments) handle xenobiotics is important to determining risks of ecological damage. Transporters could also be targets for genetic engineering or molecular breeding strategies in pest control and agriculture. Ultimately, bringing attention to a wider range of transporter-chemical interaction data will help us understand their roles in a wider variety of health-related outcomes in medical, environmental, and other contexts.

Rationale

This project aims to bring awareness to current trends in membrane transporter research, with a focus on ABC and SLC superfamily transporters, and create a long-lasting, inclusive central resource for all transporter researchers that promotes collaboration and dissemination of new data.

In chapter one, we present a descriptive analysis of the newly updated UCSF-FDA TransPortal + UCSD/UCD-NIEHS TICBase online database. The UCSF-FDA TransPortal has been a prominent resource describing interactions between drug transporters and pharmaceuticals since its creation in 2012. Its recent update, which greatly expanded its contents, provides an opportunity for a view into the state of the literature on transporter-chemical interactions. We analyze the database in its entirety and point out trends from the established literature which will inform researchers and regulators about the needs of the field. We also pay special attention to a subset of data, called TICBase, on environmental chemicals which interfere with drug transporter activity, to highlight their relevance to transporter research.

In chapter two, we introduce UC Transportal: a comprehensive database for ABC and SLC transporter superfamily proteins from all organisms and their interactions with small molecules, including drugs and environmental chemicals. UC Transportal is meant to become a central hub where several databases pertaining to different classes of transporter data come together. We focus on the first of these databases, the Kinetic Transportal, which describes direct transporter-chemical interactions, and the design, layout, and administration of UC Transportal which will serve as a framework for adding new data and databases through future collaborative efforts.

In chapter three, we describe a method development project to improve existing protocols for isolating vacuoles from leaf cells of the model organism *Arabidopsis thaliana* to pursue an

investigation of how glycosylated monolignols (precursors to the lignin polymer) move across the vacuolar membrane. This foray into a very specific transporter-mediated phenomenon is meant to support further research into vacuolar transport proteins, which are an understudied topic considering their important role in plant biology. It is also meant to showcase the inclusivity of UC Transportal by bridging the fields of vacuolar plant transporter studies with pharmacological and toxicological transporter research. The aim to develop a more efficient and accessible protocol for obtaining intact plant vacuoles will make vacuolar plant transporter studies more amenable for researchers who can produce data relevant to the mission of the UC Transportal database.

Chapter 1: Into the Transportal – Analysis and Insights from the Updated UCSF-FDA TransPortal + UCSD/UCD-NIEHS TICBase

Abstract

Drug transporters are part of the cell's xenobiotic defense system, key determinants in the uptake and elimination of small molecules, and central to understanding the biological basis for bioaccumulation of environmental chemicals and determining safe levels of administered pharmaceuticals (Giacomini and Huang 2013; Giacomini and Sugiyama 2006; Schlessinger et al. 2013). Interactions of these drugs and chemicals in the body can influence the effective dosages for drug treatment as well as safe levels of environmental exposure. A recent update to the original TransPortal database (Morrissey et al. 2012), a central repository of drug transporter interactions with pharmaceuticals, expanded the panel of kinetic interactions to include rodent and primate transporters as well as the imminent role of transporter-active environmental chemicals. Here, we present a thorough statistical analysis and graphical summary of the current database (http://transportaldev.docking.org). We highlight the importance of identifying levels and interaction potencies of Transporter-Interfering Chemicals (TICs) among the plethora of legacy and emerging chemicals and the challenges in evaluating drug-environmental chemical interactions (DECI) that could serve in clinical and regulatory decision support for improved patient care.

1. Introduction

Drug transporters are key determinants for xenobiotic uptake and distribution in all organisms, including humans (Gottesman et al. 2002; Dean and Annilo 2005; The International Transporter Consortium et al. 2010). Since the discovery of the first multidrug resistance (MDR) protein in 1978, there have been many studies on the interactions of drugs with these cellular

gatekeeper proteins (Shapiro and Ling 1994). One of the best characterized MDR proteins is the mammalian efflux transporter ABCB1 or P-glycoprotein (P-gp). Besides P-gp, many uptake transporters of the SLC superfamily have been identified in conferring MDR (Fardel et al. 2012; Nigam 2015; Nigam 2018)

Understanding to what extent xenobiotic transporters interact with drugs is of prime importance to understanding safe dosing regimens since drug-drug interactions can reduce or increase the effective concentration of an administered drug. Drug safety assessments are further complicated by the myriad of transporter-active environmental chemicals that are regularly detected in food and drinking water (Schecter et al. 2001; Schecter et al. 2001; Hites et al. 2004; Schecter, Haffner, et al. 2010; Schecter, Colacino, et al. 2010). We have recently shown that these so called Transporter-Interfering Chemicals or TICs can bind to drug transporters and inhibit their protective efflux function (Nicklisch et al. 2016; Nicklisch and Hamdoun 2020; Nicklisch et al. 2021). Notably, the levels of these chemicals in food sources can vary dramatically (Nicklisch et al. 2017a; Nicklisch et al. 2017b) and the effects of unintentional exposure to one or multiple transporter active chemicals are unknown.

The UCSF-FDA TransPortal database was announced in 2012 as a free and publicly accessible resource to inform the scientific community of transporter proteins and their interactions with pharmaceuticals affecting drug disposition (Morrissey et al. 2012). It provided kinetic interaction and human tissue expression data on 31 drug transporters extracted from more than 297 primary literature sources, and has since served as a central location for information on drug transporters and the inhibitors and substrates associated with them (Morrissey et al. 2012). In the ten years since then, significant advancements have been made in the field of drug transporter research. Transporter-xenobiotic interactions have become more widely recognized as potentially

significant determinants of health outcomes, not only by way of drug-drug interactions (DDI), but also via uncontrolled exposures and potential drug-environmental chemical interactions (DECI) from ubiquitous food-borne and anthropogenic chemicals.

The recent major update to the new UCSF-FDA Transportal + UCSD/UCD-NIEHS TICBase (https://transportal.compbio.ucsf.edu/) more than quadrupled the amount of mammalian transporter-chemical interaction data and now serves as a comprehensive reference website for the broader scientific community of environmental toxicologists, structural biologists, systems biologists, and bioinformaticians, as well as professionals in the pharmaceutical industry, regulatory agencies, and clinicians. A subset of the data has been classified as TICBase (Transporter-Interfering Chemicals database) to highlight the emerging status of anthropogenic environmental chemicals and their interactions with drug transporters. In this analysis we summarize the contents of the updated TransPortal database, which now includes 46 transporters and data from nearly 600 (n = 592) primary research articles. We point out several trends in transporter research that warrant attention from investigators in the field and assess the state of our knowledge about transporter-chemical interactions with a special emphasis on TICs.

2. Methods

2.1. General database formatting and nomenclature

Primary research articles were searched for K_i, IC_{50}, and K_m values pertaining to transporter-chemical interactions. The K_i value is a dissociation constant serving as a measure of binding affinity between the chemical and transporter. IC_{50} values are defined as the concentration at which transporter activity is reduced to 50% of uninhibited levels. The K_m value is the Michaelis-Menten constant, the concentration at which a substrate is transported at 50% of maximal transporter activity. All data was derived from *in vitro* assays on transporters from model

mammalian organisms including human (*Homo sapiens*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), and monkey (*Chlorocebus aethiops)*. The data collected pertaining to each transporter-chemical interaction is as follows: transporter name, chemical name, quantitative kinetic value (K_i, IC_{50}, or K_m), *in vitro* assay system, reporter molecule or effect, and reference number. The ABC, SLC and SLCO naming systems were used (according to the HUGO Gene Nomenclature Committee at the European Bioinformatics Institute, https://www.genenames.org/) to standardize the protein nomenclature. Homologous proteins from different organisms, such as mouse Abcb1a and human ABCB1, are indexed as different proteins in the database and their respective data are displayed on separate pages. When applicable, chemical nomenclature regarding geometric (i.e., ortho, para) and optical (i.e., -, +) stereoisomerism was accounted for. See Appendix 1, section 1, for more detailed information on the data collection process.

2.2. Data classification and normalization

2.2.1. Data clean-up

Inhibitory interactions are those associated with K_i or IC_{50} values and substrate interactions are those associated with K_m values. Values reported as an approximation by the investigator are denoted by a tilde (~) and values reported only as bounded by an upper or lower limit are denoted by a greater than (>) or lesser than (<) sign in the database. Approximated and bounded values were excluded from the statistical analyses.

Prior to the update, the original UCSF-FDA TransPortal database (https://transportal.compbio.ucsf.edu/) contained 1402 records of chemical interactions, plus gene expression and DDI data, exclusively with human ABC- and SLC-type drug transporter proteins (Morrissey et al. 2012). The new joint UCSF-FDA TransPortal + UCSD/UCD-NIEHS TICBase (http://transportaldev.docking.org) more than quadruples the content and now contains a total of

6296 interactions plus gene expression and DDI data from a total of 592 primary research articles published from 1992 to 2022 (Table S1-1). Out of these 6380 interactions, approximately one third (n = 2222) are bounded values (having a greater than or lesser than sign), of which seven are K_i values, four are K_m values, and the remaining 2211 are IC_{50} values. Of the bounded IC_{50} values, almost all (n = 2126) were obtained from a single study (Morgan et al. 2013). In that study, the authors used a high-throughput screening technique to assess a total of 2500 interactions and assigned a lower IC_{50} boundary to interactions in which the compound had no inhibitory effect at the concentrations tested. The TransPortal data also includes an additional 12 approximate IC_{50} values and 58 interactions with a "nondetermined" (ND) value for observations which only determined an interaction type (inhibition or substrate). Of these qualitative (ND) interactions, 16 are inhibition and 42 are substrate interactions. See Table S1-2 for additional summary statistics.

When subtracting all bounded, approximate, and "ND"-labeled values from the total, the updated TransPortal database contains 4004 transporter-chemical interactions with "hard values". Of these, 566 are K_i values, 2780 are IC_{50} values, and 742 are K_m values (84 interactions list both an IC_{50} and a K_i value). These 4004 interactions were used exclusively in our analysis of the data. For this reason, all data pertaining to the transporter ABCB4 was excluded, because only bounded values were associated with it, bringing the final number of transporters in our analysis to 45.

2.2.2. Assay environment

Cell systems were grouped together based on the cell type they were derived from, such as the widely recognized HEK293(Human embryonic kidney) and Sf9(Spodoptera frugiperda) cell types. Membrane environment types were grouped into five distinct categories, including crude membranes, vesicles, cells, monolayers, or liposomes, however, approximately 51% of interactions did not clearly define the membrane environment type. Therefore, our analysis of

"assay environment types" focuses on "Cell System" data, referring to the environment cell type but not to the specific membrane environment (vesicles, crude membranes etc.) or to the method of expressing the transporter protein.

2.2.3. Chemical classification

For the purposes of this analysis, we defined environmental chemicals as chemicals which are man-made (anthropogenic), are or have previously been used in consumer or industrial applications (or byproducts thereof), but which are not intended for human exposure. These are distinguished from pharmaceuticals, endogenous compounds in animals, natural chemicals from the environment such as plant or fungal compounds, and substances found in food or supplements.

2.3. Statistical and Graphical Analysis

All data analysis was performed using RStudio version 2021.09.2+382 for Windows and Microsoft Excel. Non-base R packages used were "dplyr", "tidyverse", and "ggplot2". Box and whisker plots were created using RStudio while all other graphs and tables were prepared in Excel.

3. Results

3.1. Distribution of interaction parameters

The most potent inhibitors by K_i value are the three antihypertensive agents pratosartan, losartan, and telmisartan, acting on SLC22A12 with values of .0067, .0077 and .0182µM. The most potent inhibitors by IC_{50} value are four the HIV antivirals emtricitabine, abacavir, zidovudine and tenofovir, acting on SLC22A1 and ranging from 20 to 900pM (2.0E-8 – 9.0E-7 mM). By contrast, the highest K_i value belongs to the antibiotic cefadroxil acting on SLC22A8 and the highest IC_{50} value belongs to the monoamine oxidase inhibitor tranylcypromine acting on ABCB1. The lowest K_m value recorded is 760pM for bilirubin transported by SLCO1B1 and the highest is 8.95mM for the antiviral agent oseltamivir transported by SLC15A1.

The distributions of the K_i, IC_{50}, and K_m values are all weighted towards the low ends of their concentration ranges. In each case, the larger the values become, the sparser they become in exponential fashion. Figure 1 displays the distribution of each kinetic parameter, indicating that the total range of IC_{50} values spans over nine orders of magnitude, while the K_i and K_m values span six and seven orders of magnitude, respectively.

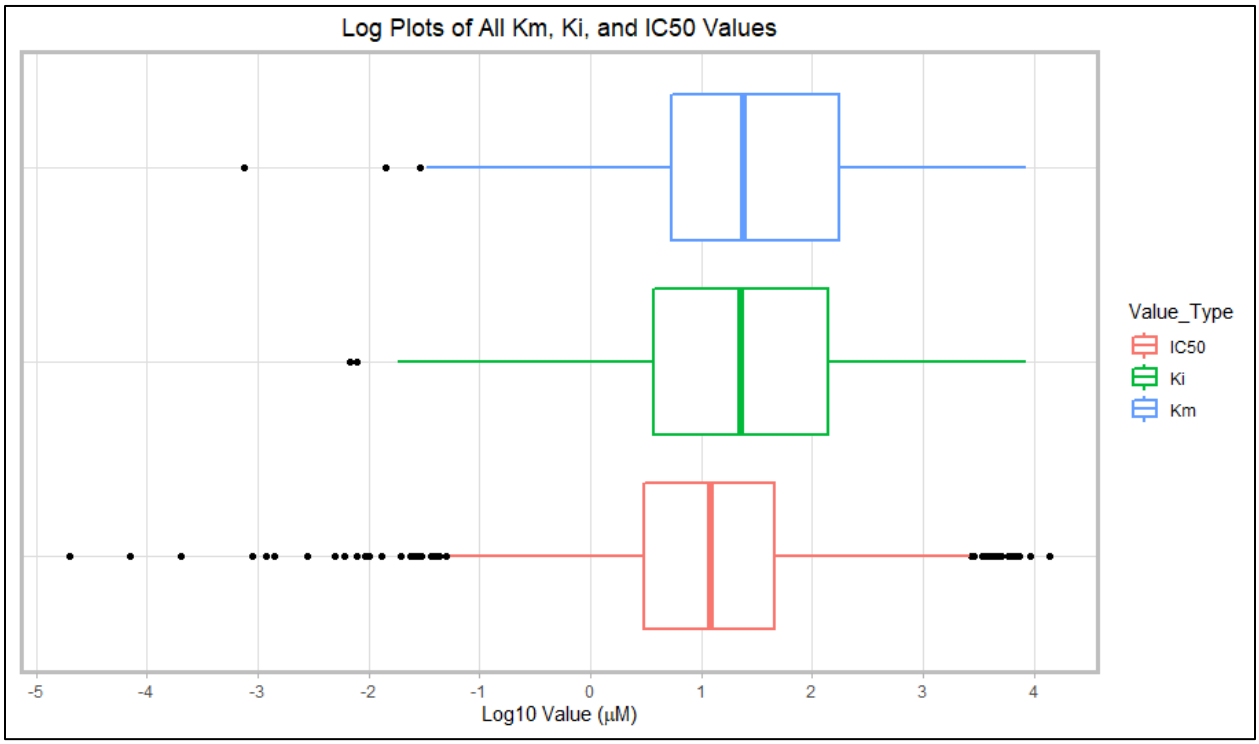


Figure 1: A boxplot of all K_m, K_i, and IC_{50} values in the database on a log base 10 scale. The boxes display the interquartile ranges (IQR) and median values. The whiskers extend from the IQR to the most extreme data point no more than 1.5 times the IQR.

3.2. Distribution of Data Among Transporters

The recent database update brings the total number of transporter proteins represented in the database from 31 up to 46. Previously, only human proteins were represented, but the database has been updated to be able to account for proteins originating in rodent and primate model organisms of which 13 are now present. Currently the non-human proteins account for just 135 interactions, or 3.4% of the database. There are 65 interactions with mouse transporters, 49 with

rat transporters and 21 with a grivet transporter. Nearly half (n = 61) of the non-human transporter interactions are currently with mouse Abcb1a and Abcb1b.

Table 1: Ranked list of the top drug transporters with >200 total interactions in UCSF-FDA TransPortal.

Transporter	Total Interactions (n)	Inhibitor Interactions (n)	Substrate Interactions (n)	Percentage of database
ABCB1	549	478	71	13.7%
SLC22A2	421	353	68	10.5%
SLC22A1	366	311	55	9.1%
SLCO1B1	293	219	74	7.3%
SLC22A8	262	221	41	6.5%
SLC22A6	235	197	38	5.9%

Table 1 shows the top six transporters represented in the database based on their total number of interactions: ABCB1, SLC22A2, SLC22A1, SLCO1B1, SLC22A8, SLC22A6. Out of the 45 transporters and 4004 interactions in our analysis, ABCB1 represents 13.7% of all data with 549 total interactions, followed by SLC22A2, containing 10.5%, and SLC22A1, containing 9.1%. Together, the top five transporters represent nearly half (47.2%) of the data (Table 1).

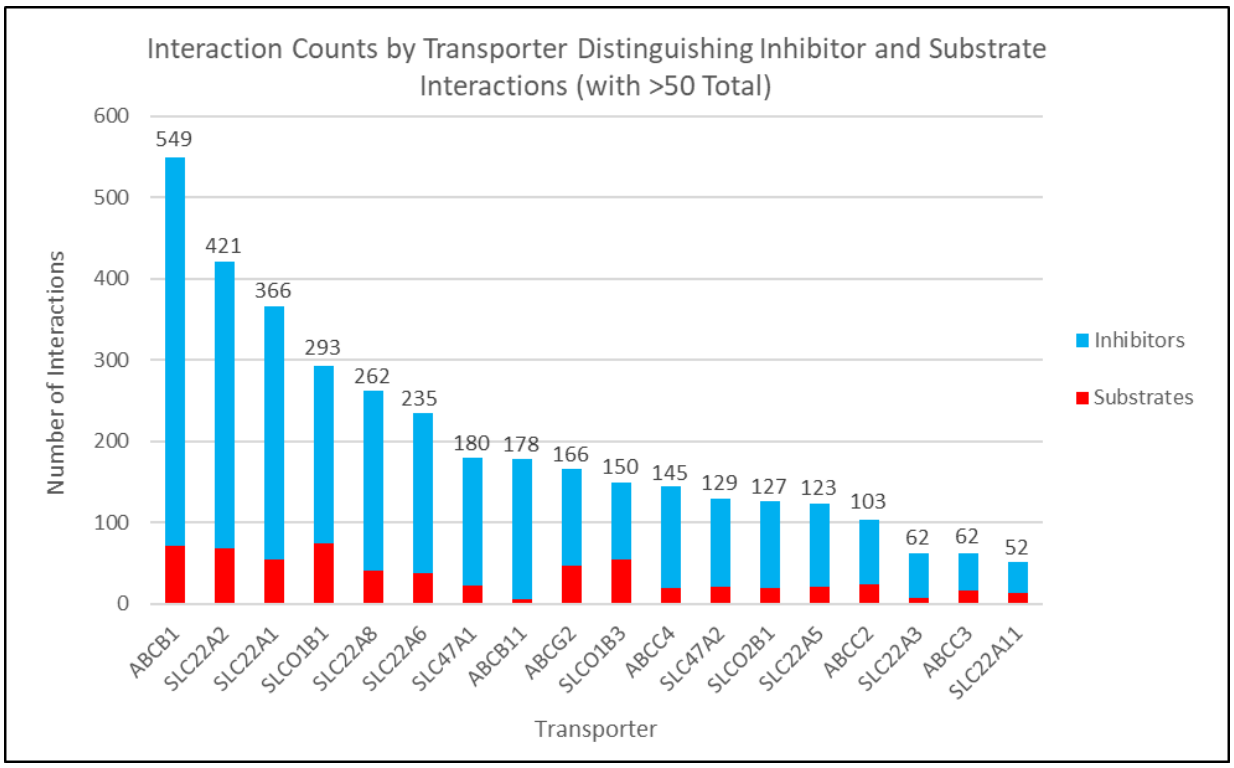


Figure 2: Ranked quantity of kinetic interactions associated with each transporter protein in the database. Stacked columns represent the sum of inhibitory and substrate interactions. Only proteins with >50 total interactions are displayed.

The top 18 transporters are all human and contain 90.0% of the data (Figure 2). In contrast, the remaining transporters (n = 27) contain a range of 1 to 47 interactions with a total of 401 interactions. Together, these interactions represent just 10.0% of the data (Table S1-3).

Additionally, out of the 45 transporters, 29% (n = 13) belong to the ABC superfamily and 71% (n = 32) to the SLC superfamily. Among the 4004 interactions, 34% (n = 1365) are associated with ABC transporters and 66% (n=2639) with SLC transporters, similar to the proportions of the proteins in each superfamily.

3.3. Chemical Interactions with Transporters

Collectively, the database contains 1101 unique chemicals and 2579 different transporter-chemical interactions. The top 20 chemicals each appear more than 19 times and represent approximately 18.0% (n = 721) of interactions (Table 2). Among these, the top five chemicals

are verapamil, appearing 86 times, cyclosporine (n = 80), quinidine (n = 53), estrone 3-sulfate (n = 41), and rifampicin (n = 40) (Table 2). In contrast, 468 of the chemicals appear just once in the database and 234 appear twice, together accounting for 63.7% of the chemicals in the database but just 17.5% of interactions.

Of the 2579 different transporter-chemical interactions, 1988 (49.7%) are completely unique (occurring only once). The top 20 transporter-chemical interactions each occur nine or more times in the database (Table 3). Of these twenty, eight include ABCB1 and represent 176 (31.4%) of ABCB1's 549 total interactions. This means that just eight out of 204 total chemicals (3.9%) interacting with ABCB1 represent 32.1% of its published interaction data. In contrast, the top transporter-chemical interaction for ABCG2 is with sulfasalazine and occurs just six times. ABCG2 appears in 166 interactions, 84% of which (n = 140) are with different chemicals.

Inhibitory chemical interactions make up 81.5% of interactions in the database (n = 3262) and substrate interactions make up the remaining 18.5% (n = 742). There are 162 different transporter-chemical interactions which have both inhibitory and substrate data associated with them. Tables 2 and 3 show that among the top 20 chemicals in the database, 14 are FDA model transporter inhibitors or substrates (Tables S1-4 and S1-5).

Table 2: The top 20 most frequently occurring chemicals in the database. FDA model inhibitors and substrates are bolded.

Chemical	Interaction Count
Verapamil	86
Cyclosporine	80
Quinidine	53
Estrone sulfate	41
Rifampicin	40
Metformin	39
Cimetidine	38
Probenecid	35
Tetraethylammonium	34
Ketoconazole	33
Ritonavir	33
Vinblastine	28
Methotrexate	28
Mk-571	24
Ranitidine	24
Valsartan	24
Quinine	22
Reserpine	20
Irinotecan	20
Imipramine	19

Table 3: The top 20 most frequently occurring transporter-chemical interactions in the database. FDA model inhibitors and substrates are bolded.

Transporter	Chemical	Interaction Count
ABCB1	**Verapamil**	43
ABCB1	**Cyclosporine**	37
ABCB1	**Vinblastine**	23
SLCO1B1	**Rifampicin**	23
ABCB1	**Quinidine**	19
ABCB1	**Ketoconazole**	16
SLC22A2	**Cimetidine**	16
SLC22A2	**Metformin**	16
ABCB1	**Reserpine**	14
SLC22A2	**Quinidine**	14
ABCB1	Nicardipine	13
SLC22A2	**Verapamil**	13
SLC22A8	**Probenecid**	13
SLCO1B1	**Cyclosporine**	13
SLC22A6	**Para-aminohippurate**	12
SLC22A6	**Probenecid**	12
ABCB1	Quinine	11
SLC22A1	**Verapamil**	10
SLC22A2	Imipramine	10
ABCB1	Paclitaxel	9

Some transporter-chemical interactions appearing many times display a wide range of experimentally derived K_i, IC_{50}, or K_m values. Two stark examples are ABCB1 with verapamil, which has IC_{50} values ranging from .2 to 446.5μM, and ABCB1 with quinidine with IC_{50} values ranging from 1 to 340μM. The most common SLC-type transporter interaction, SLCO1B1 with rifampicin, has IC_{50} values ranging from .24 to 120μM (Table 4). Among K_m values, ABCB1 tested with colchicine is a striking example: three different studies produced K_m values of 1.33, 45, and 1640μM.

Table 4: Three examples of transporter-chemical interactions in the top 20 with large ranges of reported IC_{50} values and all the cell systems and substrates from the assays producing the values in those ranges.

Transporter-Chemical Interaction	Range of IC50 values (µM)	List of Assay Cell Systems	List of Assay Substrates
ABCB1 – verapamil	.2 – 446.5	K652, MDCK, NIH-3T3, Caco-2, HeLa, B16-F10, A2780, LLC-PK1, CEM	Calcein-AM, Colchicine, Daunorubicin, Digoxin, Doxorubicin, Fexofenadine, Fluo-3, Hoechst 33342, JC-1, LDS-751, N-methyl-quinidine, Rh123, Vincristine, Ximelagatran
SLCO1B1 – rifampicin	.24 – 120	HEK293, HeLa, CHO, MDCK	8-fluorescein-camp, Atorvastatin, Bosentan, Bromsulphthalein, Cerivastatin, CGAMF, Estradiol-17-beta-glucuronide, Estrone sulfate, Fluorescein-methotrexate, Lithocholyl-lysine
SLC22A2 – quinidine	6.4 – 92.6	CHO, Oocytes, HEK293	ASP, Metformin, MPP, n-methylpyridinium, NDB-MTMA, Tetraethylammonium, YM155

3.4. Assay Environment Types

Given the wide range of kinetic parameters among some transporter-chemical interactions, we analyzed the cellular and membrane environments used for assays that produce quantitative kinetic data. These assay environments are broadly described as the "Cell System" in the database. Counts for cell types of the most common assay membrane environments are displayed in Table 5. Based on total counts, HEK293 cells are the most common cell type listed in the database, appearing in 1530 interaction assays. This is over three times more than the next most common cell system, CHO (Chinese Hamster Ovary) cells (n = 451). Overall, 18 different assay systems appear in 10 or more interaction assays and cover 3835 interactions, or 95.8% of the database. When dividing the data by transporter superfamily it becomes clear that there is a stark division between membrane environments used for studying ABC and SLC transporters (Figure 3). For ABC transporters, Sf9, Caco-2 (human colorectal adenocarcinoma), LLC-PK1 (Lewis-lung cancer porcine kidney), and NIH-3T3 (embryonic mouse fibroblast) cells are the dominant assay systems. For SLC transporters, HEK293, CHO, S2 (*Drosophila melanogaster* embryo) cells, and oocytes

from *Xenopus laevis* are the most prevalent. In fact, each of the above cell systems have a divide in utilization between the superfamilies of greater than 90 to 10 percent except Caco-2 cells, which have an 85 to 15 percent split (ABC to SLC superfamily). Only the MDCK (Madine-Darbey canine kidney) and HeLa (Henrietta Lacks) cell lineages are used more evenly while appearing greater than 100 times in the database.

Table 5: Cell environment types appearing >100 times the USCF-FDA TransPortal database. Human-derived cell types are bolded. See Table S1-6 for counts and descriptions of all cell types in the database.

Cell/Membrane Environments	**Total Count**	**ABC Transporters**	**SLC Transporters**
HEK293	1530	101	1429
CHO	451	0	451
Sf9	433	433	0
MDCK	302	114	188
S2	220	0	220
Caco-2	153	129	24
Oocytes	150	2	148
LLC-PK1	145	143	2
NIH/3T3	132	132	0
HeLa	128	37	91
All others	360	274	86
Total	4004	1365	2639

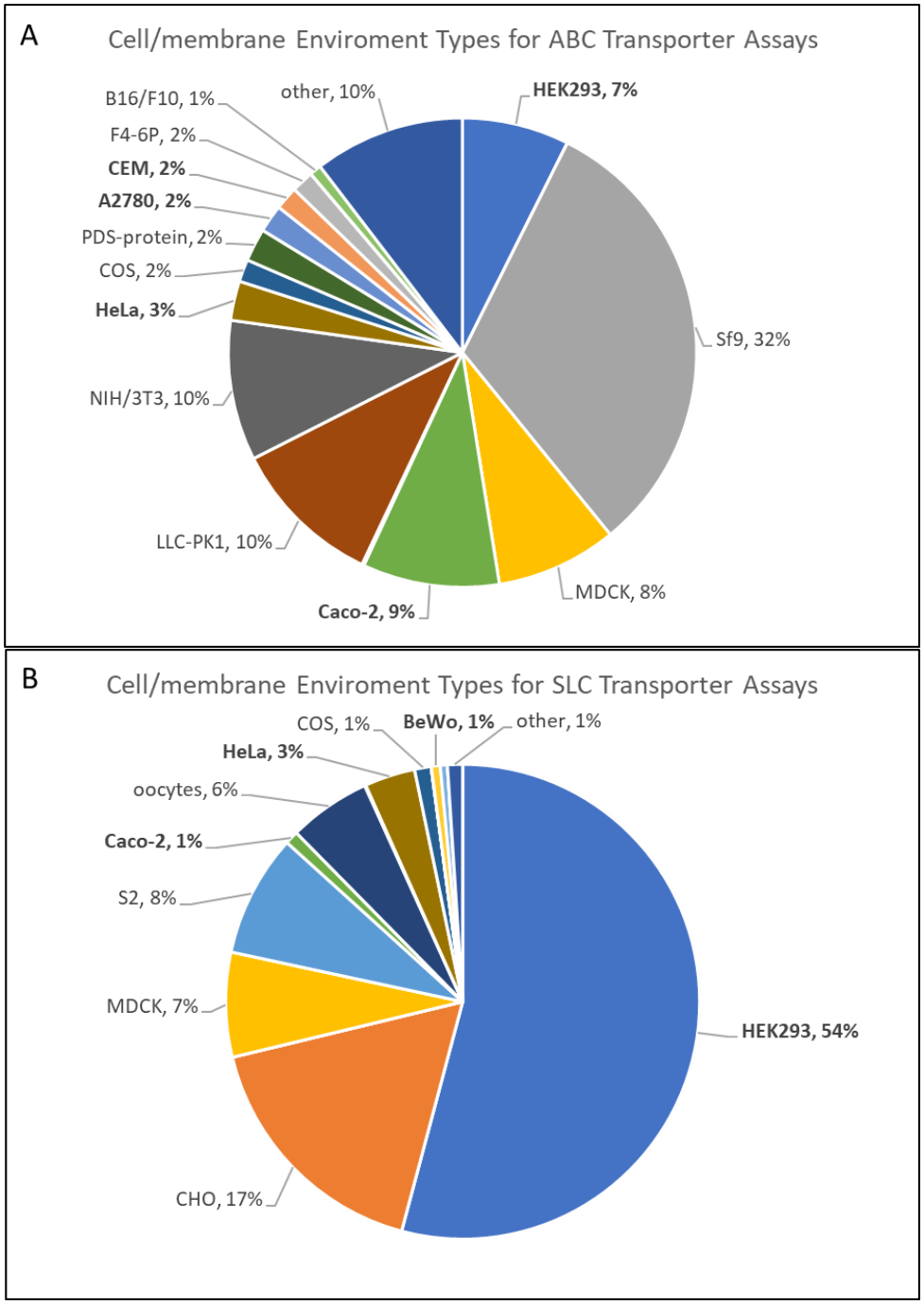


Figure 3: A pie chart of cell/membrane environments used in (A) ABC transporter assays and (B) SLC transporter assays. Human-derived cell types are bolded. Cell/membrane types constituting < 1% are not shown.

It is notable that although most research on transporter proteins is conducted with human health outcomes in mind, most transporter assays appear to be conducted in non-human cell lines. Among the cell/membrane environments appearing 100 or more times in the database (constituting the top ten, which cover 91% of interactions) only HEK293, Caco-2, and HeLa cells are human

cells. The remainder are from Chinese hamster (CHO), dog (MDCK), mouse (NIH/3T3), pig (LLC-PK1), frog (oocytes) and two different insect species (Sf9 and S2).

3.5. Distribution of Data in TICBase

Prior to the current update, no drug transporter interactions with environmental chemicals were reported. The update includes 120 inhibitory interactions of environmental chemicals (TICs) with transporter proteins and 13 substrate interactions, together covering 3.3% of the total database. Among the total, 133 interactions, 70 are with ABC and 63 are with SLC transporters, 94 of which are human, 30 from mouse, and 9 from rat.

The most potent inhibitor found among the environmental chemicals is the ionic liquid 1-hexylpyridinium-chloride (hepy-cl) acting on SLC22A2 with an IC_{50} of .35µM (Table 6). Ionic liquids are an emerging class of industrial chemicals which have been investigated for their potential to contaminate water sources (Cheng et al. 2011). The next most potent inhibitor is the antiseptic chlorhexidine acting on both SLC22A2 and SLC22A3with an IC_{50} of .4µM. The highest IC_{50} values by comparison are from three metabolites of the agricultural fungicide propiconazole: gamma, alpha, and beta-hydroxy-propiconazole, at 350.8, 366.4, and 456.3µM respectively, acting on ABCB1. The K_m values of the environmental chemical substrate interactions ranged from 10µM for metolachlor transported by ABCB1 to 212µM for paraquat transported by SLC47A1.

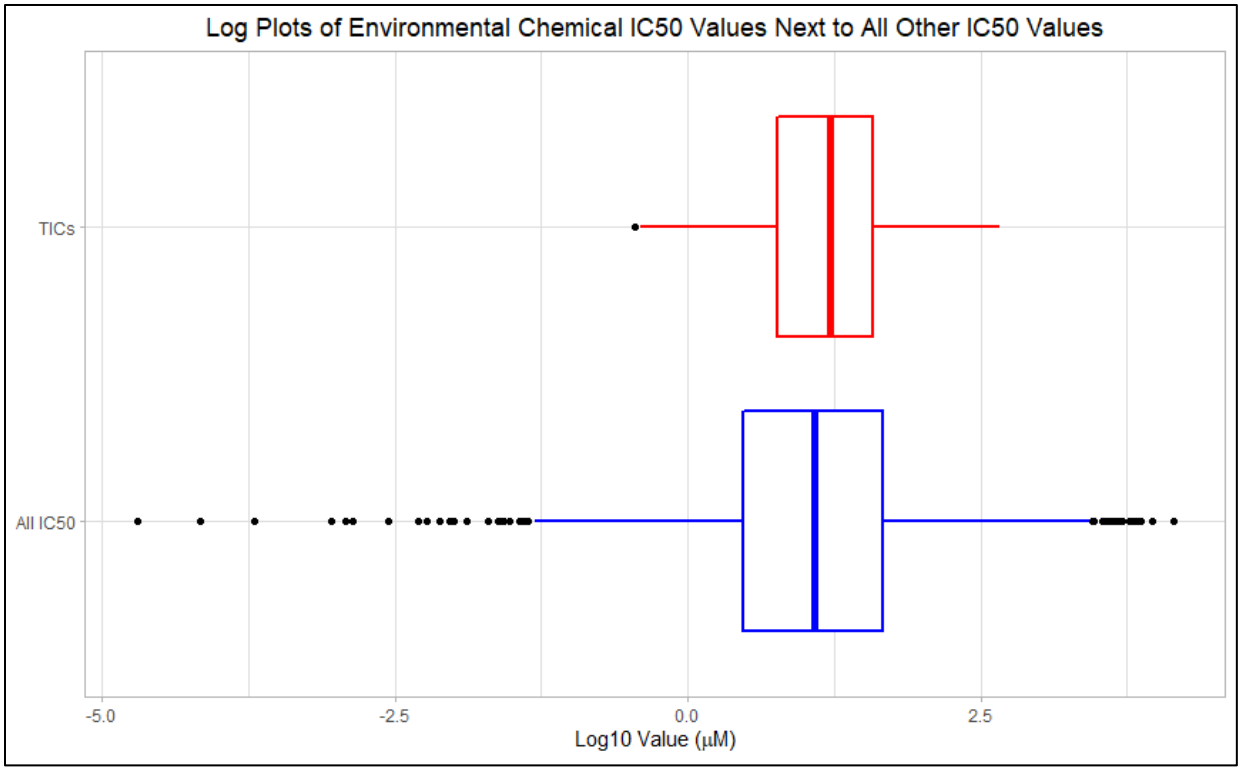


Figure 4: A boxplot displaying the range of IC_{50} values of the environmental chemicals next to a boxplot of all IC_{50} values. The boxes display the interquartile ranges (IQR) and median values. The whiskers extend from the IQR to the most extreme data point no more than 1.5 times the IQR.

Although the most potent of the environmental inhibitors come from a variety of chemical classifications, the antiseptic chlorhexidine stands out, appearing five times among the 21 most potent inhibitory interactions. Among the top twenty most potent inhibitors are three organophosphate, three organochlorine, one pyrethroid, and one triazole pesticide, five ionic liquids, and the flame retardant tetrabromobisphenol A twice. The *least* potent in the top twenty has an IC_{50} value of just 3.6µM (Table 6).

Table 6: The twenty most potent inhibitors among environmental chemicals interacting with transporters.

Environmental Chemical	Transporter	IC_{50} (μM)	Classification	Assay System
1-hexylpyridinium-chloride	SLC22A2	0.35	Ionic liquid	CHO cells
Chlorhexidine	SLC22A2	0.4	Antiseptic	HEK293 cells
Chlorhexidine	SLC22A3	0.4	Antiseptic	HEK293 cells
N-butyl-N-methylpyrrolidinium-chloride	SLC22A2	0.48	Ionic liquid	CHO cells
Tetrabromobisphenol A	SLC22A8	0.5	Flame retardant	HEK293 cells
Chlorhexidine	SLC47A2	0.5	Antiseptic	HEK293 cells
Tetrabromobisphenol A	SLCO1B1	0.6	Flame retardant	CHO cells
Chlorhexidine	SLC47A1	0.7	Antiseptic	HEK293 cells
1-methyl-3-butylimidazolium-chloride	SLC22A2	1.5	Ionic liquid	CHO cells
Pentachlorophenol	ABCB1	1.6	Organochlorine	MDCK-II cells
N-butylpyridinium-chloride	SLC22A6	1.6	Ionic liquid	CHO cells
N-butylpyridinium-chloride	ABCB1	2.29	Ionic liquid	CHO cells
Allethrin	SLC22A1	2.6	Pyrethroid	HEK293 cells
Endosulfan	ABCB1	2.8	Organochlorine	NIH/3T3 cells
Fenamiphos	SLC22A2	2.8	Organophosphate	HEK293 cells
Phosalone	ABCB1	3	Organophosphate	NIH/3T3 cells
P,P'-DDD	ABCG2	3	Organochlorine	Sf9 membranes
Mirex	mouse_Abcb1a	3	Organochlorine	Detergent-solubilized protein
Propiconazole	ABCB1	3.6	Triazole	NIH/3T3 cells
Phosmet	ABCB1	3.6	Organophosphate	HEK293 cells

Out of all 120 of the TIC interactions over half (n = 68) are with pesticides. The only disinfectant other than chlorhexidine is pentachlorophenol, which is also among the twenty most potent TICs. Chemicals used in flame retardant applications, consisting of 4 PDBEs, 6 PFAS compounds and 8 interactions with tetrabromobisphenol A (TBBPA), make up 18 more. The remaining TIC interactions include 15 with PCBs, 7 with ionic liquid solvents, and 6 with other bisphenol compounds (bisphenol A, S, and F) used as plastic monomers. In contrast, just six different chemicals are involved in the environmental chemical substrate interactions:

perfluorooctanoic acid, paraquat, cyperquat, and the three chloroacetanilide compounds metolachlor, acetochlor and alachlor (Figure5).

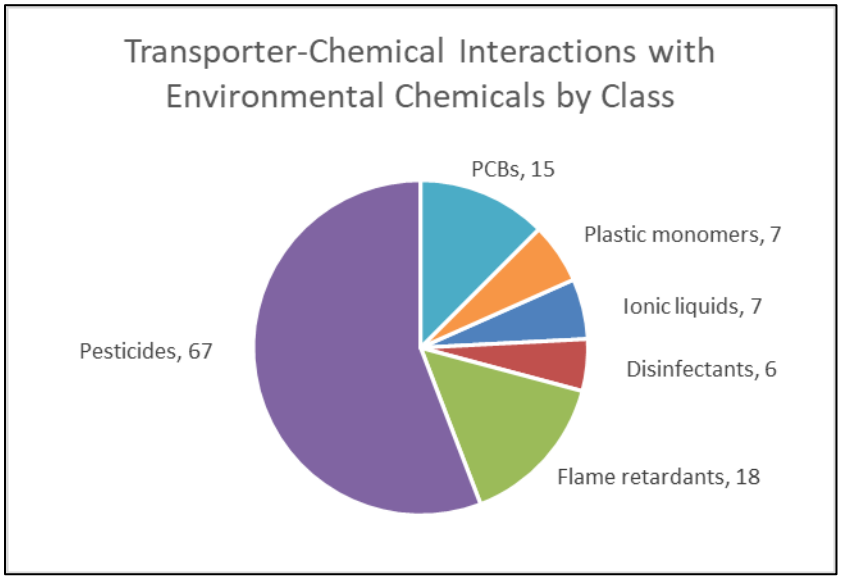


Figure 5: A pie chart of environmental chemicals in TICBase grouped by chemical class (includes both inhibitory and substrate interactions).

4. Discussion

The updated UCSF-FDA TransPortal + UCSD/UCD-NIEHS TICBase represents a substantial step forward in cataloging a wide array of transporter-chemical interactions to fulfill its original and continuing purpose, to be a useful source of information on transporters relevant to the drug discovery process. The addition of interactions that include transporters from non-human mammalian model organisms and environmental chemicals represents the expanded scope and more careful considerations now being given to transporter-chemical interactions in the healthcare field. (Fardel et al. 2012; Nicklisch and Hamdoun 2020) Data on non-human transporters from mammalian model species is an important reference for researchers because they are commonly used in research and drug development to discern the pharmacokinetics of and potential adverse responses to new compounds. It has been shown that drug transporter orthologs

have altered interaction patterns with environmental chemicals (Nicklisch et al. 2021) and this should be taken into account whenever data is extrapolated to the human clinical setting.

Environmental chemicals of anthropogenic origin can impact human health through incidental and occupational exposure. Most research on transporter-chemical interactions continues to be focused on pharmaceuticals, and to a lesser extent on endogenous substances, herbs (Li et al. 2017), and food products (Sjöstedt et al. 2017; Koziolek et al. 2019; Tikkanen et al. 2020) most likely because exposure to such chemicals is either intentional or anticipated. Human exposure to environmental chemicals is well-documented and actively monitored in some states (CDPH, OEHHA, DTSC 2022 May 10) and should also be anticipated by researchers and health professionals. More research is needed to determine whether levels of environmental chemicals currently present are cause for concern in DDIs and other effects related to transporter impairment. The recent update shows, however, that the inhibitory values of TICs fall within the ranges of several FDA model transporter inhibitors (Table S1-4), a clear indication for significant transporter-mediated effects.

Examination of the potential adverse health effects due to transporter-environmental chemical interactions is especially warranted for potent inhibitors of transporters still in use. Both mirex and pentachlorophenol were banned under the Stockholm Convention but pentachlorophenol still sees limited use in the United States as a wood preservative (US EPA 2016 Jan 3; UNEP 2019). TBBPA is widely used as flame retardant in circuit board polymers, papers and plastics. Although it is covalently bound to the base polymer in its applications, which limits release into the environment, it has been found in indoor dust and human milk samples. (Jones et al. 2014; Kodavanti and Loganathan 2014) One of the most potent TICs in the database, chlorhexidine, is widely used as a disinfectant for topical uses, surgical tools, and as a dental mouth

wash among other uses. Chlorhexidine is very poorly absorbed by the GI tract, but its direct use in dental and medical applications and potent inhibition of at least five well-known transporters is noteworthy (3M Company 2013; Karpiński and Szkaradkiewicz 2015).

5. Concluding Remarks

While the database is by no means a complete compilation of all literature in the field, the distribution of data among the transporter proteins clearly shows that previous research has focused heavily on a select few human drug transporters of interest with comparatively little data available on the interactions of others, including orthologs in model mammalian species. As with individual proteins, we see that a small number of articles also contribute a disproportionate amount of available data on transporter-chemical interactions. In some cases, the data on a particular transporter is the result of a very small number of independent investigators which makes it difficult to assess the effects of different cell systems and assay substrates on quantitative kinetic parameters. “House effects” could influence results produced under separate experimental setups, however not enough data is present to thoroughly evaluate variations caused by using different assays, substrates, or other factors. Independent replication of experiments under standardized assay conditions is essential to verifying the broad applicability of findings from transporter studies, as in other fields.

Although the quantitative kinetic parameters span between six and nine orders of magnitude (Figure 1), the largest values are so high that the concentrations they represent are not at physiologically relevant levels. This raises questions about the concentration at which it is appropriate to cease testing for weak inhibitor or substrate activity and report the interaction as “weak”, instead of having a significant effect on activity. Given high enough concentrations, an interaction could potentially be observed in many assays where researchers have kept chemical

concentration ranges limited for pragmatic reasons. Also, many transporter assays are intended for determining inhibition interactions specifically (Nicklisch and Hamdoun 2020), with substrate interactions being given comparatively less attention, hence the large difference in inhibitory and substrate interactions seen in TransPortal. Transporter-chemical interactions that were found to have both inhibitory and substrate interaction values can reasonably be predicted to be competitive inhibition interactions, even though competitive binding may not have been directly examined in studies.

We also found there are some surprisingly large ranges of kinetic parameter values within individual transporter-chemical pairings. A factor of 2232.5 separates the lowest and highest IC_{50} value for ABCB1 with verapamil and a factor of 1233 separates the lowest and highest K_m values for ABCB1 with colchicine. The wide ranges seen in these examples may be explained by the use of a variety of different reporting substrates and *in vitro* assay environments in transporter assays (Table 4). The wide array of cell systems in the database is particularly noteworthy since it has been shown that differing membrane environments can drastically alter transporter activity (Shukla et al. 2017). It would be prudent to further examine cell line effects on transporter activity, especially given the different species of origin that the most common cell systems used to study transporters have been taken from.

In summary, the results of our database analysis reveal large variations in kinetic parameters of individual transporter-chemical interactions across test systems and research labs. Standardization of these kinetic assays with a range of model compounds that could capture the multiple known modes of interactions, at environmentally relevant test concentrations, and within reproducible assay environments are needed to clearly identify and characterize environmental chemical and drug interactions with polymorphic and orthologous drug transporter variants. Given

the large set of legacy and emerging environmental chemicals to test, those assays should be optimized for high throughput screening approaches and augmented with in silico approaches that can use machine learning and network analysis tools to better predict and mitigate adverse interactions and toxic bioaccumulation of these compounds. While this database only considers drug transporter interactions with drugs and environmental chemicals, a larger, unified database with kinetic data on DECI and DDIs with other xenobiotic sensors (e.g., aryl hydrocarbon receptor) and processors, including metabolizing CYP (cytochrome P450) enzymes, could further support and aide in the development of more personalized drug therapy in humans and “greener" environmental chemicals that are eliminated more efficiently from humans and non-target organisms.

Chapter 2: Introducing UC Transportal – A Durable Data Resource to Reframe the Field of Transporters

1. Introduction

1.1. Background

The ubiquity of active transporter proteins in nature, in all domains of life from bacteria to protists, vascular plants to humans, means there is an incredible quantity of transporter functions to be characterized and transporter-chemical interactions to be examined. Even within species, active transporters act on a diverse array of substances from endogenous compounds such as inorganic ions (Na^+, K^+, Fe^{2+}, PO_4^{3-} etc.) (Zhang et al. 2019), lipid components (Wang and Westerterp 2020), amino acids (Bröer and Fairweather 2018), and glucose (Thorens and Mueckler 2010), to xenobiotics including both natural toxins and (Fischer et al. 2005; The International Transporter Consortium et al. 2010; Nicklisch and Hamdoun 2020). Due to their nature as transmembrane proteins they are considered a first point of contact with the external environment and interact with anything the cell may encounter. MDR (multi-drug resistance)-type transporters in particular serve as a first-line defense mechanism for keeping toxic or unknown substances out of the cell. Hence, the range of transporter-chemical interactions we should expect to occur is not limited to the compounds of a normal and healthy biological environment but is subject to all external exposures, the "exposome" of the organism of the transporter.

While many transporters show high specificity in recognizing and transporting endogenous substrate, others can recognize a variety of chemical structures, including xenobiotics, and transport them across the cell membrane. Two transporter superfamilies in particular, the ABC (ATPase-binding cassette) and SLC (solute carrier) superfamilies, contain many proteins known to act on wide variety of pharmaceuticals which are known as "drug transporters" (Giacomini and

Huang 2013; Schlessinger et al. 2013; Nigam 2015; Nigam 2018). The effects of drug transporter interactions on overall drug disposition and bioavailability have been well validated in pharmaceutical research. However, transporter interactions with anthropogenic environmental chemicals have been understudied and only recently gained interest in pharmacological and toxicological sciences (Smital et al. 2004; Epel et al. 2008; Luckenbach and Epel 2008; Nicklisch and Hamdoun 2020). These MDR-active transporter superfamilies are also highly conserved and appear across all domains of life, displaying a tremendous array of functionalities and ubiquity on par with or exceeding most other superfamilies of transporter protein.

Research on drug transporters and other ABC and SLC transporters has been concentrated primarily on human and model mammalian organisms. Until recently, the UCSF-FDA TransPortal (http://transportaldev.docking.org), the central repository for transporter-chemical interaction data for ABC and SLC transporters, focused solely human proteins and pharmaceuticals. However, these superfamilies are studied in other fields of research as well, as the biological function they perform is essential in all organisms. In plant biology, ABC transporters are studied for their roles in hormone transport, metal sequestration, nutrient storage, and transporting endogenous structural compounds and secondary metabolites such as monolignols, anthocyanidins, flavonoids, and aromatics (Lefèvre et al. 2015; Hwang et al. 2016). ABC and SLC transporters are of interest concerning the xenobiotic defense mechanisms of aquatic organisms, which are regularly exposed to environmental chemicals in contaminated waters and runoff and consumed by humans (Smital et al. 2004; Cunha et al. 2017; Nicklisch et al. 2017a; Nicklisch et al. 2021; Romersi and Nicklisch 2022). Similar to the term MDR, MXR (multi-xenobiotic resistance) mechanisms involving these transporters have also been described in insects and are a major concern for the agricultural and

public health fields as resistance to widespread pesticide use becomes more common (Seong et al. 2016; Mastrantonio et al. 2019; Yang et al. 2019).

A holistic observation of transporter research from across fields yields the conclusion that, as a field in and of itself, active transporter research is fragmented and divided into a series of specialty subtopics under the category of other research subjects. These include drug pharmacokinetics, veterinary medicine (Halwachs et al. 2014), food and water safety (Sjöstedt et al. 2017), the biochemistry of plant hormones and secondary metabolites (Hwang et al. 2016; Behrens et al. 2019), and aquatic toxicology. Current transporter databases either emphasize genomic data and transporter evolution and phylogeny (such as the Transporter Classification Database and TransporterDB 2.0) (Elbourne et al. 2017; Saier et al. 2021), or focus on a particular application of transporter research with a narrow subset of proteins (such as UCSF-FDA TransPortal) (Morrissey et al. 2012). Currently, there are no publicly available data resources to bring together transporter interaction data from across applications and biological fields of study.

For this reason, we introduce UC Transportal, a comprehensive and expansive data repository and bioinformatics resource, on the functional biology of all ABC and SLC transporter proteins. UC Transportal is intended to unify research on ABC and SLC transporters from across disparate fields of study in a manner that supports collaboration and advances knowledge on pharmacological and ecotoxicological phenomena these transporters are involved in. Researchers and industry professionals have an enormous amount of insight to gain and methodologies to share from cross-field collaborations in transporter research. The fact that these superfamilies are ubiquitous in nature, perform a fundamental biological function, are highly conserved, and can be examined using similar analysis methods in the laboratory make it possible to reframe transporter research as a specialized field in its own right (Xiong et al. 2015). UC Transportal collects and

categorizes data directly from *in vitro* and *in vivo* research studies for a complete picture of ABC and SLC transporter interactions with all chemicals.

1.2. Design Rationale

UC Transportal seeks to be a modern, durable, and comprehensive data resource in support of the broad community of transporter researchers that will organically grow and expand well into the future. This required creating a website using up-to-date web management software with aesthetic and functional design qualities which make UC Transportal attractive, easy to navigate, and simple for users to search. The most important quality for the success of a scientific database is that its targeted userbase finds it useful enough to refer to repeatedly over time. To achieve that goal, we designed UC Transportal to be able to possess and maintain the following characteristics: (1) a sufficient numeric quantity of information to provide users with successful search results and complete information, (2) a wide scope of qualitative data types to cover the breadth of different interests and needs held by a diverse userbase, and to give greater context to interaction data, (3) ease of use, so that information can be conveniently accessed by new and experienced users alike, and (4) up-to-date and reliable, to develop a reputation as an active and trustworthy data resource. To accommodate these characteristics, UC Transportal is envisioned to become a hub for multiple related databases on different kinds of transporter interaction data. Another major goal which will contribute to the longevity and utility of UC Transportal is to facilitate collaboration by inviting researchers of ABC and SLC transporters from a diverse set of backgrounds to contribute directly to the databases. Initiating such collaborative involvement will engage the UC Transportal userbase and promote transporter research as a self-sufficient field of study.

The flagship component of UC Transportal is the Kinetic Transportal database, which describes direct transporter-chemical interactions and their kinetic parameters. The design

rationale for Kinetic Transportal will eventually serve as a template for future databases as UC Transportal expands. To achieve clear communication and conform to the overarching design rationale for UC Transportal, the following characteristics were pursued when creating Kinetic Transportal. (1) *Expandability:* we designed the database to hold an unlimited amount of data without sacrificing aesthetics or user-friendliness when searching the data. (2) *User-friendliness:* the database must be intuitive to search, and the data must be easy to read with a clean interface. (3) *Accuracy and Precision:* the standards set for naming and displaying values are intended to be as clear and unambiguous as possible with a rigorous curation process to ensure fidelity to the source literature. (4) *Comprehensiveness:* because assay results can be affected by numerous factors, and users may search the database for different purposes, we included as much data as necessary to give a complete description of each transporter-chemical interaction. (5) *Compatible with in silico analysis:* the database was formatted to be ready for computer-driven analyses to derive high-level insights from the data in the future.

Several metadata tables in UC Transportal standardize the nomenclature of the database and contribute to the achievement of these attributes. While the Kinetic Transportal database derives all its data directly from primary research articles, the metadata refers to reputable external web resources to confirm the identity and proper nomenclature of data values throughout UC Transportal. The metadata tables are intended to apply to all individual databases and analytics tools which may be added and will act as an additional connection between distinct these resources aside from simple association through UC Transportal as a hub.

The creation of UC Transportal serves as a solid foundation for establishing a new and lasting data resource which we hope will prove uniquely valuable for unifying transporter research from across disciplines into an integrated hub for ABC and SLC transporter protein interactions.

The project is expected evolve over time as more data becomes available and is curated into Kinetic Transportal and future components of the system. Hereafter, we give a detailed description of the UC Transportal website and its features, the Kinetic Transportal database and all administrative processes associated with it, instructions for their use, explanations for the structure and meaning behind each component, the current state of the data, and an outlook for the future of UC Transportal.

2. Description of the UC Transportal Website

2.1. Technical Website Information

The UC Transportal is currently in an off-line development site (https://sandbox-3.metro.ucdavis.edu/) and will be hosted at uctransportal.ucdavis.edu on servers maintained by the Metro IT group at UC Davis. UC Transportal uses the WordPress (https://wordpress.org) as content management system software and several plug-ins for functional purposes: Max Mega Menu (https://www.megamenu.com) for menu modifications, Post SMTP (https://postmansmtp.com) for automated email management, Ultimate Member (https://ultimatemember.com) to manage user registrations, Wordfence Security (https://www.wordfence.com) for site protection, wpDataTables (https://wpdatatables.com) for data table functionality, and WPForms Lite (https://wordpress.org/plugins/wpforms-lite/) for the site contact form. The WordPress theme is a modification of Iceberg by Nord Studio (https://themesinfo.com/iceberg-wordpress-blog-template-hk2v/mbennardo.com-vzzyy). All backend database management is done with MySQL (https://www.mysql.com) software. UC Transportal has three levels of access and usership: users, registered members, and administrators.

2.2. Website Description and Functionality

The landing/home page (Last accessed: August 11th, 2022) welcomes users with a brief description of the website's purpose and content next to an animation of a transporter interacting with a chemical at its docking site. A counter at the bottom of the page displays the numbers of proteins, chemicals and organisms represented in the databases. The entire UC Transportal site can be navigated using the universal side menu on the left-hand side of the screen. The UC Transportal logo links back to the landing/home page while all other links lead to other pages (Figure 2-1).

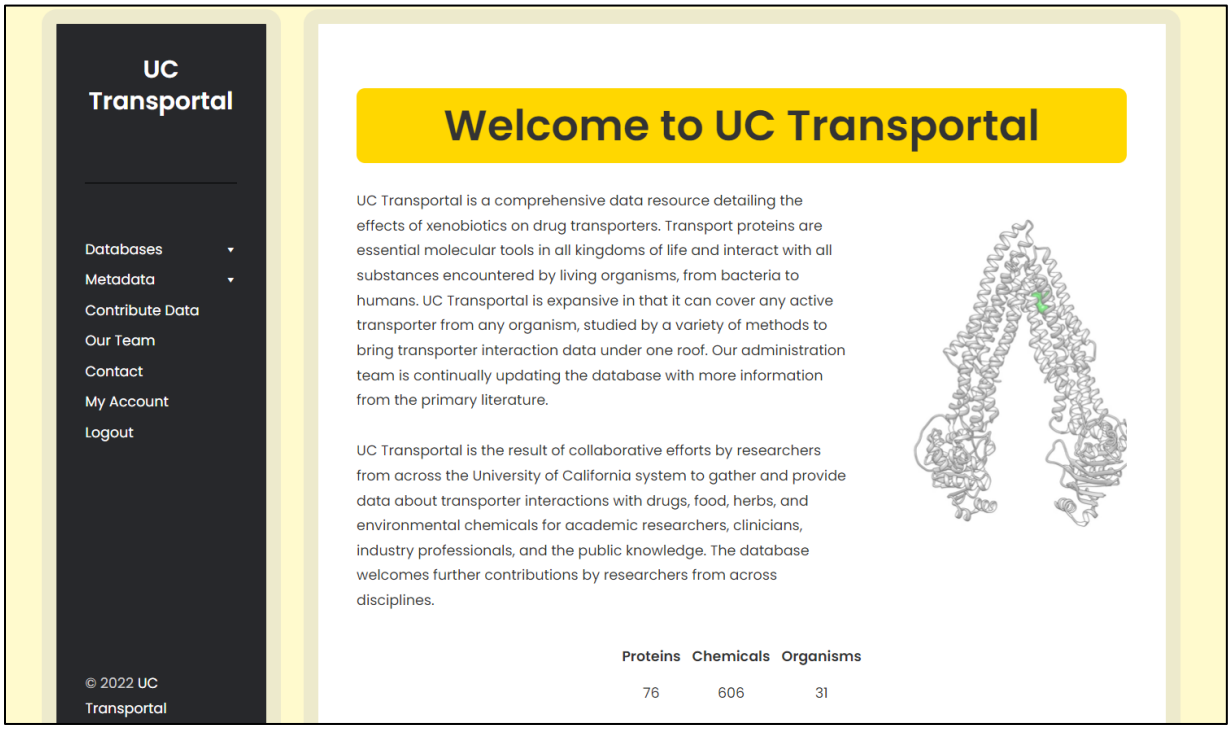


Figure 2-1: A screenshot of the landing page for the UC Transportal website for a registered member who is logged in.

Unregistered users have access to the entire site except for the *Contribute Data* pages and *My Accoun*t page. When a user is not logged in, the *My Account* link does not appear in the side menu and the *Contribute Data* link redirects to the user to the *Login* page (Figure 2-2). Users can login or register from the *Login* page. When registering, prospective members will be asked to enter a username, their first and last name, email address, institution of affiliation, and a brief

description of their research. All registrants will be briefly vetted by the UC Transportal administration team (Currently: Matthew Michel and Sascha Nicklisch) before being approved.

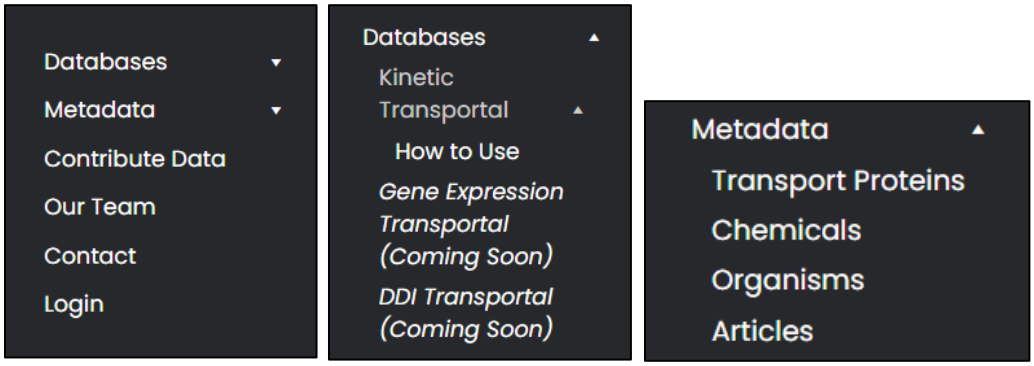


Figure 2-2: From left to right: the side menu when a user is not logged in, options under the Databases drop-down list, and options under the Metadata drop-down list.

To access the *Kinetic Transportal* database, navigate to its front page using the side menu. To search the database, enter your search parameters into the input fields for transporter protein, chemical, organism, or any combination thereof, or select your search parameters from the dropdown menus (Figure 2-3, 2-4). The search will return all entries containing all parameters entered in their respective fields. Entering only a single parameter may return an excessive number of results because all interactions associated with it will be returned. If a search term is not present, users can try searching the appropriate metadata table in case it is listed as an alternative name. To include MXR/MDR-type interactions (explained below in section 3.4.1) in your search, users can click on the checkbox at the bottom of the search form. The *How to Use* page gives a brief set of instructions about how to use the database, a description of each data field in *Kinetic Transportal*, and a description of the technical conventions used for database nomenclature.

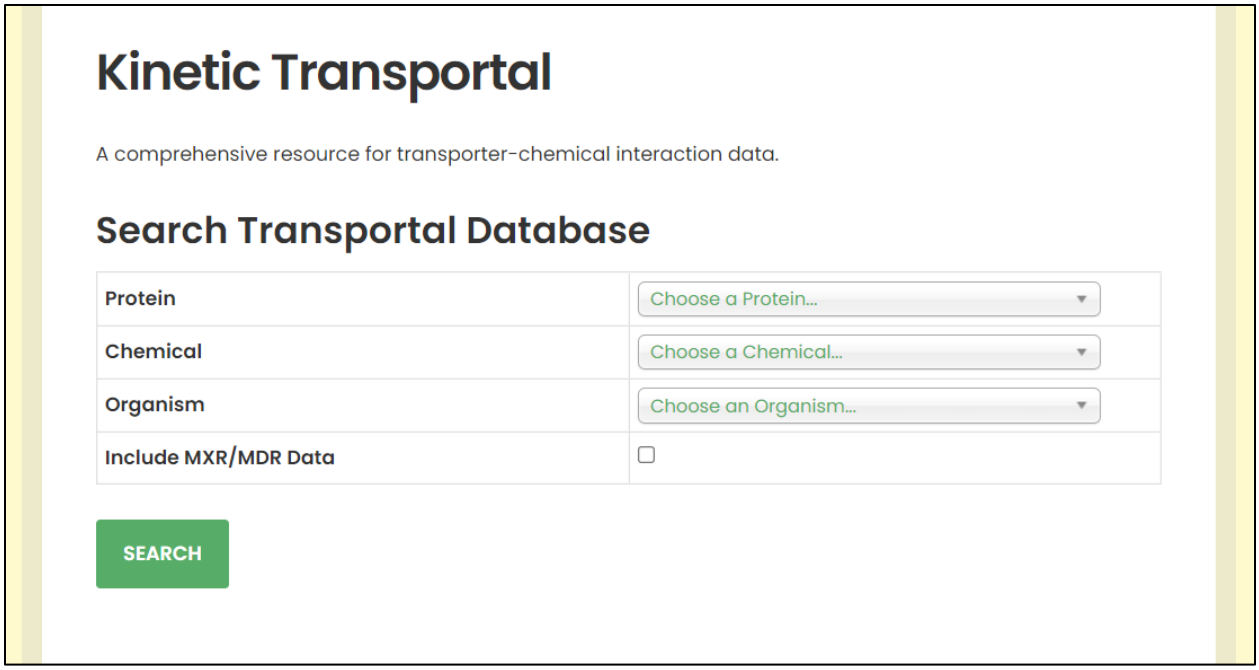


Figure 2-3: The front page of the Kinetic Transportal database and its search form.

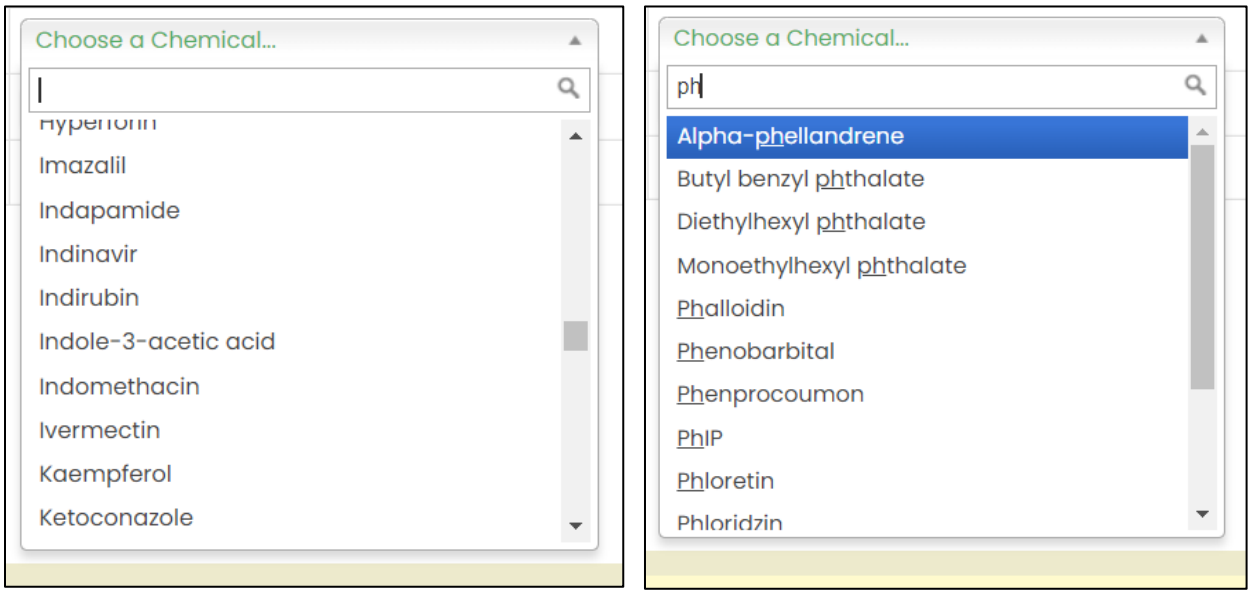
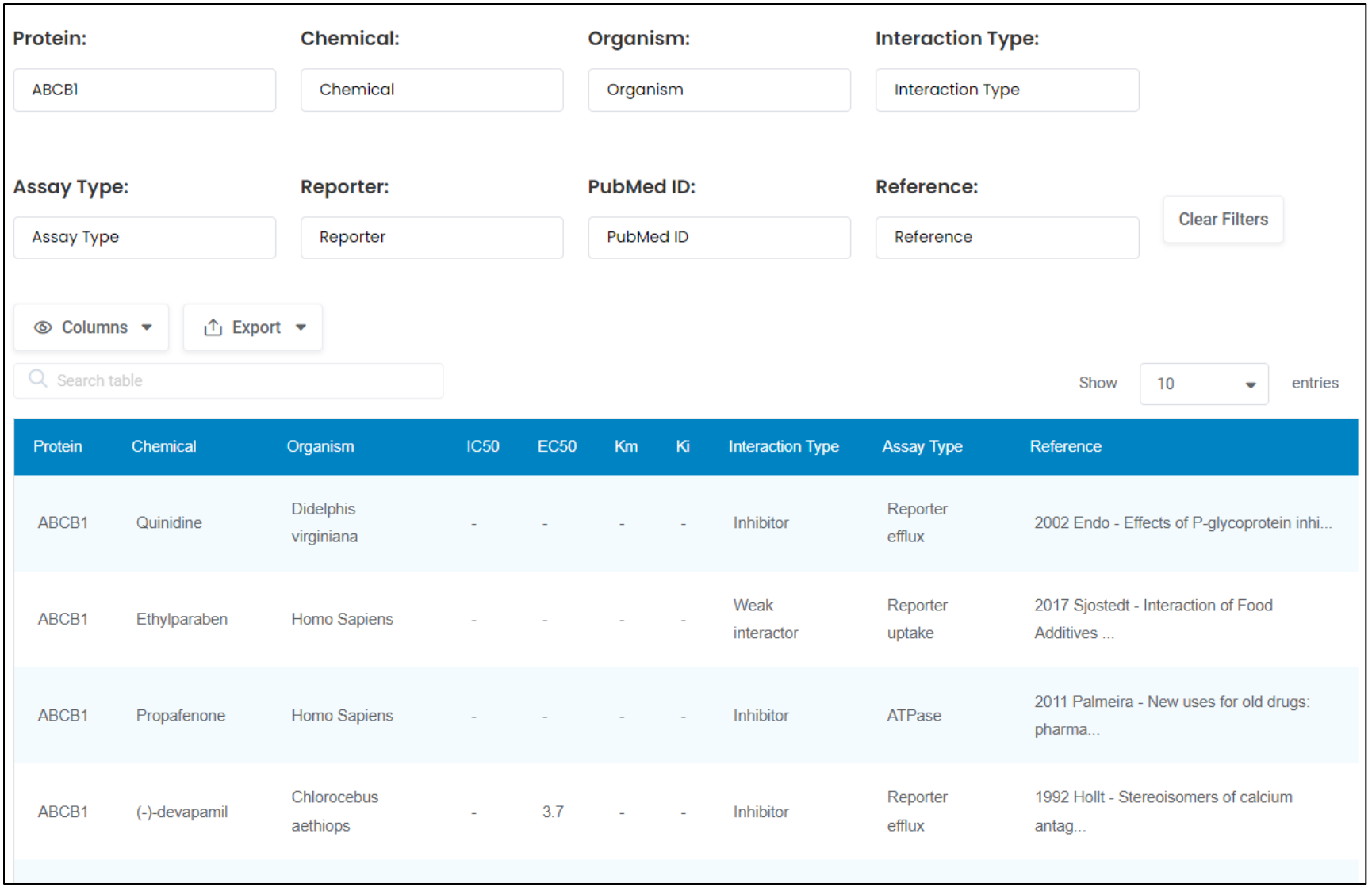


Figure 2-4: Drop-down selection lists for the Chemical parameter in the Kinetic Transportal search form. All entries are listed in alphabetical order and can be filtered by entering the beginning of your search terms (right).

The search results page displays the results in a data table that can be filtered, sorted, and exported if desired (Figure 2-5). To filter search results, enter terms into the filters at the top of the page. To sort the table based on a particular field, click on the blue field header. The export button allows users to export the data displayed under the current settings as a .csv file, Excel file (.xlsx),

or copy it to the clipboard. The default display does not include all fields in Kinetic Transportal; column displays can be toggled on or off using the drop-down menu under the Columns button on the left side of the screen (Figure 2-6). There is additionally a universal search bar atop the table which will filter the entire table for any term entered. Options to adjust the results shown per table page are located on the right side of the screen, and the table page can be selected from the bottom side of the data table.



Protein	Chemical	Organism	IC50	EC50	Km	Ki	Interaction Type	Assay Type	Reference
ABCB1	Quinidine	Didelphis virginiana	-	-	-	-	Inhibitor	Reporter efflux	2002 Endo - Effects of P-glycoprotein inhi...
ABCB1	Ethylparaben	Homo Sapiens	-	-	-	-	Weak interactor	Reporter uptake	2017 Sjostedt - Interaction of Food Additives ...
ABCB1	Propafenone	Homo Sapiens	-	-	-	-	Inhibitor	ATPase	2011 Palmeira - New uses for old drugs: pharma...
ABCB1	(-)-devapamil	Chlorocebus aethiops	-	3.7	-	-	Inhibitor	Reporter efflux	1992 Hollt - Stereoisomers of calcium antag...

Figure 2-5: A screenshot of search results for the transporter protein ABCB1 in Kinetic Transportal displaying field filters at the top, the results table at the bottom, and table display options just above the table.

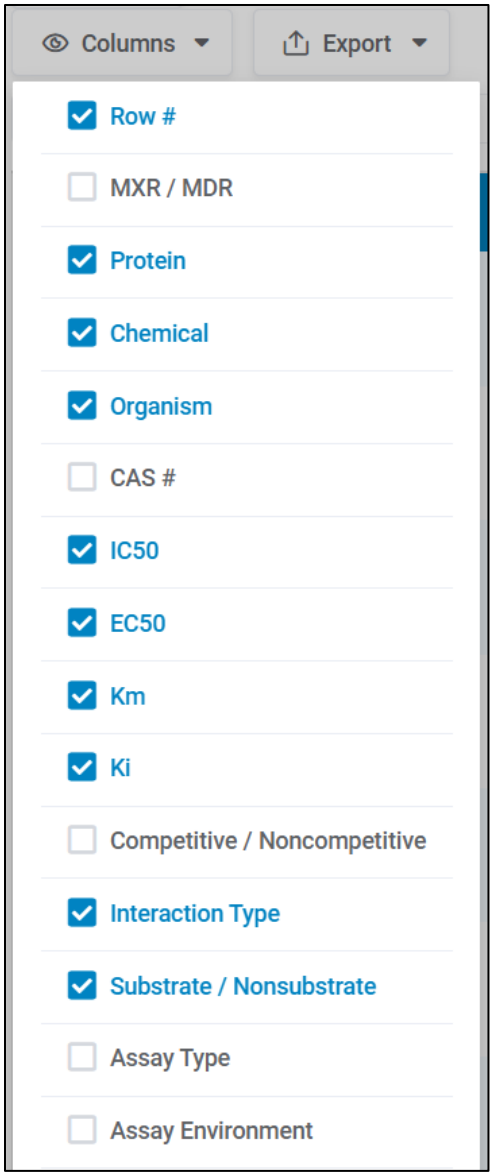
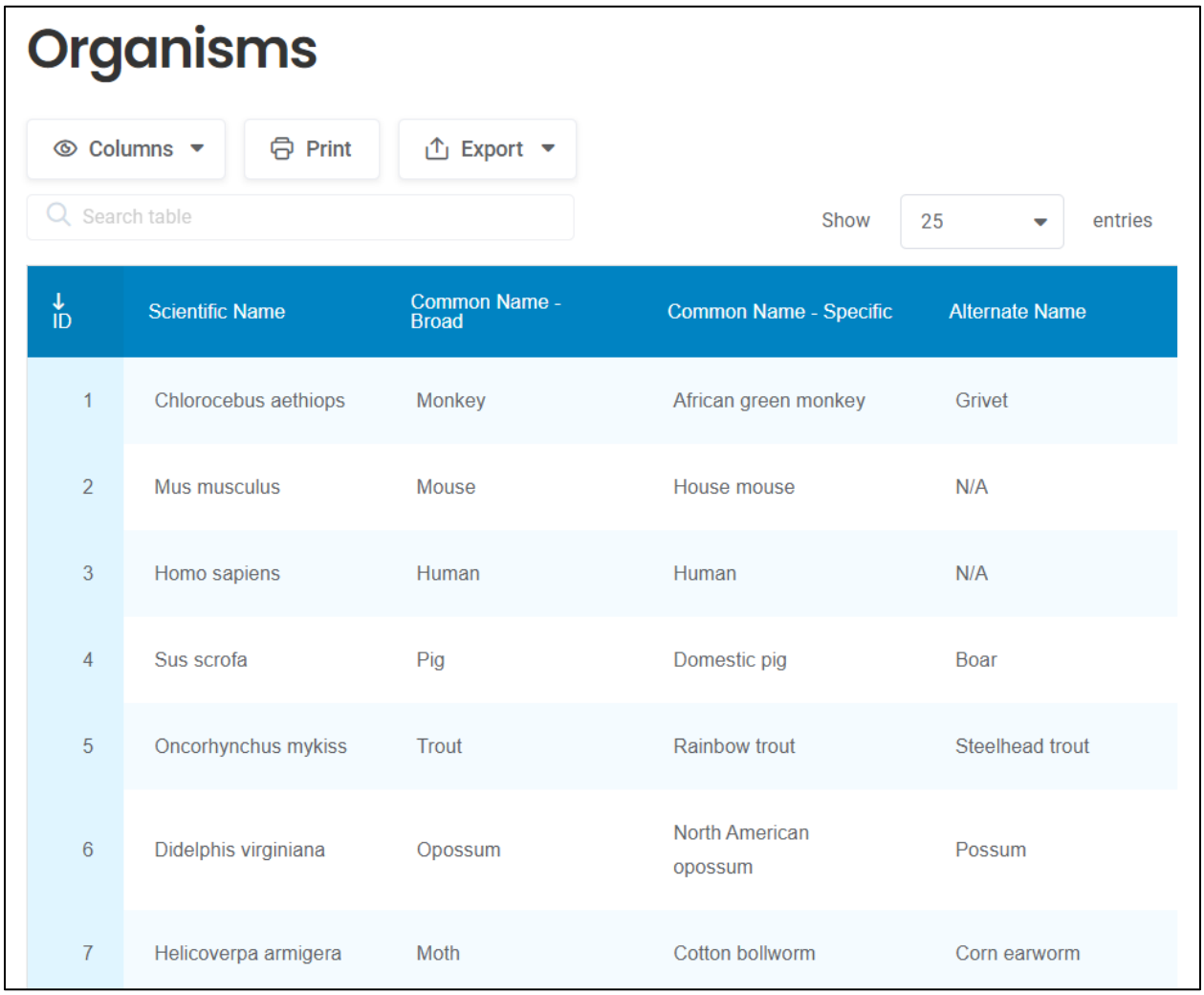


Figure 2-6: A screenshot of the column display filter drop-down menu.

The metadata for UC Transportal consists of four separate data tables: Transporter Proteins, Chemicals, Organisms, and Articles (Figure 2-2). The first three display alternate names for their search parameters to assist users in identifying the data they are interested in. All metadata tables contain information from other publicly accessible and well-established databases like PubChem (https://pubchem.ncbi.nlm.nih.gov) and NCBI Gene (https://www.ncbi.nlm.nih.gov/gene) which verify the identities of values displayed in Kinetic Transportal. This is to ensure absolutely clarity about the information being presented in UC Transportal by preventing ambiguity regarding

nomenclature and maintaining accurate data communication. Like the primary data table, the metadata tables have a set of table display options at the top, including the universal search bar which is the primary means of searching them (Figure 2-7).



ID	Scientific Name	Common Name - Broad	Common Name - Specific	Alternate Name
1	Chlorocebus aethiops	Monkey	African green monkey	Grivet
2	Mus musculus	Mouse	House mouse	N/A
3	Homo sapiens	Human	Human	N/A
4	Sus scrofa	Pig	Domestic pig	Boar
5	Oncorhynchus mykiss	Trout	Rainbow trout	Steelhead trout
6	Didelphis virginiana	Opossum	North American opossum	Possum
7	Helicoverpa armigera	Moth	Cotton bollworm	Corn earworm

Figure 2-7: A screenshot of the Organisms metadata table in UC Transportal.

The *Contribute Data* link connects to area of UC Transportal containing all functionalities related to data management, including importing new data. This section contains six pages, four of which are restricted to administrators (Figure 2-8). All registered users have access to the *My Entries* and *Contribute Data* pages. The *My Entries* page allows members to view all entries they have contributed to the database and check their approval statuses. It remains invisible until the user has submitted at least one entry. The *Contribute Data* page is where users can contribute new

data to UC Transportal. The *New Entries*, *Export data to CSV*, *Import Approved Data*, and *Import Metadata* pages are all for administrators to carry out the data approval process and perform quality checks for all incoming data. All these pages and their functionalities are described in greater detail in section 4 – *Uploading Data to UC Transportal*.

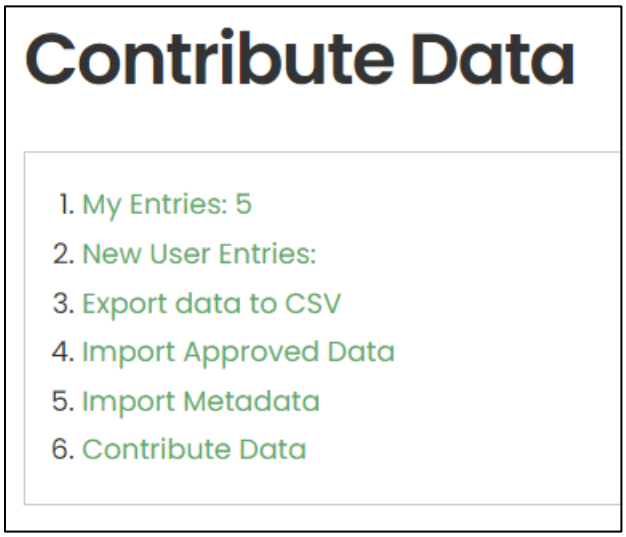


Figure 2-8: A screenshot of the Contribute Data menu for users with administrator access. Links numbered 2-5 are only available to administrators.

The *Our Team* page contains short paragraph about the Nicklisch Lab at UC Davis (https://nicklischlab.faculty.ucdavis.edu). In the future it will also display some information about each of the UC Transportal administrators and their roles. The *Contact* page holds a contact form for users to send messages to the administration team at uctransportal@ucdavis.edu. Lastly, the *MyAccount* page holds only the menu in Figure 2-8 to be used for navigation.

2.3. Backend Administrator Functions and Email Automation

All data management tasks can be accomplished through functions directly available in UC Transportal on the "front-end". However, certain tasks related to site administration require administrators to use "back-end" site management features in WordPress and its plug-ins. This includes i) approving new users, ii) altering the content of the website, iii) the appearance of the background, tables and menus, iv) updating the data templates, and v) managing email automation.

Users will receive email notifications in the following instances: applying for UC Transportal membership, membership approval, when submitting data, and upon data approval or other status changes. An example email is shown in Figure 2-9.

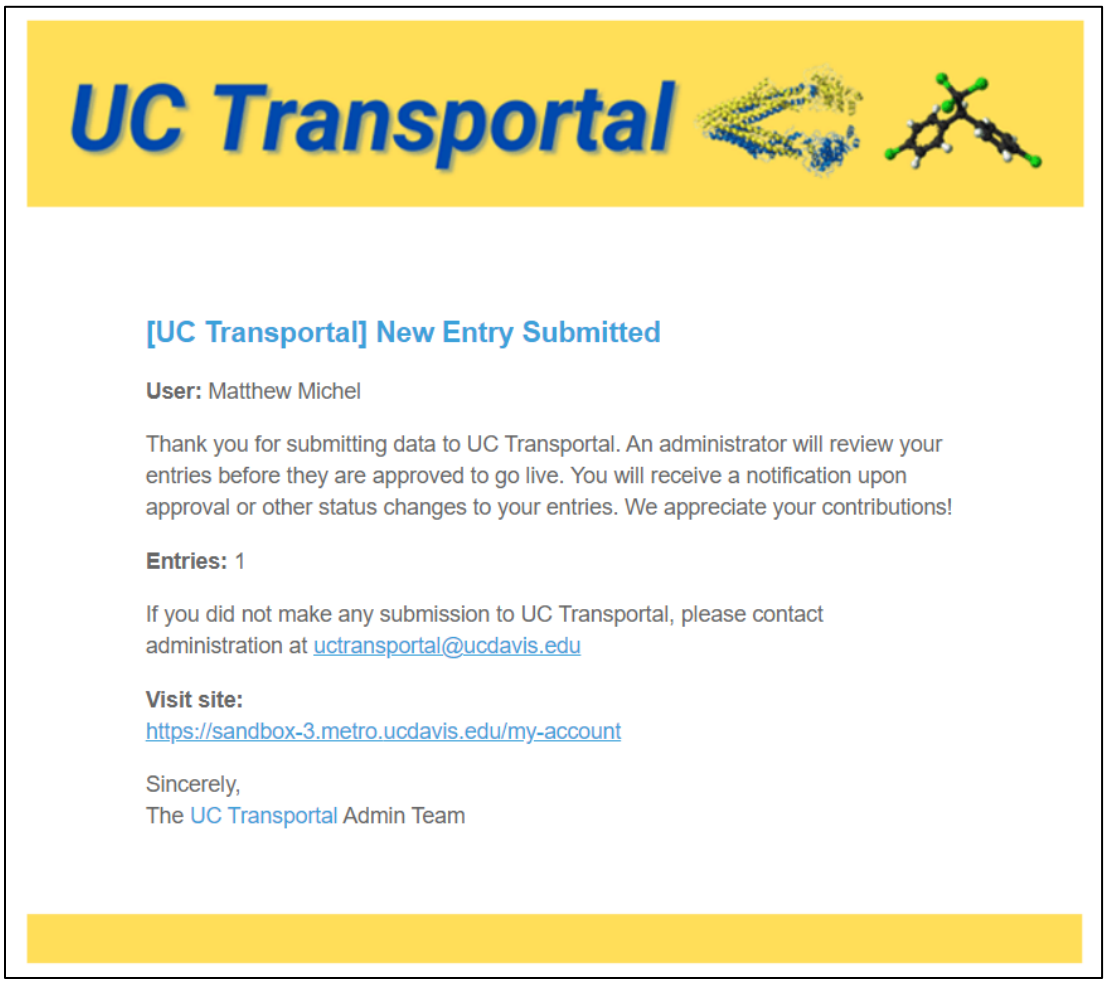


Figure 2-9: A screenshot of an automated email sent upon the submission of data to the Kinetic Transportal.

3. Kinetic Transportal

The Kinetic Transportal database is the first and primary database presented in UC Transportal, intended to be the "flagship" database of the system. It was designed to align with the goals and rationales of UC Transportal. The following is a detailed description of the database, presented as an instruction document for incoming data curators.

3.1. Introduction

The role of data curators for the Kinetic Transportal database is to thoroughly review primary scientific research articles on active transporter proteins and incorporate data on transporter-chemical interactions into the database. High value is placed on the accuracy and completeness of the data. Therefore, it is important to gain a thorough understanding of the standards and conventions of the database, the meaning of its fields and datatypes, and how to efficiently curate appropriate information from source materials. Interpreting article results from across the literature is not always a straightforward task, but knowledge about how literature from fields relevant to transporter research is written will eventually make interpreting and extracting the relevant information about transporter-chemical interactions much easier.

The purpose of Kinetic Transportal is to serve as a central location for data describing direct transporter-chemical interactions. This data consists of **primary data** and **secondary data**. The primary data is the information that gets curated directly from primary research articles into the database and which characterizes the transporter-chemical interactions. The secondary data is metadata about the primary data: detailed information defining, cataloging, and classifying values from important data fields contained within the primary data. The metadata largely has to do with standardizing nomenclature within the database to make it simple to search and analyze, and is pulled from other well-established databases (PubChem, for example, supplies most of the metadata on chemicals).

All the primary data in Kinetic Transportal fits within the framework of a single spreadsheet (Figure 2-10). To adequately describe and contextualize an observed transporter-chemical interaction from the literature, this spreadsheet contains 22 data fields. One row (an "entry") constitutes a description of a transporter-chemical interaction from the literature and each field per row (a "cell") contains a piece of information characterizing that interaction. Some of

these data fields may be left blank for a given interaction either because the article does not provide a value or it is not applicable, but most fields are always filled in. To help keep track of the different data fields, they have been grouped into five categories and will be described later in-depth. Table 2-1 provides an overview of the fields and their groupings.

In contrast to the primary data, the secondary data is organized across several spreadsheets, one for each class of metadata. There are individual metadata tables pertaining to the Transporter, Chemical (Figure 2-11), and Organism fields, one pertaining to the Assay Information field category, and one pertaining to the Reference Information field category. For every unique value from these data fields that is introduced into the primary data, one row in its corresponding metadata table should be filled out to define that value. A detailed description of each of these tables will be given in section 3.5.

Table 2-1: All 22 primary data fields in the Kinetic Transportal database and their five categories. The Notes field does not fit into a field category

Primary Parameters	Transporter Protein
	Chemical
	CAS Number
	Organism (protein source)
Quantitative Kinetic Data	IC_{50} (µM)
	EC_{50} (µM)
	K_m (µM)
	K_i (µM)
Interaction Types	Competitive / Noncompetitive
	Interaction Type
	Substrate / Nonsubstrate
Assay Information	Assay type
	Assay environment
	Expression
	Chemical Concentration (µM)
	Reporter
	Reporter Concentration (µM)
	Measurement Method
Reference Information	Pubmed ID
	DOI (Direct Object Identifier)
	Reference Name
Notes	Notes

Transporter Protein	Chemical	CAS #	Organism (protein source)	IC_{50} (μM)	EC_{50} (μM)	K_m (μM)	K_i (μM)	Competitive / Noncompetitive	Interaction Type	Substrate / Nonsubstrate	Assay Type
SLCO1A2	Perfluorooctanoic acid	335-67-1	*Homo Sapiens*						Inhibitor		Reporter uptake
SLCO1A2	Perfluoropentanoic acid	2706-90-3	*Homo Sapiens*						Weak interactor		Reporter uptake
SLCO1A2	Perfluoroundecanoic acid	2058-94-8	*Homo Sapiens*						Inhibitor		Reporter uptake
Abcb1a	Aldrin	309-00-2	*Mus musculus*	26.2					Inhibitor		ATPase
Abcb1a	Alpha-hexachlorocyclohexane	319-84-6	*Mus musculus*	26.8					Inhibitor		ATPase
Abcb1a	Beta-hexachlorocyclohexane	319-85-7	*Mus musculus*						Weak interactor		ATPase
Abcb1a	Bisphenol A	80-05-7	*Mus musculus*						Weak interactor		ATPase
Abcb1a	Dieldrin	60-57-1	*Mus musculus*	21.8					Inhibitor		ATPase
Abcb1a	Diethylhexyl phthalate	117-81-7	*Mus musculus*						Weak interactor		ATPase
Abcb1a	Gamma-hexachlorocyclohexane	58-89-9	*Mus musculus*	82.6					Inhibitor		ATPase
Abcb1a	Heptachlor	76-44-8	*Mus musculus*	10.4					Inhibitor		ATPase
Abcb1a	Hexachlorobenzene	118-74-1	*Mus musculus*						Weak interactor		ATPase

Assay Environment	Expression	Chemical Concentration (μM)	Reporter	Reporter Concentration (μM)	Measurement Method	PubMed ID	DOI
HEK293 cells	RE	100	[3H]estrone sulfate	0.007	Reporter radioactivity	20639259	10.1093/toxsci/kfq219
HEK293 cells	RE	100	[3H]estrone sulfate	0.007	Reporter radioactivity	20639259	10.1093/toxsci/kfq219
HEK293 cells	RE	100	[3H]estrone sulfate	0.007	Reporter radioactivity	20639259	10.1093/toxsci/kfq219
Purified, detergent-solubilized protein	RE	.017-333.33**	Verapamil [prestimulation]	100	Phosphate complex absorbance	27152359	10.1126/sciadv.1600001
Purified, detergent-solubilized protein	RE	.017-333.33**	Verapamil [prestimulation]	100	Phosphate complex absorbance	27152359	10.1126/sciadv.1600001
Purified, detergent-solubilized protein	RE	.017-333.33**	Verapamil [prestimulation]	100	Phosphate complex absorbance	27152359	10.1126/sciadv.1600001
Purified, detergent-solubilized protein	RE	.017-333.33**	Verapamil [prestimulation]	100	Phosphate complex absorbance	27152359	10.1126/sciadv.1600001
Purified, detergent-solubilized protein	RE	.017-333.33**	Verapamil [prestimulation]	100	Phosphate complex absorbance	27152359	10.1126/sciadv.1600001
Purified, detergent-solubilized protein	RE	.017-333.33**	Verapamil [prestimulation]	100	Phosphate complex absorbance	27152359	10.1126/sciadv.1600001
Purified, detergent-solubilized protein	RE	.017-333.33**	Verapamil [prestimulation]	100	Phosphate complex absorbance	27152359	10.1126/sciadv.1600001
Purified, detergent-solubilized protein	RE	.017-333.33**	Verapamil [prestimulation]	100	Phosphate complex absorbance	27152359	10.1126/sciadv.1600001
Purified, detergent-solubilized protein	RE	.017-333.33**	Verapamil [prestimulation]	100	Phosphate complex absorbance	27152359	10.1126/sciadv.1600001

Figure 2-10: Two screenshots of a small portion of the primary data table. Twelve chemical-transporter interactions are displayed. The top panel shows data fields 1-12; the bottom panel shows data fields. 13-20. Data fields 21-22 (Reference and Notes) to the right are not shown.

Primary Identifier	First Alternative	Second Alternative	Third Alternative	CAS Number	PubChem ID
Quintozene	Pentachloronitrobenzene	Brassicol		82-68-8	6720
Resveratrol	Trans-resveratrol	3,4',5-Trihydroxystilbene		501-36-0	445154
Rhodamine 6G	Basic Red 1	Rhodamine F5G		213-584-9	13806
Thiabendazole	Tiabendazole	Mintezol	Equizole	148-79-8	5430
Trifloxystrobin	Stratego			141517-21-7	11664966

Figure 2-11: A screenshot of a small portion of the Chemical Metadata spreadsheet. Shown are five chemicals and their descriptive parameters. The Abbreviation and Notes data fields to the right are not shown.

3.2. Curation Process Overview

Prior to the beginning of the curation process, a literature search is carried out to identify primary research articles which may contain data that can be added to the database (see section 5.1). All such articles are collected from reputable journals and kept in a central repository until the time of curation. The only tools necessary for data curation are Microsoft Excel, software to open PDF files, and internet access to refer to certain publicly available databases such as PubMed and PubChem.

There are three main steps to the data curation process:

1) Selecting an article and recording the reference metadata
2) Reading the article and extracting information into the primary data spreadsheet
3) Updating the metadata tables and standardizing data from the new article

The curation process is performed one article at a time. Once all appropriate data has been entered into the primary and secondary data spreadsheets, the article's PDF file should be named according to the Reference Name field from the primary data table and stored in the central article repository for fully curated articles in the Nicklisch Lab file system.

Step one takes no more than five minutes. Step two usually contains 80-90% of the work and can take anywhere from fifteen minutes to two hours depending on the length of the article, how much relevant data it contains, the number of different assays performed, and the clarity of the writing. Step three varies significantly in time and difficulty: it can take very little effort if the nomenclature from the article already matches the database conventions, and their values are already present. However, it can be a very time-consuming step if the article contains many proteins or chemicals new to the database, especially if they are obscure or not clearly defined in the article. Much care needs to be taken to record the correct nomenclature of new entities in the

Primary Parameter fields because that nomenclature essentially defines the identity of the interactions in the database.

Curators will work exclusively with a working template containing all the primary and secondary data (metadata) spreadsheets. A set of master backup spreadsheets containing all data in Kinetic Transportal exists separately and is updated with new data from the working template periodically. These are called "Kinetic Transportal Data Archive" appended with the date of backup and serve as a form of version control.

3.3. Recording the Reference Metadata

To record an article's reference metadata, first open the Article Metadata table in Microsoft Excel. Then, search for the article by its title in PubMed. When found, a new row in the spreadsheet is started and the full title of the article, the PubMed ID, the DOI (digital object identifier), year of publication and the last name of the first author are copied and pasted into the proper fields. Initials are entered into the Curator ID field and "In progress" is entered into the Curation Date field as a temporary placeholder. The date of curation is entered only *after* finishing curation of the article. The curation date entry is the indicator that an article has been fully curated, renamed as described above, and moved to the central article repository for storage. Since it is ideal not to create backups with articles in the middle of the curation process, curators should attempt to not leave articles unfinished over multiple days.

3.4. Curating Data from Primary Research Articles

3.4.1. Kinetic Transportal Data Fields

This is the main substance of the data curation process. Once an article has been identified to curate data from, information from any conclusive result characterizing direct transporter-chemical interactions will be extracted. The 22 data fields give an adequate description of almost

any interaction and are grouped by their purpose into five categories (Table 2-1). Below is a detailed description of the contents of each data field in order, its meaning, and an explanation of the purpose of each field category.

Primary Parameters: Transporter Protein, Chemical, CAS #, Organism (protein source)

The Primary Parameters are the defining elements of each interaction. Users search the database by entering a transporter, chemical, organism or combination thereof into the three search bars of the Kinetic Transportal main search form. These parameters are searched in combination to display the interaction(s) the user is looking for if they exist.

Transporter Protein: This is the transporter protein involved in the transporter-chemical interaction. UC Transportal is focused on *active* transporters of the ABC and SLC superfamilies, not on ion channels or other passive transporter proteins. There are multiple nomenclature systems for different classes of transporter, which can be confusing. Therefore, all transporter protein nomenclature in the database is standardized to the ABC/SLC/SLCO nomenclature system whenever possible. For human proteins this is in accordance with the naming of the HUGO Gene Nomenclature Committee (https://www.genenames.org/). Non-human organisms usually have their own convention(s) which may or may not include ABC/SLC nomenclature. Whenever *new* transporters (including homologues of existing proteins) are entered into the database, you will need to update the protein metadata. Refer to the NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene) whenever possible.

Oftentimes with *in vivo* studies or studies on organisms with scarce transporter data, the specific transporter involved in an interaction cannot be identified. Instead, the measurement refers to an interaction between a chemical and an unknown transporter or a combination of several possible transporters. These interactions are classified as “MXR-activity” and their results can be

toggled on and off from showing up in the search results by database users searching online. Other than the non-specificity of their transporter, these interactions are the same as other interactions, hence the MXR-activity spreadsheet is essentially the same as the transporter-specific spreadsheet. The hierarchy of specificity for these interactions is as follows:

1. "MXR activity" designates no determination of a specific transporter.
2. ABC- or SLC-superfamily specifies determination of the superfamily only.
3. "ABCC-family" or similar entries indicate a more specific family of protein(s) was determined.
4. ABCB1-like, SLC22A6-like or similar entries indicate a specific protein resembling a known protein but without a confirmed identity (this can happen when a protein is identified using an antibody for a homologous protein).

Chemical: This is the chemical involved in the transporter-chemical interaction. The chemical's role in the interaction is described by the Interaction Type category fields; it is the chemical which is "acting on" the transporter (it is an *actor*, as opposed to the Reporter which is *affected*). Chemicals are reported in their most specific stereoisomeric forms when possible. Occasionally, assumptions about isomerism are unavoidable because one form or a mixture is predominant in nature or in human use. Such assumptions are stated in the Chemical metadata when a new entry is added (see section 3.5, Chemical Metadata). When the stereoisomeric form is unstated and cannot be reasonably assumed, the designation "(nonspecific)" is added to its name.

CAS #: Most chemicals in the database have a CAS (Chemical Abstract Service) number, a unique identification for every (registered) substance, including isomers and mixtures. These are entered into the chemical metadata and retrieved by the primary data table using a VLOOKUP function in Excel.

Organism: This is the organism of origin of the transporter protein in the interaction. It is *not* always the same organism as an *in vivo* species being used or the organism from which *in vitro* cells came from in the study (i.e., homo- and heterologous expressions). The scientific name of the genus and species is used, but subspecies are disregarded. Common names for use in searching the database are included in the Organisms metadata table (see section 3.5, Organism Metadata).

Quantitative Kinetic Data: IC_{50}, EC_{50}, K_m, K_i

When an interaction is studied in detail, researchers may characterize its kinetics quantitatively and produce an IC_{50}, EC_{50}, K_m, or K_i value. A large proportion of studies do not produce quantitative kinetic data; therefore, most rows will be left blank in these data fields. Other quantitative values, such as ratios of transport between different cell lines, are not included in the database. Quantitative kinetic values are usually stated plainly in articles and easily recorded. All quantitative parameters in the database should be reported in micromoles per liter (µM), the default unit. If unit conversion is not feasible, the reported units with the numeric value are stated. Bounded values reported with a greater-than symbol can be recorded but should always be interpreted as a "weak interactor" for inhibition interactions (See Interaction Types). Below is a brief definition of each of the quantitative kinetic parameters.

IC_{50}: This is the half maximal inhibitory concentration of the chemical upon the transporter activity, as reported by the article.

EC_{50}: This is the half maximal effective concentration of the chemical either effecting stimulation or inhibition of the transporter. The exact way this is defined is dependent upon the context of the experiment and how the author of the article has analyzed and presented the data.

K_m: This is the Michaelis-Menten constant of transport kinetics for a chemical substrate with the transporter protein. It is the concentration of the chemical at which its transport activity will be half of its maximum velocity (V_{max}).

K_i: This is the inhibitory constant, or the concentration at which a chemical binds to 50% of the transporter protein.

Interaction Types: Competitive/Noncompetitive, Interaction Type, Substrate/Nonsubstrate

The Interaction Type fields are the main characterization of the transporter-chemical interaction, and therefore are arguably the most important data in Kinetic Transportal aside from the Primary Parameter fields. Due to the experimental setup in each publication, it is rare for all three data fields to contain data for a single interaction. However, at least one of the three data fields in this group *must* be filled out for any interaction. The data in these fields is not based on preexisting knowledge of the curator; it is based on information directly from the article and will always be associated with an assay described in the article (see Assay Information). The only exception is for cases when an interaction type result is not directly stated but can be very reasonably inferred through interpretation of the article (see Technical Conventions).

Competitive/Noncompetitive: This refers to whether the chemical interacts with the protein at the transporter protein's active site (competitive) or somewhere else on the protein (noncompetitive). It is analogous to orthosteric or allosteric binding, respectively. Most articles do not investigate competitive versus noncompetitive binding, so this field will be left empty most of the time. This field receives the most usage of asterisked values because even though competitiveness may be determined by an assay it is often not of interest to authors and therefore may not be directly stated. This field should never be filled in for a "Weak interactor" because to for competitiveness to be a relevant descriptor, a significant interaction must be observed.

Interaction Type: This is the single most important descriptor of the transporter-chemical interaction. The three possible values for this data field are “Inhibitor”, “Weak interactor”, and “Stimulator”. An inhibitor significantly decreases the transport activity of the transporter protein, a weak interactor does not have a significant effect, and a stimulator significantly increases transport activity of the transporter protein. Compared to inhibitors and weak interactors, reported stimulators are rare. If the article specifies cis/trans inhibition or stimulation, record it in the Notes field. Cis/trans stimulation differentiates between stimulation of the transporter from the same or opposite side of the membrane as the substrate chemical and only applies to SLC-superfamily symporters or antiporters.

Substrate/Nonsubstrate: This describes whether the chemical was determined to be a substrate of the transporter protein or determined not to be a substrate of the transporter protein. Typically, substrates are unambiguously identified in cellular assays where a given chemical needs to cross a membrane barrier facilitated by the investigated transporter.

Assay Information: Assay Type, Assay Environment, Expression, Chemical Concentration, Reporter, Reporter Concentration, Measurement Method

The Assay Information fields give context to how the interaction data (the results) were obtained and can be very important for interpretation because observations of interactions display a large amount of variation depending on the assay type, sensitivity, environment, detection method, chemical concentrations used, and other decisions made by the investigator.

Assay Type: This is the general assay type that was used to produce the data being reported. There is a limited number of predefined assay types which an assay can be categorized as, listed below. It is important to become familiar with all of them because knowing how they work will greatly assist your ability to read and interpret articles. Sometimes the Reporter, Reporter

Concentration, and Measurement Method data fields need to be filled out differently based on the Assay Type (see respective sections).

- ATPase assays
    - Transporter activity is measured indirectly through the level of ATPase activity during the reaction. This can be done by measuring the release of phosphate, the depletion of ATP or by other methods. The reporter for ATPase assays is a baseline level of activity and not a molecule, unlike most other assay types.
- Cell-viability assays
    - Transporter activity is measured by using cell growth or cell mortality as a proxy. A stressor chemical is used to challenge cell growth or survivability and the chemical of interest interferes with transporters' ability to remove the stressor. Cancer cell lines or single-cell organisms such as yeast may be utilized.
- Monolayer assays
    - A chemical is introduced to one side of a cell monolayer and its transport to the other side of the monolayer is measured. Transport of the chemical is affected by the presence of transporters in the cell membrane and often a ratio between apical-to-basal and basal-to-apical transport is produced.
- Ligand binding assays
    - A ligand molecule binds to a site on the transporter of interest and either its binding is measured or its effect on the transport or binding of other molecules is measured. Most ligand binding assays use radio-labelled ligands.
- Reporter efflux assays
    - The transport of a reporter molecule out of a cell or vesicle is measured.

- Reporter uptake assays
    - The transport of a reporter molecule into a cell or vesicle is measured.
- Tissue distribution assays
    - The accumulation of a substance in one or more body tissues is measured in an *in vivo* environment after administration. Measurement of chemical accumulation in tissues may be performed using a variety of techniques, with or without sacrificing the organism.
- Toxic endpoint assays
    - A challenge chemical is introduced to an organism *in vivo* and the toxic response is evaluated. Mortality may be an endpoint, but for single-cell organisms such as yeast these assays are categorized as cell-viability assays.

Assay Environment: This is the medium of the assay. For *in vitro* studies it is usually the cell line that was used but it can be any membrane environment in which the protein is situated. It is exceedingly common for cell lines to be modified by transfection, upregulation of a transporter, or by other means (see Expression) and given extended, modified names. Only the basic cell line name should be recorded, otherwise there would be too many variation to feasibly record in the assay metadata (see section 3.5). For example, HEK293 cells may be transfected with ABCB1 and called HEK293-MDR1 cells. The basic cell line is HEK293, so "HEK293 cells" is the correct value to enter. Be sure to state whether they are cells, membranes, vesicles etc. and refer to the Assay Metadata sheet to match them with previous entries when appropriate. For *in vivo* studies, the assay environment is the live organism on which the assay is being performed. The scientific name should be given and "(in vivo)" appended to the name. Ex-vivo tissues should be given a precise description (for example "isolated gill tissue").

Expression: This is the manner of expression of the transporter protein under investigation. It is usually also the method by which the researcher assures that the protein under investigation is responsible for the interaction being observed. The expression types have the following abbreviations and definitions:

- EE = endogenous expression
    - Transporter expression is from endogenous genes in a normal cell environment.
- RE = recombinant expression
    - The cell line was transfected with the gene of the transporter protein.
- UR = upregulated
    - For *in vitro* assays, the cell line was cultured in a substance known to induce greater expression of the transporter. For *in vivo* assays, a strain of organism confirmed to express greater amounts of the transporter (often a pesticide resistance mechanism) was used.
    - UR is always jointly stated with another expression type (ex. EE, UR)
- RC = reconstituted
    - The transporter has been purified or partially purified from its membrane environment and then returned to a functional membrane-bound state
    - PF = purified, may be jointly stated with RC (PF, RC)
- KO = knockout
    - The organism has been genetically manipulated in some manner to disable the expression of the transporter. The nullizygous genotype falls under this category.

Chemical Concentration: This is the range of concentrations at which the chemical being tested was used in the assay or set of assays performed. The default unit of concentration is

micromole per liter (μM). There may be just one concentration or a range of concentrations. The lower limit of a range is only counted as zero if a zero concentration is stated or shown explicitly, even though practically the concentration in negative controls is often zero. This is to prevent the lower limit of testing from becoming meaningless.

The chemical concentration data is frequently the most obscure to find in an article. Whenever concentrations must be inferred from visual graphs or charts, such values should be reported as approximate values (see Technical Conventions).

Reporter: This is what was used to make measurements in the assay. In most assay types it is a molecule that is quantified through mass spectrometry, radioactivity, fluorescence etc. However, for certain assay types, the reporter should be identified differently.

In ATPase assays the reporter is a response baseline. It is either the "Basal activity" of ATPases in the assay or the level of prestimulation activity provided by a prestimulator, in which case enter the name of the prestimulator followed by "[prestimulation]". (Example: "Verapamil [prestimulation]")

In cell-viability assays the reporter can be the chemical used to directly measure cell-viability. Examples of these are dyes like MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) and WST-1 (sodium 4-[2-(4-iodophenyl)-3-(4-nitrophenyl)tetrazol-2-ium-5-yl]benzene-1,3-disulfonate) which are measured using spectrometry. However, the reporter could also be a toxic challenger to the cell (i.e., cytotoxic substrates), which is typically acted on by the transporter (doxorubicin and paclitaxel are good examples). This is preferred if the challenger concentration is available because the Reporter Concentration field can then be filled in with more meaningful information. Also, in this case, the Measurement Method is usually Cell Mortality.

In toxic endpoint assays, the reporter could be the toxic challenger molecule, or it could be the toxic endpoint directly being examined. In this case, the toxic endpoint is preferred, and a concise description should be given (Example: “Lung injury”).

In the case of assays meant to determine if a substance is a transporter substrate, the reporter is often the same as the chemical or is a radio-labelled variant.

Reporter Concentration: This is concentration of the reporter if the reporter is a molecule. It is much less common for there to be a range of concentrations for the reporter than for the chemical of interest.

Measurement Method: This is the method of measuring the outcome of the assay results or the kind of measurement of the reporter that was taken. When possible, refer directly to the “reporter” in this field (Example: “Reporter absorbance”). It is typically the most direct data that would be obtained from the assay.

Reference Information: PubMed ID, DOI, Reference Name

Each row of data must be directly connected to the reference it was gathered from to ensure users of its veracity. For all data from a given article, all reference information is identical and can simply be auto-filled in Excel.

PubMed ID: This is a unique identification number given to every article in PubMed. Not all articles can be found in PubMed, however this is a very small proportion. Please note, the PubMed ID is not the same as the PubMed Central ID.

DOI: DOI stands for Digital Object Identifier. It is an identification system for documents adopted by most prominent scientific publications. Almost all articles you will come across will have a DOI which directs you to their location on the internet.

Reference Name: The article from which the data was curated is displayed in the primary data table as the year of publishing followed by the first author of the article and the title truncated to thirty characters (ex. “2020 Tikkanen - Food Additives as Inhibitors o...”). This data is retrieved from the Articles secondary data table using a series of VLOOKUP functions connected to the DOI column, so this field can be filled by copying the formula once the DOI has been entered.

Notes:

The Notes field does not neatly fit into another category. It is a space to enter any information that might be useful, but which does not neatly belong to any one data field or requires a written explanation. Notes can include personal interpretations and communications related to the data point and article. Notes with a double asterisk in front pertain to a specific cell in that row marked by the curator (see Technical Conventions).

3.4.2. Practical Tips for Curating Articles

1. Most articles are broken into sections based on the different assays that the investigator ran, one by one. It is most efficient to curate article data one assay-based section at a time to stay organized. Oftentimes this allows copying sets of values multiple times as you fill out rows because researchers will perform the same experimental setup several times with a different proteins or chemicals.
    a. As a result, it is far more efficient to fill in sets of entries at once instead of one row at a time, making extensive use of the autofill features in Excel.
2. The abstract can give a useful overview of the data but information should not be taken directly from it. It can be informative about what to look for as you read. The introduction section can be skimmed or skipped, as they rarely contain collectable information, However, abbreviations and definitions are occasionally given. All the information you

gather should come from the materials and methods and the results sections. The discussion section also rarely contains useful information.

3. The first information to search for (after the reference data) is usually the cell line or other assay environment data. This is usually clearly stated near the beginning of the materials and methods section.
    a. Expression data is often found with the cell/assay environment data as part of a description about how the transporter protein was verified to be present in the environment.
    b. Western blotting procedures usually contain information verifying the identity of one or more transporters of interest but do not contain information that can recorded in UC Transportal. Therefore, western blotting procedures can be skipped over.
4. If the specific protein responsible for the observed results is unclear throughout the article, it is a good indication that a specific protein may not have been investigated and the data will be MXR-type. Sometimes a specific protein is suggested to have been involved but direct evidence is not given.
    a. Often the clearest indication of this is through the cell line and expression data. When the expression type is EE (endogenous expression), the interaction often belongs with the MXR-activity data.
5. Researchers will often communicate assay results in the form of charts and only highlight certain data in the article text. When interpreting assay results from charts, only count statistically significant results as inhibitors or stimulators. Statistically nonsignificant results compared to controls are considered weak interactors. Significance levels are usually indicated by a symbol.

6. When researchers find a substrate of particular interest, they will often repeat their initial assay at a range of substrate concentration to determine the K_m value. This is often stated separately from the initial assay, but a separate row indicating another interaction should not be made. If all other fields are the same as the initial experiment, the initial row can simply be updated with the K_m value and the expanded chemical concentration range used in the assay.
    a. Similar cases occur for inhibitors and their IC_{50} or K_i values as well.
7. When populating primary data table, it is okay not to follow the protein and chemical nomenclature conventions at first. If the article uses a different nomenclature, it is okay to use that nomenclature for ease of keeping track of information while curation is ongoing. When it is time to update the metadata tables and make sure everything adheres to the proper nomenclature standards, then values can be changed to the proper names.

3.4.3. Technical Conventions

Consistent with the level of detail given to establishing proper naming conventions and standardized values in the data fields, the following is a set of technical conventions for all individual cell values in the primary data table:

- For stereoisomers, the (+)/(-) convention is preferred to the R/S or D/L conventions when the meaning would be equal (ex. (+)-niguldipine, (-)-epigallocatechin).
- Abbreviations for chemical names are avoided in most cases, but overly technical organic nomenclatures are also avoided. The common name of a chemical is used whenever possible, and abbreviations and more technical names are stored in the metadata tables. See section 3.5, Chemical Metadata, for examples.

- Values always begin with capital letters except when the first character is not a letter. In such a case the first letter is not capitalized unless another convention says to (e.g. element abbreviations).
- Radioactively labeled compounds, often used as reporter substances, are written with the radioactive labeling in brackets:
    - [3H]verapamil for tritium-labeled verapamil
    - [14C]parathion for carbon-14-labelled parathion
- Conjugated complexes are hyphenated according to common convention if one exists
    - Examples: Cadmium(II)-glutathione, Bodipy-verapamil
- A tilde indicates a value that is approximate.
- An asterisk is used to indicate values which are inferred by the curator but not clearly indicated by the author of the paper. It is used for cases when communication in the article is unclear or ambiguous and the reader must make an inference.
- A double asterisk indicates additional information for the value in the Notes column.
- UC Transportal uses American English, and all spellings match accordingly (ex. “sulfur”, not “sulphur”, and cyclosporine, not cyclosporin).

3.4.4. Other Data Types

Although Kinetic Transportal is focused exclusively on describing direct transporter-chemical interactions kinetically, many articles contain data that will be included in future databases. The other data types to watch out for are the following:

1) Chemical-gene expression interaction data.
2) Transporter expression and localization data (transcriptomics and proteomics).

3) Clinical drug-drug interactions (DDI) or drug-environmental chemical interactions (DECI) involving transporters.

If any of these categories of data are present, a note should be made in the spreadsheet titled "*Articles with Other Data Types of Interest*".

3.5. Updating the Metadata Tables

Once the primary data table has been populated with all relevant information from an article, the secondary data tables (metadata) need to be checked. Any values new to the database need to be added to the appropriate table and values in the primary data table need to be standardized if not already. Updating the secondary data tables with new values involves (1) making sure the new entity is actually new, (2) identifying the best name for the value (the "primary identifier"), (3) identifying useful alternative names, and (4) verifying that the values in the primary data table match primary identifier. The following is a description of each metadata table, presented in the suggested order of updating them.

Assay Metadata

The Assay Metadata sheet holds information on several data fields in the Assay Information category. Metadata for the Assay Environment field is stored here as well as a list of valid values for the Assay Type and Expression fields (Figure 2-12).

The Assay Environment field has two additional metadata fields corresponding to it: Assay Environment Type and Assay Environment Name. The Assay Environment Type has a very controlled and limited number of values. The main values are "Cell", "Membrane", "Vesicle", "Tissue", and "In-vivo". There is also a set of "Other Environment Types" for uncommon assay environments. The Assay Environment Name is the designation or "name" of a membrane, cell, tissue, etc. type which distinguishes it from others. The proper Name rule depends on the Type.

For Cell, Membrane, and Vesicle, the Name is the standard written designation. For example, the name for "Sf9 cells" would be "Sf9", and for "HEK293 membrane vesicles", the name would be "HEK293". For our definitions, all Tissue is considered *ex vivo* and the name should be the same as the Type: a short, concise description of the tissue. For in *vivo* environments, the Name is the scientific name of the organism. For Types that fall in the "Other" category, the name field is left blank.

New additions to the assay metadata should be considered carefully since these values are very controlled (see Appendix 2, section 2). Almost all assays and expression types can be categorized within preexisting values, therefore, adding to them should be done rarely and represents a significant change. Only the Assay Environment data should receive new values with any plausible regularity.

Organism Metadata

The Organism Metadata spreadsheet contains the fewest entries of the three tables for the Primary Parameters (Figure 2-13). UC Transportal uses the scientific name (genus + species) of an organism as its primary identifier. The other organism name fields are for a broad common name, a specific common name, and an alternate name. An example of a broad common name is "Bear" while a specific common name would be "Black bear". More levels of specificity are possible but out of practical consideration are not included. Subspecies are not considered in the database. Not every field needs to be filled out: enter "N/A" if no applicable name exists for a given column. There is also a series of fields for designating the taxonomic classification of the organism. Fill these columns with a one or zero depending on whether the organism falls into the accepted taxonomic category or not.

Assay Environment	Assay Environment Type	Assay Environment Name	Main Environment Types	Other Environment Types	Assay Types	Expression
B16/F10 cells	Cell	B16/F10	Cell	Reconstituted proteoliposomes	ATPase	EE
Brain capillary endothelial cells	Tissue	Brain capillary endothelial cells	Tissue	Purified, detergent-solubilized pr	Cell-viability	RE
Caco-2 cells	Cell	Caco-2	Membrane	Micelles	Ligand binding	RC
CEM cells	Cell	CEM	Vesicles	Nanodiscs	Monolayer	KO
CH-1 cells	Cell	CH-1	In-vivo		Reporter efflux	EE, UR
CHO cells	Cell	CHO			Reporter uptake	RE, UR
COS-7 cells	Cell	COS-7			Tissue distribution	KO, UR
Daphnia magna (in vivo)	In-vivo	Daphnia magna			Toxic endpoint	RE, KO
DU145 cells	Cell	DU145				RC, PF

Figure 2-12: A screenshot of the Assay Metadata spreadsheet. Data fields colored yellow catalog assay environments from the primary data and their characteristics. Data fields colored blue list allowable values in the assay environment Type and Name fields. Data fields colored light orange list allowable values for the Assay Types and Expression data fields in the primary data table.

Species Identification								
Scientific Name	**Common name - broad**	**Common name - specific**	**Alternate name**	**Mammals**	**Birds**	**Reptiles**	**Amphibians**	**Fishes**
Chlorocebus aethiops	Monkey	African green monkey	Grivet	1	0	0	0	0
Mus musculus	Mouse	House mouse	N/A	1	0	0	0	0
Homo sapiens	Human	Human	N/A	1	0	0	0	0
Sus scrofa	Pig	Domestic pig	Boar	1	0	0	0	0
Oncorhynchus mykiss	Trout	Rainbow trout	Steelhead trout	0	0	0	0	1

Figure 2-13: A screenshot of the Organism Metadata spreadsheet. Five entries are displayed. The Scientific Name field is the primary identifier for organisms in UC Transportal. Taxonomic classification fields not shown to the right include Vertebrates, Invertebrates, Insects, Mollusks, Plants, Fungi, Bacteria, and Protozoa.

Protein name	Alternate name 1	Alternate name 2	Alternate name 3	Superfamily	Organism	NCBI Gene ID	Full Common name
ABCB1	MDR1	P-gp		ABC	*Homo Sapiens*	5243	P-glycoprotein
ABCB11	BSEP	SPGP		ABC	*Homo Sapiens*	8647	Bile salt export pump
ABCC1	MRP1			ABC	*Homo Sapiens*	4363	
ABCC2	MRP2			ABC	*Homo Sapiens*	1244	
ABCC3	MRP3	cMOAT2	MLP2	ABC	*Homo Sapiens*	8714	
ABCC4	MRP4			ABC	*Homo Sapiens*	10257	
ABCC5	MRP5			ABC	*Homo Sapiens*	10057	
ABCG2	BCRP	BCRP1	MXR1	ABC	*Homo Sapiens*	9429	Breast cancer resistance protein

Figure 2-14: A screenshot of the Protein Metadata spreadsheet. Eight entries are displayed.

Protein Metadata

The Protein Metadata contains the most important metadata in UC Transportal. Identifying each transporter protein correctly is extremely important. Each transporter is identified by both the Protein Name and Organism data fields (Figure 2-14) because homologous proteins from different organisms are considered distinct. Currently, UC Transportal does not distinguish between different isoforms of transporters within the same species. NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene) is consulted whenever a new transporter is entered into the database or whenever unfamiliar nomenclature for an existing protein is encountered. If there is a common name for the protein, like "P-glycoprotein", it should be included in the "Full Common Name" field. Many proteins do not have a common name.

The NCBI Gene database contains many transporter proteins, but certainly not all, especially when departing from human and model mammalian species. In such cases it may be challenging to find authoritative references to the transporter protein nomenclature outside of a few primary research articles. If there is truly no standardized nomenclature system that the protein belongs to and no database affirming the protein name, it is entered as seen in the article. However, alternative genomics databases for different species do exist and may contain nomenclature information on obscure organisms and proteins. A working list of such databases can be found in the *"Primary Parameter References"* document in the UC Transportal project folders (Appendix 2, section 3).

Chemical Metadata

The Chemical Metadata spreadsheet is the metadata table with the most entries (Figure 2-11). The primary identifier of a chemical should be as specific as necessary to distinguish it from any other chemical while the alternate name fields can possess less specific names. Identifying the

best names to use is not always straightforward because of the great amount of variation in mixtures, isomers, and salts that chemicals sharing the same common name possess. The following examples and rules provide guidance about how to select the best primary identifier for chemical names and what to include as alternative names:

- Primary identifiers should be "common" names and not technical designations like IUPAC nomenclature or database codes (unless the only name available is a technical name). A good rule of thumb is they should be easily verbalized as an ordinary word.
- When a chemical isomer is specified, the isomer must be indicated in the primary identifier. For example, there are entries for both "(-)-verapamil" and "(+)-verapamil". However, compounds with multiple isomers are commonly used as racemic mixtures (50:50) and are often assumed to be so when isomerism is not specified. Therefore, "Verapamil" is also an entry, and is much more common than one isomer. A comment can be made in the Notes field of the chemical metadata that when unspecified, a racemic mixture is to be assumed.
    - In less common cases, chemical isomerism is known significant in interactions with transporters (e.g. DDT, DDE, methoxychlor), but an author has not specified the isomer, mixture, or formulation used. In such a case, the term "(nonspecific)" is added to the chemical to distinguish it. (ex. "DDT (nonspecific)").
- In some cases, a compound may technically have multiple isomers, but practically speaking, the general name only refers to one isomer unless explicitly stated otherwise. An example is glutathione, which practically only refers to "L-glutathione" and never "R-glutathione". In such cases, "Glutathione" should be the primary identifier, and a note saying "Assumed to be L-glutathione" is added in the Notes field.

    - This is similar to cases in which compounds can be assumed to be a racemic mixture when unspecified (e.g. verapamil).
- Many chemicals are formulated as organic salts which dissociate, for example, sodium methotrexate. However, the sodium cation is generally not of interest in transporter-chemical interactions, so quite often articles will not specify the salt. In such instances, "Methotrexate" can be the primary identifier. When articles specify the salt, it should be added to the metadata as a distinct chemical (with "Sodium methotrexate" as the primary identifier, for example), but when not, the generic compound (e.g. "Methotrexate") must be used.
- The primary identifier for acids in the database is always the acid and not the conjugate base (for example, use "Succinic acid" instead of "Succinate"). The conjugate base is included as an alternate name.
- When the primary identifier is very specific, the alternate names should include more generic forms of that name.
    - For example, if the primary identifier is "(+)-camphor", then the alternative names need to include "Camphor" even if "Camphor" is the primary name of another metadata entry (for example, if a study used racemic camphor).
    - If the primary identifier is "Sodium methotrexate", the alternative names need to include "Methotrexate".
- Alternative names should also be easily verbalized and not technical strings of characters (unless no other alternatives exist).
- Alternate spellings of the primary name are acceptable, but different names are preferred.

- Commercial or brand names are acceptable, but different names and alternate spellings are preferred.
- Very simple IUPAC names (e.g., 2-chlorobutanol) are acceptable, but complicated names (e.g., 1,4,5,6,7,7-hexachloro-2,2-bis(chloromethyl)-3-methylidenebicycloheptane) should not be added unless no other alternative names are available.
- Each chemical metadata entry should have at least one alternative name.

Due to the quantity, diversity and technicality of the chemical metadata, the technical conventions set forth for the primary data table also apply to the chemical metadata. All entries are required to include a PubChem identifier when possible. If a chemical cannot be identified in PubChem, a SMILES code should be added in its place. Values in the Abbreviation field should only be included if they were seen an article or they are well-known. The CAS# field in the primary data table is linked to the metadata and automatically populates based on the corresponding value in the Chemical field.

Validation Sheet

After the secondary data tables have been updated, curators must make sure the newly entered values in the primary data table match the metadata. Use the Validation Sheet to check fields with values that should be present in the metadata. The Validation Sheet checks whether the corresponding row in the primary data table matches *any* value in the appropriate metadata field and throws an error if it does not. If an error is displayed, either the metadata was not updated with that value or the value in the primary data table is incorrect. The Transporter Protein field for the MXR-activity sheet is unique in that it always *should* throw an error. If an error is not shown in a given cell, it does not necessarily mean that the data is correct, however, it is a very useful tool for

catching simple errors. Also make one last check in the primary data table for spelling errors among fields not checked by the Validation Sheet.

1	Transporter-specific Data	Proteins	Chemicals	Organisms	Assay Environment	Assay Type	Expression	DOI
2668		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
2669		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
2670		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
2671		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
2672		TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
2673		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
2674		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Figure 2-15: A small screenshot of the validation spreadsheet. The red cells indicate there is no match between corresponding cells in the primary data table and any values in the corresponding metadata primary identification field.

3.6. Uploading Data to the Website and Creating Backups

3.6.1. Selecting Data to Upload

Once data from one or more research articles has been fully curated, it needs to be uploaded to the UC Transportal website and subsequently approved by a database administrator. Uploading data does not need to be done every time an article is curated; data uploads should be performed periodically as needed. There is no limit to how much data can be uploaded at one time; however, the website and Excel sheet backups should be updated at minimum once every three months.

When submitting curated data to the UC Transportal website, it must go through the same submission and review process as data submitted from outside contributors as a form of quality control. Data submission can be done by any registered database user, however only administrators can complete the review process. The review and approval process is described in detail in section 4, "*Uploading Data to UC Transportal*". When determining which data needs to be uploaded, refer to the Articles Metadata sheet and look at the Submission Date field. Upload any data associated with articles for which this field is blank and be sure to submit *all* data associated with a particular article together. Once the data for an article has been submitted, enter the date in the Submission Date field.

To create a backup of the offline database, first make sure that no articles listed in metadata are currently in the process of being curated by checking the Curation Date field in the Articles Metadata spreadsheet. Also check the Validation Sheet and fix any errors in the primary data or metadata tables before creating a backup. Make a copy of the Kinetic Transportal Working Curation Sheet document, move it to the Master Version Archive folder in the Nicklisch Lab file system and rename it "*Kinetic Transportal Data Archive – [YYYY.MM.DD]*".

3.6.2. Backing Up Data from Outside Contributors

Database administrators have the responsibility of reviewing all data submitted to the online database whether submitted by a trained curator of the UC Transportal administration team or by outside contributors wishing to add data. Such data, although submitted through the website, must also be added to the offline working curation sheet, and eventually backed up. When the data is ready to be uploaded, indicating final approval (see section 4.2), the administrator should copy the data into the appropriate primary data sheet (either Transporter-specific Data or MXR-activity Data). Since outside contributors do not contribute to the metadata, any updates to the metadata tables should occur during this review as well.

4.Uploading Data to UC Transportal

4.1. Introduction

The ability to regularly upload new data to UC Transportal is an essential feature for it to stay relevant and well-maintained. A key goal of UC Transportal, facilitating collaboration and input from researchers from across the field, also requires a mechanism for uploading and sharing data. Currently, the upload functions are tailored to Kinetic Transportal because it is the first and only database in UC Transportal. Data can be submitted for incorporation into the database by any registered member. Nonregistered users can search the database, but the ability to contribute data

is reserved for registered users to protect the reliability of incoming data. All data goes through a series of quality assurance steps by an administrator before “going live”.

4.2. Uploading Primary Data

Data submission is accessed through the Contribute Data page where contributors have two options: Manual Entry and CSV Import (Figure 2-16). The Manual Entry option provides a web-based form for contributors to fill in cell-by-cell, pertaining to a single transporter-chemical interaction. Manual entry is a useful option when a contributor has just one or two interactions that they want to add without having to download or upload any forms. When a contributor would like to upload a larger amount of information to the database, the CSV Import option is needed. For this option, the “Transportal_data_upload.xlsx” file should be downloaded and used as a template for entering data. For security purposes, a maximum of 1000 entries can be uploaded at once. Sample data is also available for new contributors to follow as a model. The CSV Import function makes it possible for contributors to submit organized sets of data without having to enter each value individually.

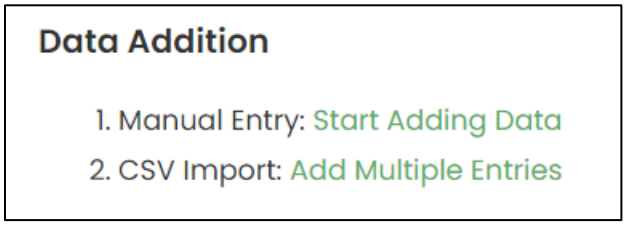


Figure 2-16: A screenshot of the option on the Contribute Data page to choose the Manual Entry or CSV Import upload methods.

Upon submission to UC Transportal, the data begins a review process which is carried out by an administrator. A flowchart of the entire process is shown in Figure 2-17 All registered users can view the My Entries page, which displays all entries they have submitted and their statuses but does not allow users to edit their entries. Administrators have access to the New Entries page, which contains all new data submitted to UC Transportal (currently only Kinetic Transportal), and

which allows administrators to make minor edits to the data. The New Entries table can be searched by administrators and filtered using the webform displayed in Figure 2-18.

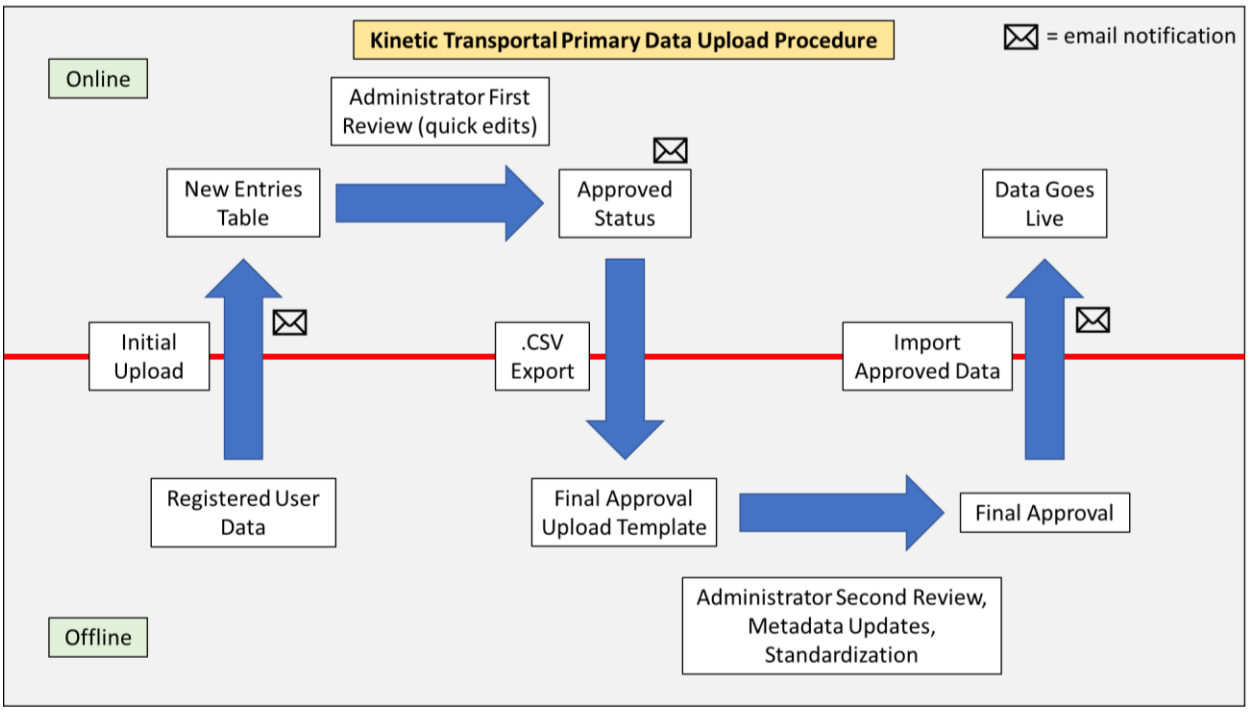


Figure 2-17: A flow chart of the uploading procedure for primary data to the Kinetic Transportal database. Steps above the red line occur online, steps below occur offline, and steps in the middle are uploading/downloading steps.

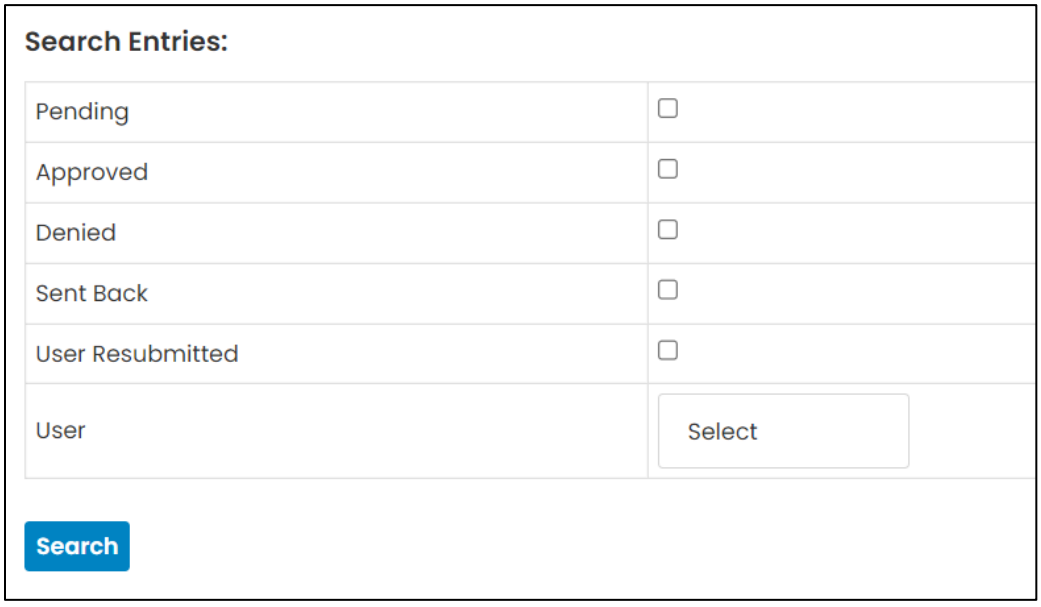


Figure 2-18: A screenshot of the webform for filtering the New Entries table into which all new entries which have just been submitted by registered users are deposited.

All records in the New Entries table have one of five statuses: Pending, Approved (not live), Denied, Sent Back, or User Resubmitted. Upon arrival, all entries receive the Pending status which they retain until an administrator acts on it. In this first stage of review, the administrator gives each entry a quick first assessment and verifies its connection to a valid reference article which should be provided by the user through the PubMed ID or DOI fields. The administrator can make quick edits or change the status of an entry to Approved, Denied, or Sent Back. If approved, the entry goes onto the second stage of review for formatting and standardization, double checking for errors, and entry into the official database. If denied, the entry will be held for 14 days and then deleted. If sent back, the contributor will receive an email requesting that they review and alter the selected entries. If the would-be contributor resubmits an entry, it is given the User Resubmitted status and treated the same way as a Pending entry. However, if a sent back entry sits idle for 30 days it well be treated as a Denied entry and eventually deleted.

Approved entries can proceed toward the second stage of review, which is completed with Microsoft Excel. Administrators can download entries with the "Approved" status to a .csv file, filtering by user if desired (Figure 2-19). This download contains two fields not present in the Kinetic Transportal curation sheets: an ID field and the MXR/MDR field. The MXR/MDR field allows the transporter-specific and MXR-type data tables to be merged in UC Transportal. At this stage, the administrator performs a more thorough review of the data and edits values to conform to technical conventions and nomenclature, match metadata values, fix spelling errors or other necessary edits. Once the administrator is satisfied with the state of the submitted data it can be reuploaded using the "Import Approved Data" function (Figure 2-20). Importing data through this function takes the data "live" and indicates "Final Approval" has been given; it can now be searched for using the Kinetic Transportal search form.

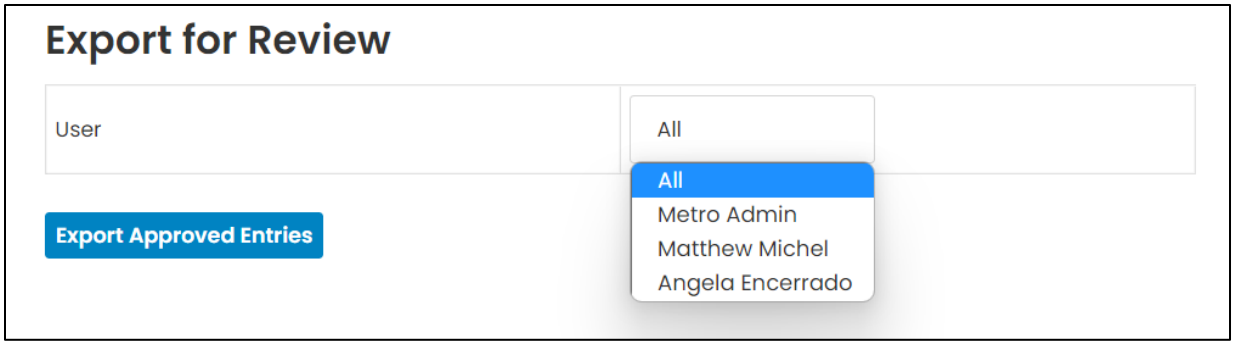


Figure 2-19: A screenshot of the "Export data to CSV" function. The drop-down only shows users with data with the "Approved" status.

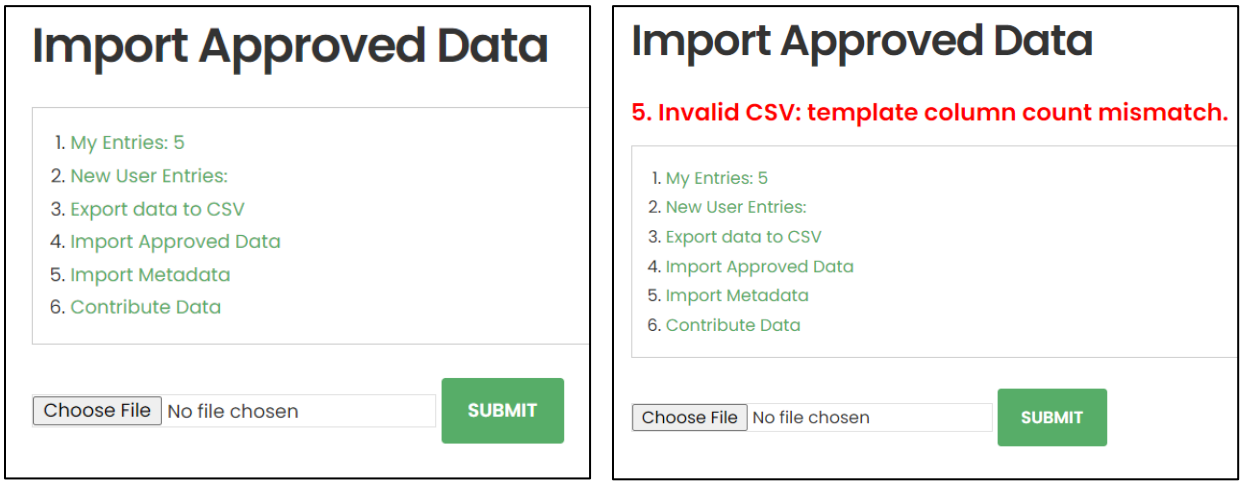


Figure 2-20: Left: Screenshot of the Import Approved Data function; Right: Screenshot of the function after throwing an error due to the file having an incorrect number of columns.

It is possible for an administrator to circumvent the UC Transportal submission system by making a replicate .csv file with the correct number of columns and directly uploading the data. However, the upload process is deliberately designed to require the quality checking steps described above and this *should not be done*. Circumventing the system would require making up a new ID number or overriding an existing one, which poses a risk for data integrity. This would have to be done deliberately and would be against protocol, since the upload for final approval contains one more data field than are contained in the curation spreadsheets and it only accepts .csv files. If a file with an incorrect number of fields is submitted for final approval, the system will throw an error, shown in Figure 2-20.

4.3. Uploading Metadata

The metadata tables must also be updated whenever new transporter proteins, chemicals, organisms, or articles are added to the system. To accomplish this, first select the appropriate metadata upload template from the Import Metadata page and fill in the data you wish to upload. Leave the ID field blank when uploading new metadata (unlike when uploading primary data for final approval, where the ID field will already be filled in); the system will automatically assign a new ID to each metadata entry. Then, upload the template to the appropriate metadata table using the Metadata Import function (Figures 2-21, 2-22). Like the final upload for primary data, the metadata upload will throw an error if the .csv file has an incorrect number of columns.

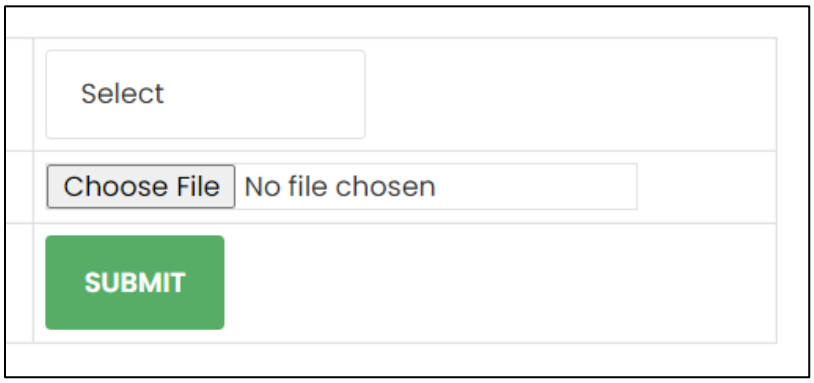


Figure 2-21: A close-up of the Metadata Upload function without the selection drop-down menu displayed

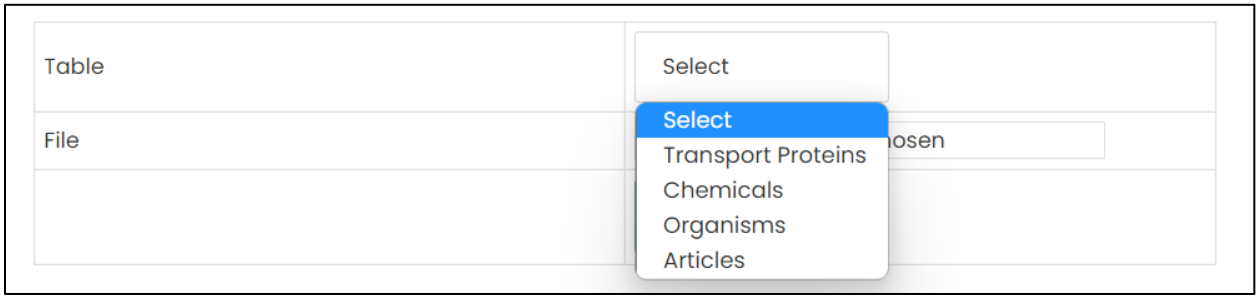


Figure 2-22: The Metadata Upload function with the metadata table selection drop-down menu displayed.

4.4. Updating Live Entries

To update entries that have already gone live in the database, whether belonging to the primary data table or any of the secondary data tables, simply reupload the data using the proper template .csv file and the ID value of the entry. The existing data in the database will be overridden.

Therefore, it is extremely important enter the correct ID. Otherwise, the intended entry will not be updated, and another entry will be overridden instead.

5. Initial Data Collection and Summary

5.1. Data Collection

UC Transportal is meant to be maintained on a routine basis, with new data uploaded in a timely manner after being curated, or at minimum every three months. Therefore, UC Transportal will not undergo major version updates on account of new data collections, as was the case with the recent update to the UCSF-FDA TransPortal (http://transportaldev.docking.org). Version updates will be implemented when adding new features to UC Transportal, however.

The initial pool of data used populate the Kinetic Transportal database and test aspects of its design was obtained through an extensive literature search focusing on interactions of environmental chemicals, drugs, phytochemicals, and other classes of chemicals with active transporter proteins. PubMed (https://pubmed.ncbi.nlm.nih.gov), Google Scholar (https://scholar.google.com) and Web of Science (https://www.webofscience.com) were searched using combinations of keywords to find data on direct transporter-chemical interactions. Keywords such as “environmental chemical”, “pesticide”, “inhibition”, “substrate” and “IC_{50}”, were combined with protein nomenclature terms, such as “ABC transporter”, “SLC transporter”, “ABCB1”, SLC22A1 or others. Occasionally organism terms such as “fish”, “sea urchin”, “murine”, or “human” were also used to narrow the search. At times, non-mammalian organisms and less commonly studied transporters were deliberately sought out to demonstrate the inclusivity of UC Transportal toward the entire range of active transporters across species.

Search results were given a brief examination and those containing direct transporter-chemical interaction data were archived for curation. Articles containing data on chemical-gene

expression interaction, transporter localization, clinical drug-drug interaction or *in silico* data instead of information on direct transporter-chemical interactions were archived separately for use in future expansions of UC Transportal. A set of 100 primary research articles collected by Travis Fleming (Hamdoun lab, Scripps Institution of Oceanography, UCSD) was also included in the initial pool of data used to populate the database.

5.2. Brief Data Summary

As of June 2022, the Kinetic Transportal contains 2039 transporter-specific interactions and 162 MXR/MDR-activity interactions from 99 primary research articles published between 1992 and 2021. There are 52 different transporters, 664 chemicals, and 24 organisms represented. Among the transporter-specific interactions, a majority (n = 1686) are with human transporters. We expect this to change as the database continues to expand and encompasses data pertaining to transporter-chemical interactions in many organisms. We also expect the proportion of MXR/MDR-activity interactions to grow because the in the initial stages of development *in vivo* studies were excluded and began to be added once development was further underway. *In vivo* studies produce MXR/MDR-activity interactions at a far higher rate than *in vitro* studies because of the practical difficulty in pinpointing a specific protein as the cause of an observed effect. There are currently 58 different assay environments in the database, a number that also expected to rapidly increase with the addition of more *in vivo* studies.

6. Outlook and Future Development

The outlook for UC Transportal is to continue developing as a resource for transporter protein research, updating to include the latest data and adding new capabilities well into the future. The design rationale for both the website of UC Transportal and the initial Kinetic Transportal database included the requirements of expandability without sacrificing user-friendliness,

inclusivity for additional data categories, and structuring data with future analyses in mind. A small, static database would not need to consider these design elements, however, the vision for UC Transportal is to grow beyond its initial scope and become a comprehensive data resource for transporter protein research.

The Kinetic Transportal, which focuses on direct transporter-chemical interactions, is only the first of multiple connected databases planned for the UC Transportal hub. We plan to add databases cataloguing data pertaining to interactions between chemicals and transporter gene expression, pharmacokinetic data for clinical drug-drug interactions, and transporter localization data as well. These datasets will be treated as entities distinct from the existing Kinetic Transportal but will be closely associated using UC Transportal as a hub and will share applicable metadata resources. The addition of crystal structure information (i.e., PDB database, https://www.rcsb.org) to the protein metadata for any transporter proteins for which high-quality structural data is available is also part of the vision for additional datasets. It should be noted that each of the major additions will require a new design suitable for displaying their distinct sets of information clearly and accurately, however all databases within the UC Transportal hub will be prepared with the key design elements listed in the design rationale in mind.

The existing Kinetic Transportal will need further expansion through a large curation effort. One ready source of primary research articles to curate data from is the recently updated UCSF-FDA Transportal, containing 592 articles confirmed to have transporter-chemical interaction data. The Nicklisch Lab has also compiled a large repository of primary research articles ready for curation. The current dataset has already demonstrated the ability of Kinetic Transportal to effectively describe data for transporters from across species, however, human transporter interactions with pharmaceuticals still make up a large portion of the current data. To

showcase UC Transportal as an all-inclusive transporter database, examples of non-mammalian proteins should be deliberately included in near-term curation efforts, including from fish, mussels, bacteria, and model plant species. This may accelerate our ability attract transporter researchers from different backgrounds to create a diverse transporter researcher community anchored by UC Transportal as a common collaborative resource.

Although the outlook for UC Transportal is principally to become a comprehensive data resource, the broader vision behind it is to facilitate collaboration and move to reframe transporter studies from dissimilar topics as all a part of one unified field. This is rooted in the fact that active transporters tend to be highly conserved, perform a fundamental biological function, and similar methods can be used to study transporter proteins involved in dissimilar species or which transport dissimilar compounds (Tirona et al. 2003; Palmeira et al. 2011; Nicklisch et al. 2021). To this end, a UC Transportal Forum for communication between all interested parties would be a highly beneficial addition to UC Transportal. An active forum anchored to the UC Transportal system would not only provide a dedicated platform for pan-transporter protein discussion, but also ensure utilization of the database system as research progresses. While the current Kinetic Transportal database and framework for UC Transportal provide a method of collaboration and distribution of data through the database, a forum is a more familiar channel of communication capable of providing greater visibility to users and gathering ideas.

Additional features to improve user experience and aesthetics and to enhance data analysis will also be made in the future. Class-based searches for the primary parameter fields in Kinetic Transportal will be enabled by expanding the scope of chemical and organism metadata on the website and modifying the current search form. Proteins will be able to be searched by superfamily, chemicals by selected uses and other classifications, and organisms by select taxonomic categories.

The assay metadata (see section 3.5) is currently not accessible to users but may also be made available in the future. Adjustments to the data tables, background images and other minor design features will also likely be made going forward to enhance user experience and improve features.

Lastly, we intend to add a network analysis tool to visualize existing and predict new transporter-chemical interactions, their type (inhibitor, stimulator) and potency. Users will be able to upload exposome datasets or focus on specific transporter proteins and discover potential drug-drug, drug-environmental chemical, drug-food, and other interactions based on their individual interactions with transporter proteins in the Kinetic Transportal database. The network analysis tool will be interactive, use color-coding to display the interaction types and parameters, and its user-interface will use the network analysis tool for the protein-protein interaction database hosted by string-db.org (STRING Consortium 2022) as a creative starting point. This tool will make suggestions predicting increased, decreased, or unchanged transport activity on selected substrates due to exposure to chemicals from user-uploaded exposome datasets with potential uses for clinical, regulatory, and academic researchers.

Chapter 3: Vacuolar Monolignol Transport – Developing a Protocol for Efficient Vacuole Isolation in *Arabidopsis thaliana* for Transport Studies

1. Introduction

Active transmembrane transport proteins are ubiquitous across organisms in all domains of life (Xiong et al. 2015). They perform the essential task of regulating the movement of substances across cell and organelle membranes regardless of the presence of a chemical gradient of the substrate. Due to the particularly high number of active transporters found in terrestrial plants compared to other kingdoms of life, it has been suggested that they play a uniquely important role in terrestrial plant biology (Hwang et al. 2016). Indeed, plant physiology would suggest a need for such machinery. Terrestrial plants are sessile, meaning they cannot move to a more favorable environment and must make the best of their immediate surroundings, and they are non-aquatic, which necessitates an exterior designed to protect them from a harsh, dry environment and which greatly limits the exchange of compounds through fluids. Hence, there is a need for efficient and substance-specific mechanisms of material exchange with the surrounding environment. Additionally, terrestrial plants are autotrophs, synthesizing most of their own components rather than consuming them, and are known for containing an incredible variety of unique secondary metabolites in addition to the four standard classes of organic macromolecule (proteins, lipids, carbohydrates, and nucleic acids). They are also highly compartmentalized, and having inherited a circulatory system that is generally more specific in function than that of animals (Susann and Biddulph 1959), the need for abundant molecular machinery to regulate the distribution and partitioning of substances within terrestrial plants is apparent (Do et al. 2021). Active transporters appear to have filled this role to such a degree that the model plant *Arabidopsis*

thaliana contains over 2.5 times as many ABC (ATP-Binding Cassette) transporters and over 1.5 times as many secondary active transporters as humans (Hwang et al. 2016).

Within the plant cell, the storage vacuole stands out as a compartment whose functions are directly tied to partitioning and sequestering substances from the rest of the cell. The vacuole has long been associated with regulating turgor and solute concentrations but is now known to accumulate a myriad of endogenous substances including energy stores like sucrose, dyes such as anthocyanins (Shitan and Yazaki 2020) and flavonoids (Martinoia 2018), defensive compounds such as nicotine (Shitan and Yazaki 2020) and cyanogenic glycosides (Etxeberria et al. 2012), nutrients like phytic acid (Nagy et al. 2009), and countless others. It also has been known to sequester toxic metals such as cadmium (Gao et al. 2017) and arsenic (Maciaszczyk-Dziubinska et al. 2012). Researchers have identified and characterized a large number of active transporters embedded in the vacuolar membrane, called the tonoplast, which carry out these partitioning and sequestering functions(Carter et al. 2004; Shimaoka et al. 2004; Jaquinod et al. 2007). However, identification and characterization of vacuolar transporters is an ongoing challenge that is far from complete.

Among the many compounds known to accumulate in the plant storage vacuole are glycosylated monolignols. Monolignols are the monomeric precursor to lignin, a complex heteropolymer that is crucial for the development of vascular plants. Although most prominent in woody tissues, lignin is an important structural element of the secondary cell wall throughout both woody and nonwoody tissues (Barros et al. 2015). In addition to providing structural support, lignin plays roles in water retention, pest resistance, salt tolerance and climate adaptation(Bhuiyan et al. 2009; Liu et al. 2018) among others. As a testament to its importance, lignin is the second most abundant organic polymer on Earth, after only cellulose (Barros et al. 2015; Liu et al. 2018).

The complexity of lignin is derived from the varying composition of its monomeric constituents, the monolignols, of which there are three main types: coniferyl alcohol (G units), sinapyl alcohol (S units), and para-coumaryl alcohol (H units) (Guragain et al. 2015; Lourenço and Pereira 2018) (Figure 3-1). These units are synthesized within the cytosol of the plant cell but must make their way to the outside of the cell to be incorporated into the polymeric lignin mesh which makes up a significant portion of the secondary cell wall. To date, researchers have managed to identify a single ABC-type transporter which transports para-coumaryl alcohol from the cytosol to the cell exterior in *Arabidopsis thaliana* (Alejandro et al. 2012) but have not made further definitive progress in uncovering the mechanisms by which monolignols are transported to the locality of their polymerization (Perkins et al. 2019).

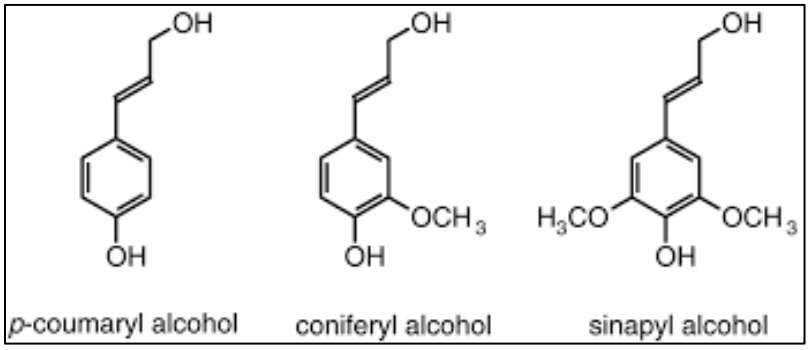


Figure 3-1: Structures of the three main monolignols.

At high concentrations, monolignols in the cytoplasm can have cytotoxic effects and are glycosylated by UDP-glucosyltransferase (UGT) enzymes to protect the cell (Le Roy et al. 2016; Perkins et al. 2019). Glycosylated monolignols do not appear to be transported across the plasma membrane of *Arabidopsis thaliana* rosette leaf cells by any transporter proteins (Miao and Liu 2010), nor is passive diffusion a likely method of crossing the membrane for a polar glycosylated compound. Glycosylated monolignols and related polyphenolic oligomers have, however, been detected inside the plant vacuole (Dima et al. 2015), and are thought to be transported there for sequestration by an active transporter embedded in the tonoplast (Miao and Liu 2010; Tsuyama et

al. 2019; Väisänen et al. 2020). It has been proposed that transport into the vacuole may be an intermediate step on the way to the outside of the cell or for storage as a defense mechanism against consumers and pathogens. The mechanism of transport may also differ across species, and it is unknown how these compounds exit the vacuole once inside (Liu 2012).

Regulating lignin content in agricultural and silvicultural commodities is of interest in several potential applications. Lignin has not historically been a major target for plant breeders, however in forage crops reduced lignin is desirable because lignin is considered an antinutrient for its negative effect on digestibility. Low lignin transgenic strains of alfalfa and maize have been developed with the goal of improving digestibility, and natural varieties of several other forage crops containing low lignin have been identified (Frei 2013). Lignin is also highly significant to plant-based biofuel production both because of its high energy content and resistance to degradation, which poses a challenge to its effective utilization. Currently, extensive chemical or enzymatic pretreatments are required to extract useful end products which potentially include valuable aromatic compounds (such as xylene and benzene) and alcohols that are currently primarily derived from petroleum (Frei 2013; Zeng et al. 2014; Guragain et al. 2015; Welker et al. 2015) in addition to biofuels. As previously mentioned, lignin plays a role in a number of plant responses to both abiotic and biotic stress factors, which almost always induce lignin production (Frei 2013; Liu et al. 2018). Regulating lignin content could open ways of tailoring crops to specific environments with greater resistance to stressors, greater yield or a more desirable material composition.

Researchers have already discerned the entirety of the monolignol synthesis pathway beginning with the amino acid phenylalanine, and several factors regulating it (Barros et al. 2015). Progress has also been made in understanding how monolignols polymerize at their site of

incorporation into the cell wall by way of laccase and peroxidase enzymes (Tobimatsu and Schuetz 2019). The intermediate step of transport from the cytosol to the outside the cell, however, continues to represent a wide gap in our knowledge about how lignin is formed (Sibout and Höfte 2012; Perkins et al. 2019). For this reason, it is still necessary to explore multiple possible routes that monolignols and their glycosylated derivatives may take and where they end up in different contexts.

Our initial project toward this end is to isolate whole, intact vacuoles from the model plant species *Arabidopsis thaliana* to examine the transport of monolignols, their direct derivatives, and structurally related compounds into and out of this vital cellular compartment and to measure the activity of any transport proteins involved. Using intact vacuoles carries distinct advantages over membranes prepared from microsomal fractions of plant cells, the first and foremost of which are the greatly reduced chances of damaging or altering the composition of the tonoplast and ensuring that assays are performed using the native membrane environment, which has a significant effect on transporter activity (Shukla et al. 2017). With intact vacuoles, accumulation assays and leakage assays are an option, and the native contents of the vacuole can also be analyzed. In addition, the experimental methods described hereafter for examining questions related to monolignol transport can be modified and applied to a large variety of compounds for which the vacuole is a relevant storage compartment.

2. Growth Conditions for *Arabidopsis thaliana*

All plants were grown in chambers at the Controlled Environment Facilities (CEF) at UC Davis. The chamber was set to a 16-hour photoperiod at 20°C and 50-60% relative humidity. The soil mix, provided by CEF, was Sun Gro Sunshine Mix #1. Plants were typically watered every other day with nutrient water containing Grow More 4-18-38 fertilizer without boron and

supplemented with calcium nitrate and magnesium sulfate. Watering was done by subirrigation of pots in 11 by 21 inch “1020 flats” except when seedlings were particularly small, in which case a spray bottle was used. The proper amount of watering was determined by feeling the weight of each pot. Before potting, dry soil was soaked in water to regain the ability to hydrate properly and was gently compacted into the pots to reduce air pockets.

Arabidopsis thaliana seeds for method development were kindly provided by Dr. Nitzan Shabek. For transport experiments, seeds of the Columbia ecotype were purchased from the Arabidopsis Biological Resource Center (ABRC) at Ohio State University (stock number CS1092). To eliminate dormancy and synchronize germination times, seeds were stored in a refrigerator at 4°C for at least one week before sowing. Due to the exceptionally small size of Arabidopsis seeds, they were sown by the following method: a small amount of seed was sprinkled onto a petri dish and gently sprayed with water until they floated in large water droplets. A 200µL pipette with the last .25 inches of the pipette tip cut off to increase bore size was used to visually aspirate seeds one at a time. Seeds were deposited on top of moist soil and not covered. A clear humidity dome was used to retain soil and air moisture until the cotyledons were visible, after which the dome was lifted slightly until plants reached the four-leaf stage and then domes were removed (Figure 3-2). In the case that multiple seeds were deposited close together, transplanting could be performed with tweezers from between 7 to 14 days after sowing with a near 100% survival rate. We used pots that were 4 inches square and 3.5 inches tall, and each pot had one to four plants.

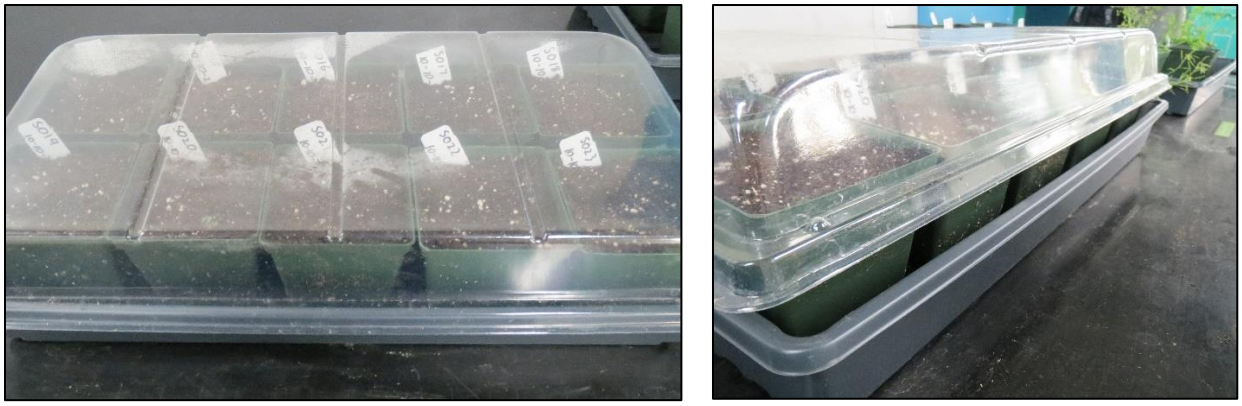


Figure 3-2: A humidity dome is placed over pots with seeds recently sown. Right: the humidity dome fully covering the tray; Left: The humidity dome is lifted slightly to help the plants adjust to growth chamber environment before full removal.

The resources used to develop the growth protocol were documentation from the ABRC at Ohio State University (Arabidopsis Biological Resource Center 2002), the *101 Ways to Try to Grow Arabidopsis* project at Purdue University (Eddy and Hahn 2012), and the advice of Dr. Bo Liu in the Department of Plant Biology at UC Davis.

3. Vacuole Isolation Protocol Development

3.1. Overview and Starting Point

Obtaining purified, intact plant vacuoles is the major hurdle to carrying out our intended experiments on monolignol transport. To achieve this, we started with the methods of Carter et al. 2004, Robert et al. 2007, and Zouhar 2017, which came out of the same cohort of researchers and are very similar. The goal was to adapt this procedure to work with our set of constraints as a laboratory and to make it as efficient as possible. As development and verification of the procedure continued, significant challenges were encountered which shifted the goal from utilizing the original protocol to creating a more adaptable protocol which could reliably produce vacuoles with less material expense, more accessible equipment, and greater ease of technique.

The protocol can be divided into three main steps: 1) protoplast isolation, 2) cell lysis, and 3) vacuole purification. For a brief explanation, sliced leaf tissue is placed in an enzyme digest

solution which is incubated with gentle shaking for 4 hours at 21°C. This removes the cell wall and other components of the extracellular matrix to produce protoplasts (free plant cells lacking the cell wall). Then, the solution is filtered, then centrifuged, decanted, and washed with wash buffer, twice. Lysis buffer is then added to rupture the cells, releasing the vacuoles, and introducing neutral red dye to stain them pink. The lysis solution is then placed at the bottom of a Ficoll density gradient and centrifuged at high speed to separate the vacuoles from all other cell components and debris. The vacuoles appear as a faint red band at the bottom of the uppermost density layer of the gradient and are collected with a micropipette. See Carter et al. 2004, Robert et al. 2007, and Zouhar 2017 for the full original protocols. Figure 3-3 displays an overview of the procedure.

Alterations and refinements were made to each major step in response to challenges encountered, as well as opportunities for improvement. Initially, only minor changes were made to the original protocol to streamline the steps and cater to our lab setup. Major modifications came later to address challenges and potentially expand the usefulness of the protocol. Determining which factors were most important for successful vacuole production took a significant amount of time, and some outstanding issues were never adequately solved. However, some changes represented a significant improvement upon existing protocols, which, if built upon, may prove valuable to future researchers in need of intact vacuoles.

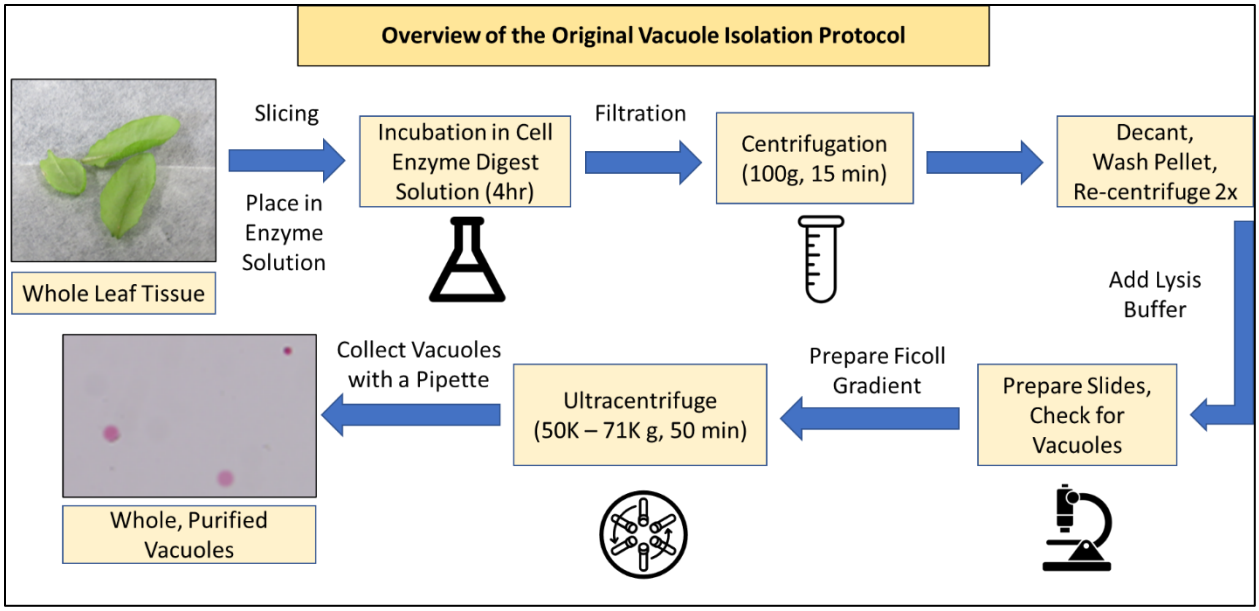


Figure 3-3: A flow chart summarizing the original vacuole isolation protocol as derived from the Carter, Robert, and Zouhar protocols (Carter et al. 2004; Robert et al. 2007; Zouhar 2017).

3.2. Minor Adjustments to Original Protocols

3.2.1. Solutions and Plant Material

The original protocol called for 2 grams of thinly sliced Arabidopsis rosette leaves to be placed in 30mL of an enzyme digest solution. The solution called for 1% (w/v) Cellulase "Onozuka" R-10, 1% (w/v) Macerozyme R-10, .4M mannitol, 10mM MES buffer at pH 5.7, 25mM calcium chloride, and 5mM beta-mercaptoethanol (Zouhar 2017). Due to the expense of the enzymes, we ran most experiments at 1/20 scale, reducing the materials needed to 100mg of tissue in 1.5mL total solution. This required an alteration in preparation techniques (see section 3.3 and Appendix 3, section 1) and caused scalability to become a development goal.

To avoid using beta-mercaptoethanol, which is volatile, toxic and must be vacuumed off, we substituted it with a safer reducing agent, tris(2-carboxyethyl)phosphine hydrochloride, (TCEP-HCl). Adding TCEP-HCl, however, overwhelmed the MES buffer and lowered the pH to below 3. Although the lowered pH did not appear detrimental to the process, experiments in which the concentration of TCEP-HCl was reduced or removed entirely showed that the addition of a

reducing agent is not critical to producing healthy protoplasts. Therefore, to achieve a more desirable pH, we reduced the amount of TCEP-HCl to just .025mM and doubled the concentration of MES buffer to 20mM, as in Carter et al. 2004. This put the pH near 4.7, within the optimal pH ranges of both macerozyme (3.5 - 7.0) and cellulase (4.0 - 5.0), as stated by the manufacturers.

The most critical component of all buffers in the vacuole isolation procedure is the mannitol concentration. Our first isolation attempts mistakenly excluded mannitol from the enzyme digest solution. The protoplasts from these early attempts were large, bloated, and fragile, and the solution turned green from loose chloroplasts in solution during the 4-hour incubation period. The lysis step failed completely in these experiments because it is driven by osmotic pressure, and any surviving protoplasts had already been incubating in a minimal osmolality environment. Enzyme solution with the proper mannitol concentration produces healthy protoplasts which sink to the bottom of the tube.

The lysis buffer consisted of .2M mannitol, 10% (w/v) Ficoll, 10mM EDTA, 5mM sodium phosphate, pH 8.0, and .001% (w/v) neutral red dye. The concentration of neutral red was doubled from Robert et al. 2007 to assist with staining and to reach a more convenient volume for pipetting (3µL); otherwise the solution was left unaltered. The basic pH of the lysis buffer is critical for the efficacy of the neutral red dye, which has a pK_a of 6.8. At basic pH, neutral red has a light-yellow color and no charge, allowing it to diffuse across the vacuolar membrane. Once inside, the pH of the vacuole, which is around 5.2 (Shen et al. 2013), causes neutral red to become positively charged with a deep pink color, trapping it inside the vacuole for an effective stain. We found that the volume of lysis buffer used was unnecessarily large and that it could be reduced by at least 1/3 without any noticeable effect. Lysis buffer volumes were therefore reduced to more convenient volumes; checking for vacuoles in more concentrated solution was also easier.

The three remaining solutions had no need for adjustments. The wash buffer consists of .4M mannitol and 10mM MES, pH 5.7. The vacuole storage buffer contains .45M mannitol, 5mM sodium phosphate, pH 7.5, and 2mM EDTA, pH 8.0. The 4% Ficoll solution contained 4% (w/v) Ficoll, 5.2mM EDTA, .35M mannitol, and 5mM sodium phosphate, pH 7.5. During later experiments on centrifugation speed and scalability, the Ficoll concentrations of the lysis buffer and vacuole storage buffer were changed and the 4% Ficoll solution was disused (see section 3.4), but all other components remained the same. Table 3-1 lists the concentrations of all stock solutions that were prepared.

Table 3-1: A list of stock solutions used for all working solutions in the vacuole isolation protocol.

Stock Solution	**Concentration**
Mannitol	1M
Sodium phosphate, pH 7.5	.2M
Sodium phosphate, pH 8.0	.2M
Calcium chloride	.1M
MES, pH 5.7	.1M
EDTA, pH 8.0	.1M
TCEP-HCl	.1M
Ficoll	20% (w/v)
Neutral red (in .2M sodium phosphate, pH 7.5)	.33% (w/v)

Table 3-2: Compositions of all working solutions in the vacuole isolation protocol.

Enzyme Digest Solution	
Component	**Concentration**
Cellulase "Onozuka" R-10	1% (w/v)
Macerozyme R-10	1% (w/v)
Mannitol	.4M
MES, pH 5.7	20mM
Calcium chloride	25mM
TCEP-HCl	.025mM

Lysis Buffer	
Component	**Concentration**
Mannitol	.2M
Ficoll	10% (w/v)
EDTA, pH 8.0	10mM
Sodium phosphate, pH 8.0	5mM
Neutral red	.001% (w/v)
Wash Buffer	
Component	**Concentration**
Mannitol	.4M
MES, pH 5.7	10mM
4% Ficoll Solution	
Component	**Concentration**
Mannitol	.35M
Ficoll	4% (w/v)
Sodium phosphate, pH 7.5	5mM
EDTA, pH 8.0	5.2mM
Vacuole Storage Solution	
Component	**Concentration**
Mannitol	.45M
Sodium phosphate, pH 7.5	5mM
EDTA, pH 8.0	2mM

Several experiments indicated that the quality of the leaf material matters significantly when preparing vacuoles. When deciding when to harvest plant material, a balance must be struck between harvesting early enough for leaf quality to be good, and late enough to have sufficient material. An ideal time is around 30 days after sowing (Yoo et al. 2007) (Figure 3-4); however, plants can vary significantly in growth rate (Appendix 3, section 4), and growing a large number of plants can solve any problems regarding obtaining enough tissue. Younger leaves that have not yet elongated with age, retaining a round appearance, produced smaller, healthier, more durable protoplasts than older leavers, and subsequently more intact vacuoles as well. Younger leaves are also softer and easier to slice into digestible strips than tougher older leaves.

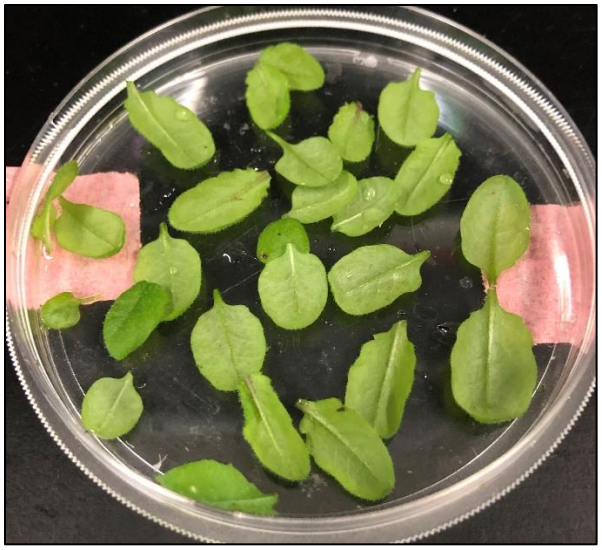


Figure 3-4: Approximately 500mg of rosette leaves from 30-day old *Arabidopsis thaliana* on a petri dish. The petri dish is 9cm in diameter.

3.2.2. Scaling and Techniques

At full scale, the original protocol calls for the enzyme digest solution to be gently shaken at 70rpm in a 250mL flask, and for 50mL tubes to hold samples in subsequent steps. To save material, most of the experiments were run at 1/20 scale, and some at 1/4 scale, which necessitated slightly different technique (Figures 3-5, 3-6). At 1/20 scale, 2mL tubes were used for the entire process until ultracentrifugation, and at 1/4 scale, 15mL tubes were used. During the incubation, the tubes were placed in a tube holder and laid sideways to allow for gentile agitation. The 2mL tubes were shaken at 70rpm as normal, but the 15mL tubes needed slower shaking at 40 to 60rpm to prevent wave action. Tilting the tubes upward at a shallow angle helped with this problem. Results did not show that shaking at 70rpm noticeably disrupted protoplasts, but it is nonetheless important to prevent too much agitation.

Figure 3-5: Enzyme digest solution with Arabidopsis rosette leaf material. Left: before the incubation period, Right: after four hours of incubation at 21°C.

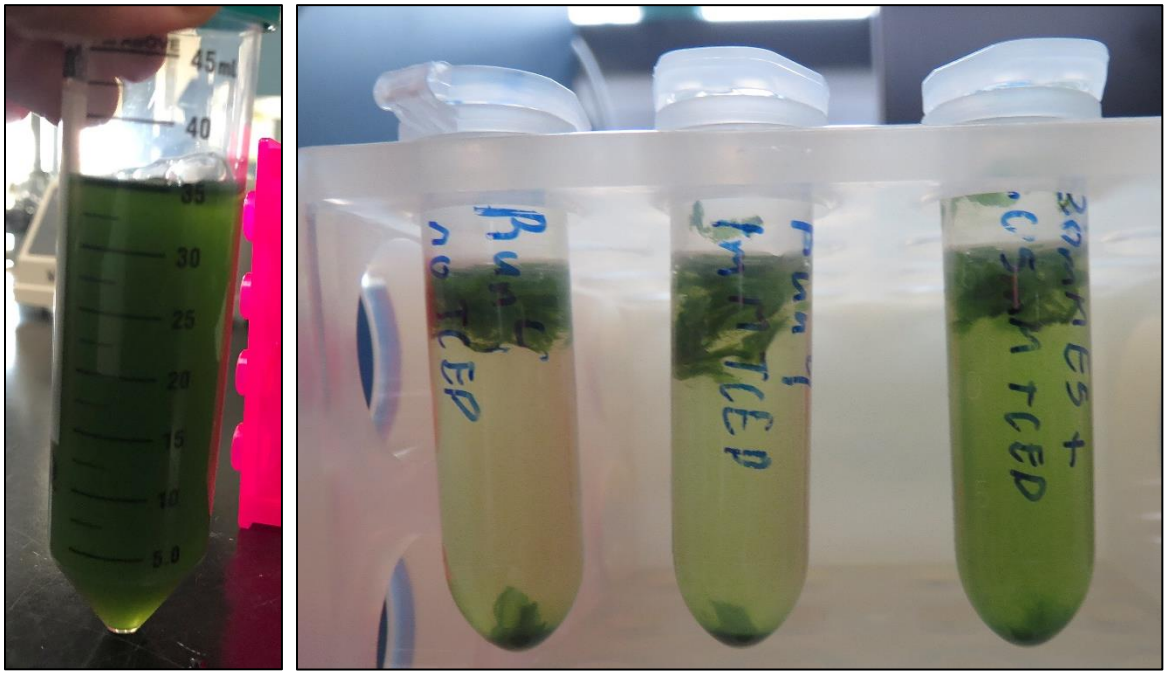


Figure 3-6: Left: A 50mL tube of enzyme digest solution from a full scale run after 4 hours of incubation and filtration; Right: Three 2mL tubes of enzyme digest solution at 1/20 scale after 3 hours of incubation but before filtration during an early experiment testing differences in TCEP-HCl concentration, which affects pH.

For all scales of experiment, we used 100µm cell filters to strain the enzyme digest solution after the incubation period. The cell filters were placed in small funnels atop the transfer tube (Figure 3-7). After transferring the material, wash buffer was used to rinse residual material from the original tube and passed through the filter as well. When using 2mL tubes, it is important to avoid accidentally overflowing the tube due to displacement caused by the tip of the funnel. Also,

when discarding liquid from the 2mL tubes after the subsequent centrifuging and washing step, it helps to aspirate the supernatant using a pipette tip instead of decanting.

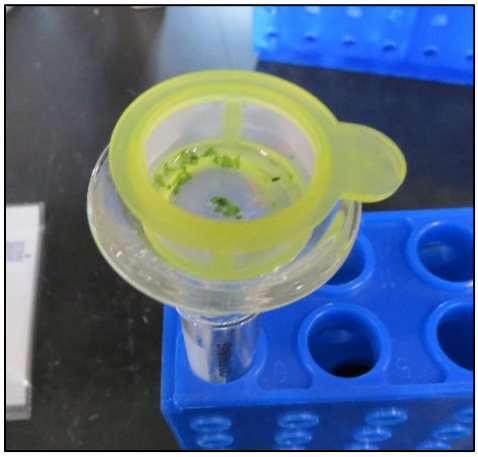


Figure 3-7: A 100µM cell strainer atop a small funnel atop a 15mL tube, with debris from digested plant material.

One of the most difficult parts of the procedure to learn was making Ficoll density gradients. The original protocol called for the 10% Ficoll Lysis buffer to be set on the bottom of an ultracentrifuge tube, with 4% Ficoll solution layered on top, and ice-cold vacuole storage buffer (0% Ficoll) layered on top of that. Once made, the density gradient is remarkably stable, however, the layering process requires patience and a very steady hand. The layers will partially mix if the liquid is introduced at with any appreciable speed, ruining the gradient. To prevent this, tilt the tube at a steep angle without spilling, and dispense the liquid slowly from just above the top of the waterline, slowly tilting the tube upward and moving the pipette upward as the liquid level increases. Using an automatic pipette can work with a 50mL tube, but a manual pipette is recommended for control. When using 26.3mL ultracentrifuge tubes, this was a very difficult process, however, with smaller volumes in 15mL and 2mL tubes, the process was much easier. It is also easier to make Ficoll gradients when the difference in concentration between consecutive layers is greater, such as a 10% to 0% Ficoll transition than a 10% to 4% transition.

All volumes were adjusted proportionately with the scale of experiment until the ultracentrifugation step. The model of ultracentrifuge available to us was a Beckman Coulter L-70 with a Ti-70 rotor and 26.3mL ultracentrifuge tubes, far larger than the equipment used in the original protocols. Due the larger size, 1/20 and 1/4 scale experiments were not viable to continue after checking the lysis buffer for vacuoles. Full scale experiments adjusted the volumes such that the lysis buffer was overlayered with 5mL of 4% Ficoll solution, and the top layer of ice-cold vacuole storage buffer filling the tubes up to the top. The lack of access to a suitable ultracentrifuge was a major contributor to the need for a more heavily modified protocol.

3.3. Observations and Challenges

Using the original protocol with only minor modifications brought us most of the way to obtaining purified, intact vacuoles. We were able to obtain healthy protoplasts with consistency and successfully released stained vacuoles in many tests at 1/20 scale. We observed several different protoplast and vacuole morphologies which we used as indicators of health after incubation in the enzyme digest solution (Figure 3-8) or incomplete lysis after adding lysis buffer. Healthy protoplasts should be round and not distended. Their vacuoles should not be colored or conspicuous and the chloroplasts should not be bunched up on one side of the cell. During many of our experiments, we observed highly inflated protoplasts with bloated vacuoles, presumably due to osmotic stress. The chloroplasts are often pushed to one side of the protoplast when this happens, which we called “crescent morphology” (Figure 3-10). Isolation attempts in which most protoplasts were overly distended tended not to produce stained vacuoles. We believe that in cases with too much osmotic stress, the entire protoplast and vacuole becomes very fragile and bursts entirely (Figure 3-9). Since lysis is achieved through osmotic pressure, there is a balance that must

be struck between lysing the protoplasts but not the vacuoles. Smaller, healthier protoplasts tend to produce smaller, more deeply stained vacuoles which are easily observed.

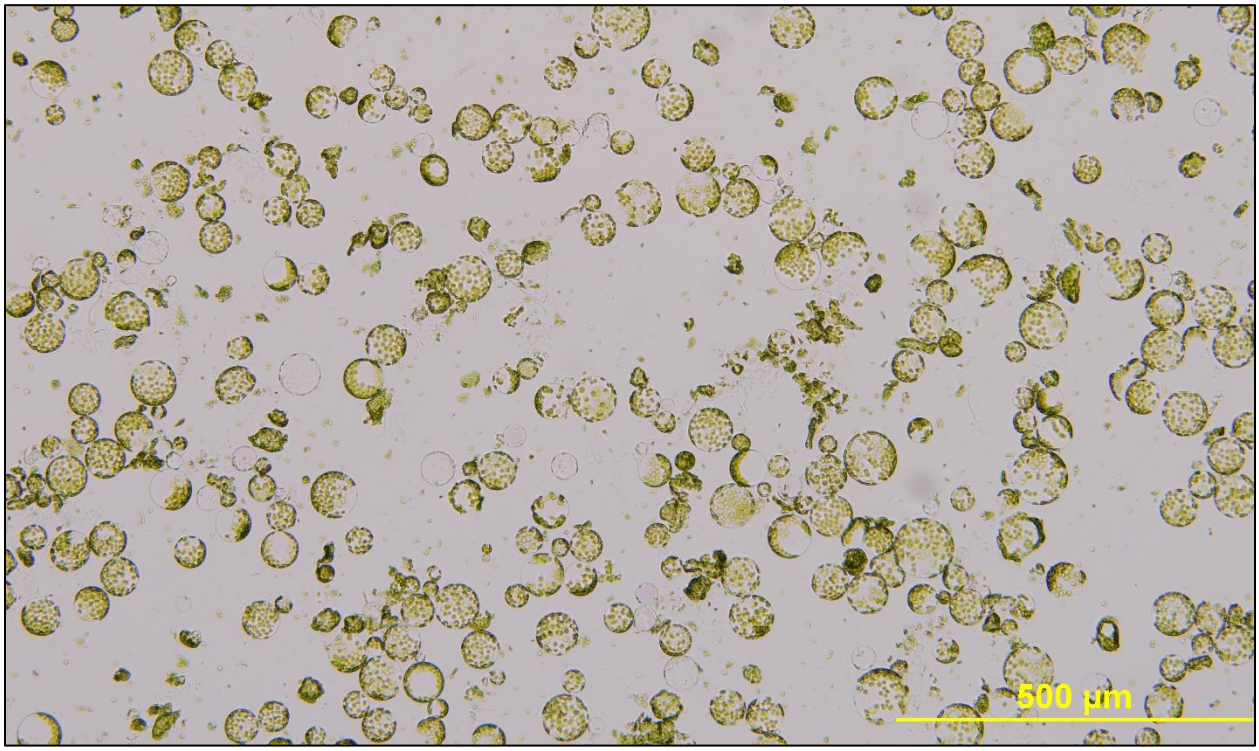


Figure 3-8: Healthy protoplasts observed after the enzyme digestion step. A few different morphologies can be observed. All microscope images were taken using an Olympus BX51 fluorescence microscope in UV-Vis mode without filters.

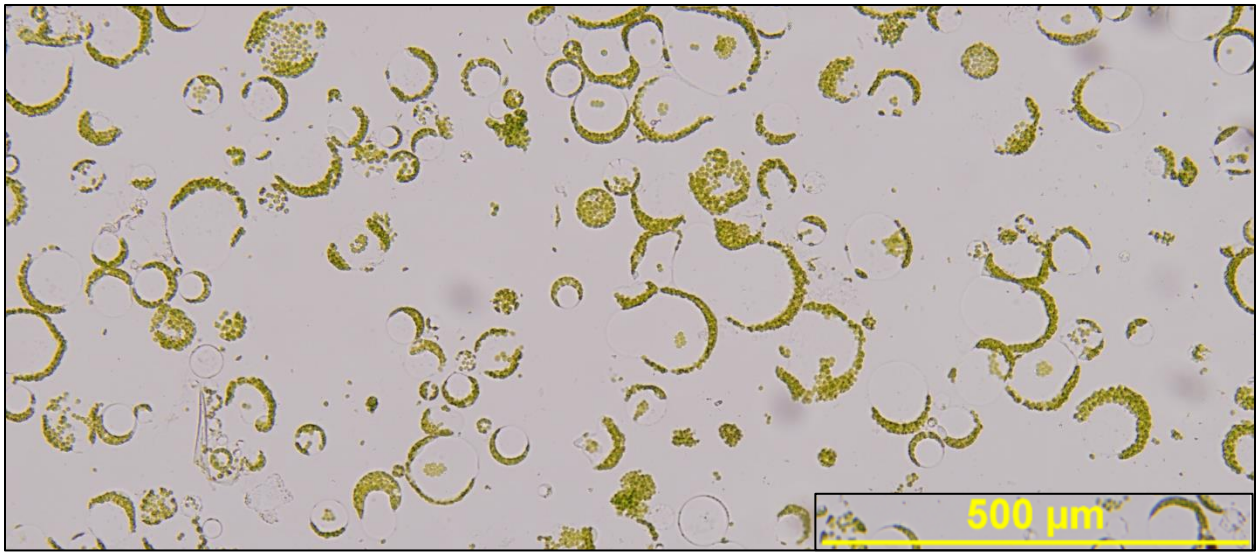


Figure 3-9: Debris left over from protoplasts which (mostly) have completely lysed without leaving any vacuoles.

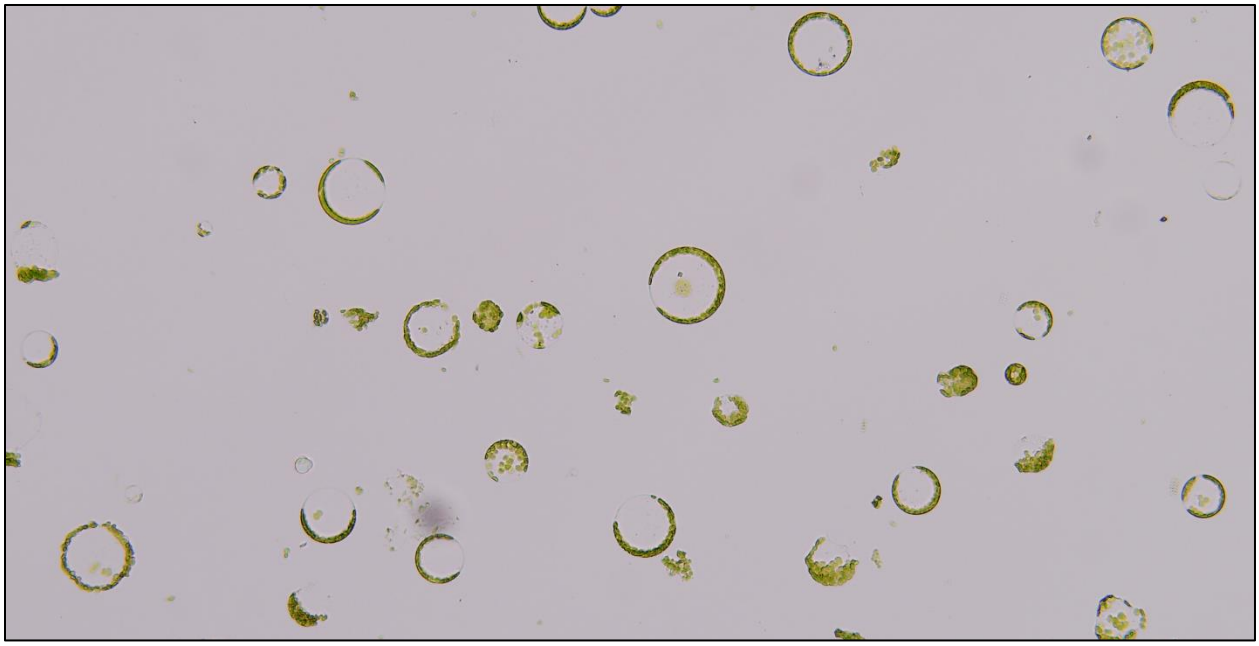


Figure 3-10: Protoplasts with “crescent morphology”. Notice the vacuoles in the middle take up most of the cell and chloroplasts are usually pushed toward one side of the cell. The larger protoplasts are between 25-50µm.

Vacuoles were observed after the addition of lysis buffer to confirm success up until that point in the protocol. Inconsistent staining in the vacuoles was a problem which, although not critical, was never fully resolved. As seen in figures 3-11, 3-12, and 3-13 below, some vacuoles only stained a very light pink or not at all. Vacuoles which did not stain pink constitute a majority of vacuoles observed throughout the project. One observation which was consistent throughout all experiments is that small vacuoles stain at a higher frequency and more intensely than large ones. This observation, along with those about the health of protoplasts, suggest that many large vacuoles may have an altered pH because of leakage, dilution, or another reason. In one experiment, 5mM ATP was added to the lysis buffer to see if vacuolar ATPases which acidify the vacuole (Gaxiola et al. 2007) could cause more vacuoles to take up the neutral red dye, however no effect was observed.

We also observed that chloroplasts are often tightly attached to the vacuoles, and this appears to be the only reason why intense or extended centrifuging is necessary for purification.

Chloroplasts were observed at varying degrees of attachment, from firmly attached after the lysis step, to barely hanging on after centrifugation in the density gradient. (Figures 3-11, 3-12, 3-13, 3-14) The sheering force of the centrifuge seems to cause the chloroplasts to progressively bunch together on one side of the vacuole, less and less in the crescent shape seen in some protoplasts and more clumped together, until presumably they are sheered from vacuole.

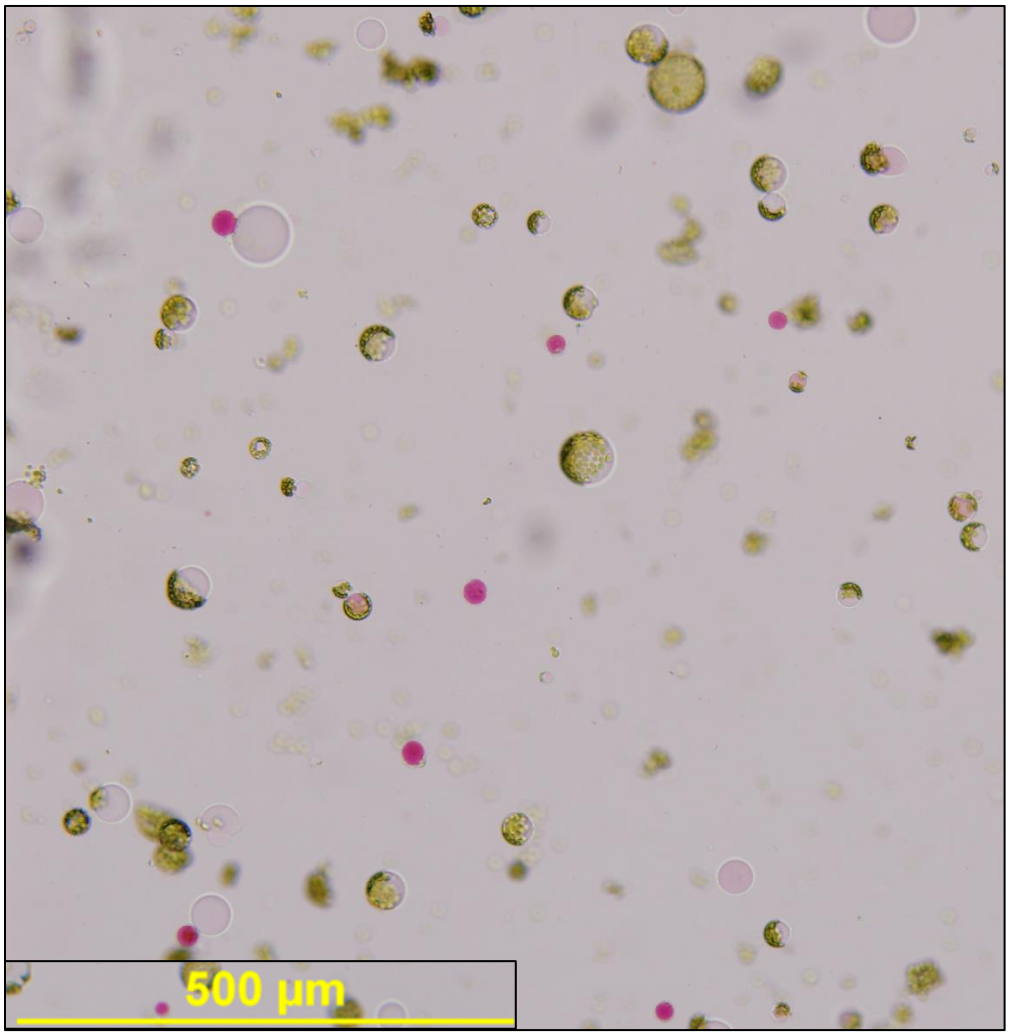


Figure 3-11: Stained vacuoles (dark pink) and unstained vacuoles (opaque) using neutral red dye plus some unruptured protoplasts after the addition of lysis buffer to protoplast-containing solution.

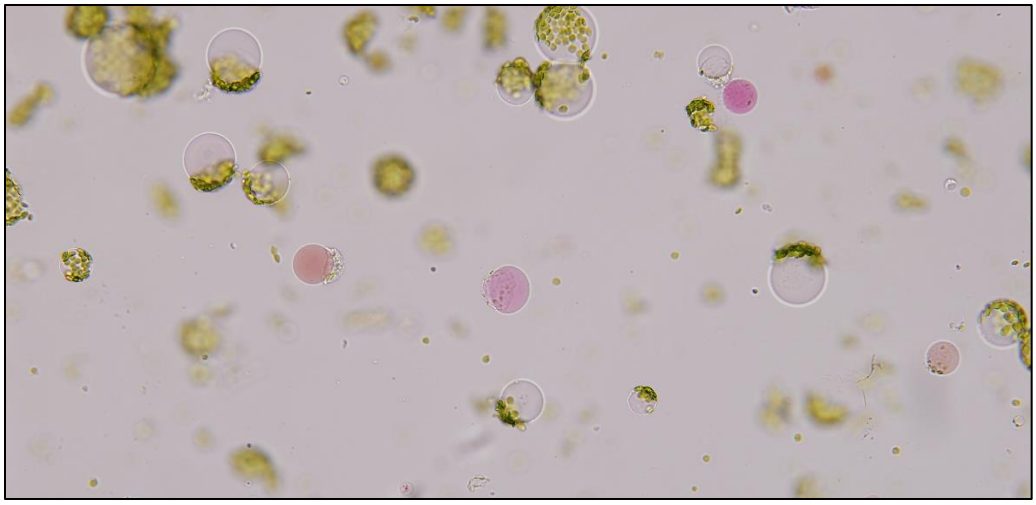


Figure 3-12: Vacuoles and debris after the addition of lysis buffer to protoplasts. The stained vacuoles are between 20-40µm.

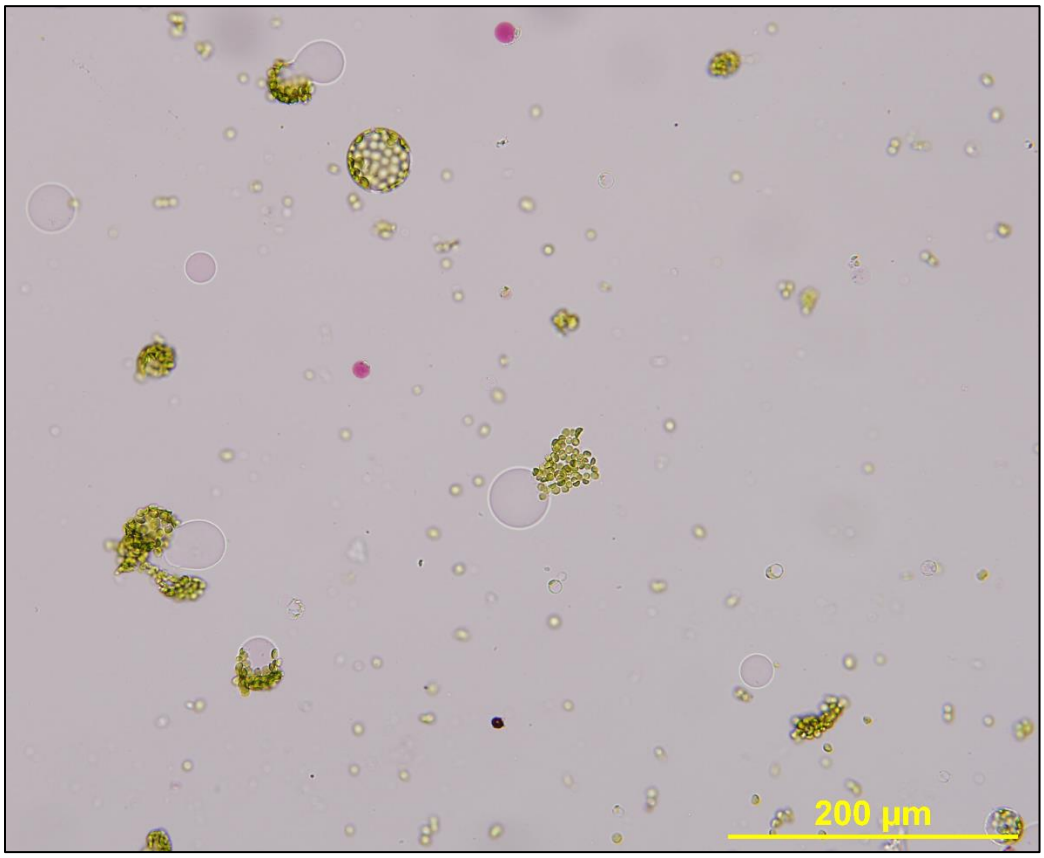


Figure 3-13: Vacuoles and debris after the addition of lysis buffer to protoplasts.

The most significant challenge which prevented us from achieving fully purified vacuoles with the efficiency we required was the lack of access to an ultracentrifuge suitable for the scale we needed. Only full-scale isolation runs could be brought to completion because the tube sizes

available were too large be able to effectively isolate and collect the final product from smaller scale isolation attempts. Filling the 26.3mL ultracentrifuge tubes requires even more material than a full-scale isolation in the original protocol. To test and then utilize the full-scale procedure repeatedly for our monolignol transport experiments would have used an excessive amount of resources, particularly mannitol, Ficoll, and the enzymes which are quite expensive. The speed at which to run the ultracentrifuge also had to be determined experimentally. The long length of the tubes means that the difference between g-forces at the minimal and maximal radii are large, which affects component separation significantly, and the gaps between conversions from rpm to g-forces in the rotor manual are large (every 5,000 rpm). Also, two of the original protocols call for different g-forces to be applied: Robert et al. 2007 calls for 71,000g while Zouhar 2017 calls for 50,000g. The optimal speed was investigated and is estimated to be 26,500-28,000 rpm but testing was not taken to completion. Without more knowledge about how long intact vacuoles could be stored, and the quantity and purity attainable from a single extraction, it was not feasible to go through all the necessary testing before even beginning our planned experiments on monolignol transport while incurring such material costs. Two micro-ultracentrifuges appropriate for our desired scale were permanently out of service but inspired the idea of trying using a microcentrifuge to circumvent these problems (see section 3.4).

The vacuole collection step at the end of the protocol was also more problematic than expected and influenced more significant protocol modifications made later on. An ordinary micropipette is supposed to be able to collect the vacuoles where they settle during ultracentrifugation: at the interface of the 0% Ficoll and 4% Ficoll layers. The stained vacuoles appear visible as a faint red band as prescribed, however, it is difficult to collect the vacuoles without aspirating a large amount of solution with it and difficult to determine when everything

has been collected, making the vacuole solution very dilute. A significant amount of Ficoll also gets aspirated, which is undesirable for handling the solution and degrades within a few days, potentially altering the osmolarity of the solution. No instructions beyond collecting the vacuole with a micropipette were given in the original protocols to address these problems.

Figure 3-14 shows the results of a full-scale isolation attempt after ultracentrifugation for 50 minutes as prescribed in the original protocols. After collection, the vacuoles were pelleted in a microcentrifuge at 13,000g for 5 minutes. This attempt was the most successful run at full scale before we began to make more significant alterations to the protocol at smaller scale. Notice how on almost all the vacuoles, the chloroplasts are narrowly bunched together in one area, yet still firmly attached. Also notice the inconsistent staining, with smaller vacuoles staining more frequently and intensely than larger ones. In this purification attempt, the speed of the ultracentrifuge was not optimized, running at 26,000g (likely below optimum), although presumably if these samples had been run at the correct speed for sufficient time, the debris and chloroplasts still clinging to the vacuoles could have been successfully removed.

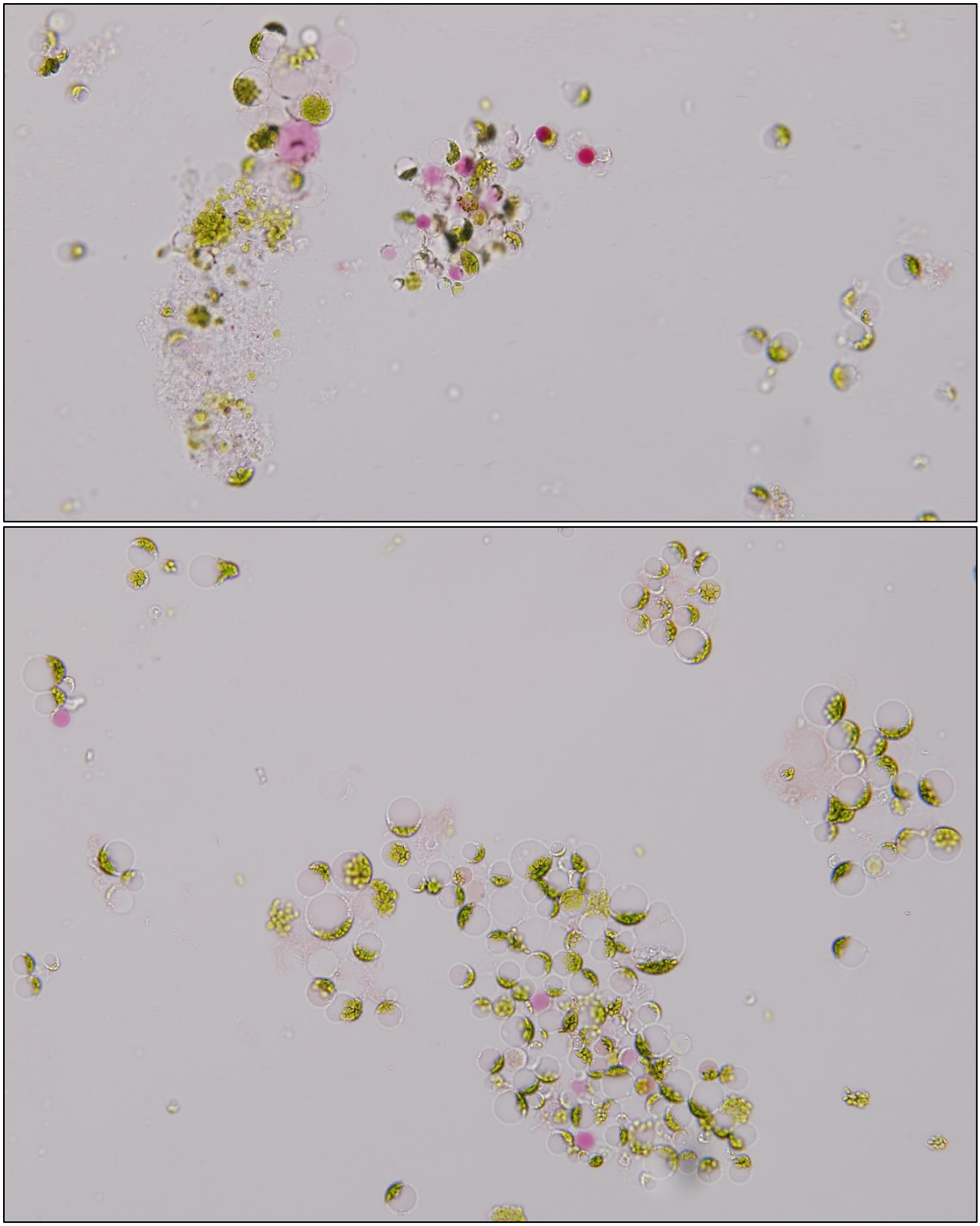


Figure 3-14: Clumps of vacuoles and associated debris from a full-scale vacuole isolation attempt, after ultracentrifugation, collection, and pelleting. Most of the vacuoles or protoplast are between 15-40µm.

3.4. New Goals and Innovations

The challenges we encountered while adapting this protocol to our lab occurred in all three main steps of vacuole isolation: protoplast isolation, cell lysis, and vacuole purification. However, these challenges culminated at the ultracentrifugation step, resulting in high material costs, a difficult technique to perform (creating Ficoll gradients in large tubes), little flexibility in scale, and difficulty collecting the final product among other problems. While it is likely that with sufficient testing large scale vacuole isolation would have been achieved, we decided to deviate from the protocol to try to circumvent the problems caused by our need for an ultracentrifuge. The goals of these new, more significant adaptations were to make the protocol more efficient by reducing material costs, more flexible by making the protocol easier to scale to researchers' requirements, easier to perform by modifying technically difficult tasks, and more accessible by eliminating the need for an ultracentrifuge, which many institutions cannot reasonably attain.

Some progress was already made toward such goals in the minor adjustments made during early small-scale tests. Therefore, we sought to test the feasibility of removing the ultracentrifuge from the protocol completely. Several papers indicated that this was a reasonable goal. A vacuole isolation protocol by Gao et al. 2017 from the leaves of *Sedum alfredii* used speeds of no more than 1,500g for purification. Another protocol for vacuole isolation by Song et al. 2014 from both barley and Arabidopsis leaves used speeds of no more than 100g, and a protocol by Shimaoka et al. 2004 used no more than 800g, although these protocols used more complicated series of Ficoll or Percoll density gradients and the level of purity is uncertain. The protocol described in Carter et al. 2004, Robert et al. 2007, and Zouhar 2017 may have been optimized for tonoplast extraction instead of keeping vacuoles intact, since Zouhar 2017 includes additional steps for tonoplast

purification and Carter et al. 2004 is a proteomics paper which did not use the vacuoles for the assays we seek to perform.

The first question to answer was whether we could achieve at least a semi-purification using a microcentrifuge, as we had done with the ultracentrifuge without optimization. To test this, two lysis buffers were made: the normal lysis buffer containing 10% Ficoll, and some lysis buffer without Ficoll. The modified buffer *without* Ficoll was added to the protoplast solution to confirm that the absence of Ficoll does not affect lysis. Normal lysis was confirmed and the 0% Ficoll lysis buffer with vacuoles was layered on top of the 10% Ficoll lysis buffer. Creating the density gradient was far easier in the 2mL microcentrifuge tubes than in the 26.3mL ultracentrifuge tubes, and the gradient was spun in an Eppendorf 5424R microcentrifuge at 20,000g (near the maximum speed) for 15 minutes at 20°C. The result was the chloroplasts and heavy debris pelleted at the bottom of the tube and a faint red band of vacuoles materialized at the interface of the 10% and 0% Ficoll solutions, exactly as prescribed by the original protocol after ultracentrifugation. Observed under the microscope, the vacuoles were extremely dilute because the collection step still proved problematic despite the much smaller tube size.

Hereafter, the 4% Ficoll solution used in the original protocol was not used, saving the work of making an extra solution. The bottom layer was changed to ice-cold vacuole storage buffer with the addition of 10% Ficoll. To solve the problem of vacuoles being too dilute after the collection step, a pelleting step was added following instructions from a yeast vacuole isolation procedure (Cools et al. 2020). Vacuoles were collected from the faint red band, placed in a 2mL tube, and spun at 13,000g for 20 minutes (far longer than prescribed). Unexpectedly, no pellet formed, but instead the red band reformed very close to the top of the solution. While collecting the vacuoles, enough of the Ficoll-containing bottom layer had been aspirated to prevent the

vacuoles from pelleting in the resultant solution. The new red band of vacuoles was carefully aspirated from the top, placed in a new 2mL tube and diluted as much as possible with fresh vacuole storage buffer (0% Ficoll) and centrifuged as before. The vacuoles successfully pelleted and were observed to have better purity than we had previously achieved (Figure 3-15).

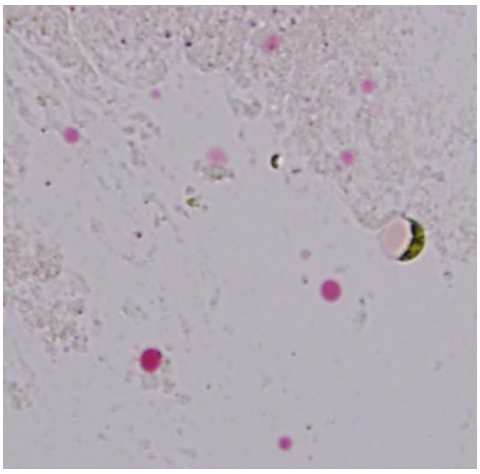


Figure 3-15: Several vacuoles-stained deep pink after purification, pelleting, and resuspension for examination. The rightmost vacuole still has some chloroplasts attached on the right side. The grainy residue may be condensed Ficoll. The vacuoles are between 10-20μm.

These results indicated that purification with a microcentrifuge is feasible, however, at 1/20 scale the pellet was too small for post-isolation testing or any transport experiments. Our larger centrifuge, an Eppendorf 5804R, has a maximum speed of just over 3,000g so slower speeds were tested before scaling up. Spinning at 3,000g for 20 minutes also successfully produced a red band of vacuoles and was able to pellet the vacuoles after collection, however, we did not achieve the same level of purity, measured by the relative absence of chloroplasts and other debris, as at higher speeds. We decided to continue testing using 1/4 scale isolation attempts in 15mL at the lower speeds. If successful, this would enable us to completely scale up to the full scale of the original protocol or even greater if desired using 50mL tubes, achieving full scalability of our protocol.

The results of our vacuole isolation attempt at 1/4 scale using 3,000g spins were inconsistent. Out of six using these parameters, only two produced vacuoles, and those which did

produced very few (Figure 3-17). The failure of these experiments can be partially attributed to newfound difficulties in protoplast isolation and lysis steps prior to the purification. Protoplasts from several attempts appeared swollen and obtained the "crescent morphology" previously observed to varying degrees (Figure 3-16). Also, fewer stained vacuoles were observed than expected. The age of the plant material, ranging from 34 to 51 days, likely contributed for those at the higher end, but cannot account for the poor quality of protoplasts or failed lysis alone, especially because some success was achieved with similar age plants when testing purification using the microcentrifuge.

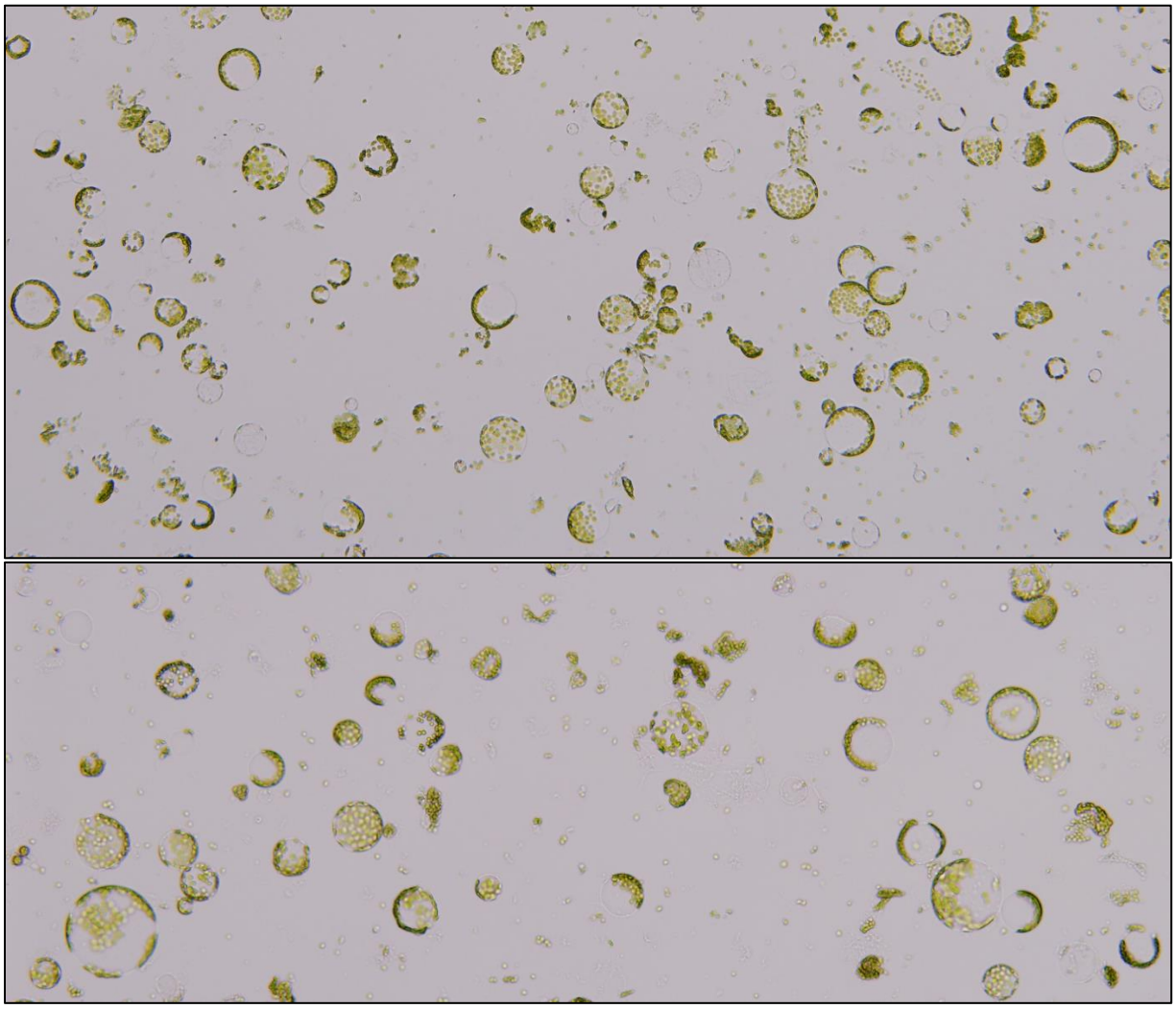


Figure 3-16: Protoplasts after the enzyme digest step from ¼ scale vacuole isolation attempts (7.5 mL total solution volume). Many of the protoplasts are swollen or have "crescent morphology", but some are healthy. Most of the protoplasts are between 20-50µm.

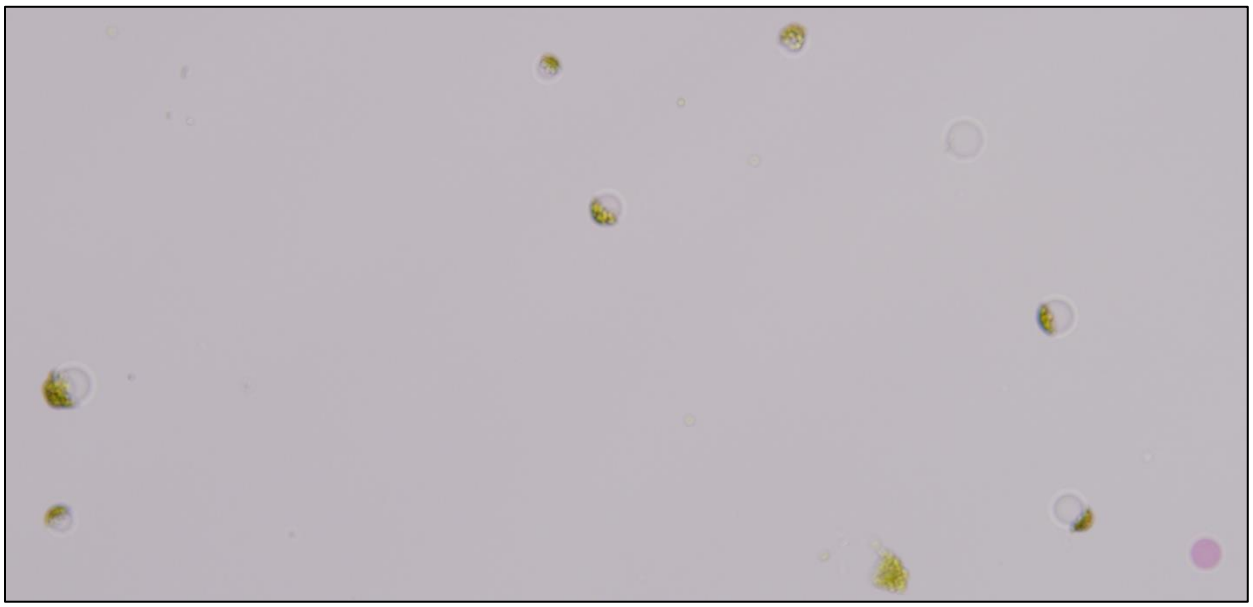


Figure 3-17: Partially purified vacuoles from a ¼ scale isolation attempt after centrifuging for 15 minutes at 3,000g. Notice how the chloroplasts are bunched up on one side of the vacuole. The protoplasts are between 15-40µm.

The barriers encountered at this stage in development remain unexplained, as testing for this stage of the project was halted in late June of 2022. This includes an explanation for the poor health of the protoplasts and the absence of vacuoles after purification in most 1/4 scale experiments. Another phenomenon that occurred during two of the attempts was that the top layer without Ficoll turned light red, while the bottom layer with 10% Ficoll turned light green (Figure 3-18). The light green color can only be presumed to be chloroplasts which did not pellet, and the red color may have come from scattered vacuoles (optimistically), or the solution somehow become acidic and changed the color of the neutral red dye. The inconsistency of this phenomenon, similar to vacuole production, prevents us from determining why this occurred for now. Another unexplained occurrence is that when 4% Ficoll vacuole storage buffer was tested as the bottom layer in the density gradient in place of 10% Ficoll and spun at 3,000g, no red band of vacuoles formed, and vacuoles could not be found in the pellet or the supernatant. This is surprising because

in the original protocol, vacuoles settle at the interface of the 0% and 4% Ficoll layers after centrifugation at very high speeds Due to this result, 10% Ficoll continued to be used.

Figure 3-18: The appearance of the Ficoll density gradient as described above, at ¼ scale in a 15mL tube after centrifugation at 3,000g for 15 minutes. It remains unclear why this phenomenon involving separation of colors occurred two times only.

3.5. Future Development

3.5.1. Completing the Isolation Protocol

In future development efforts, priority should be given to standardizing how protoplasts and vacuoles are sampled and evaluated. Additional expertise on the durability of protoplasts and vacuoles, and their health at various morphologies could assist significantly with this. Protoplasts and vacuoles were handled minimally when possible, because of the possibility that they would break if handled too much. A standardized sampling procedure and evaluation standards can produce more useful information including quantitative measure of success, which have been lacking. Checking the state of the protoplasts and vacuoles after the appropriate step every time an isolation experiment is done will also close information gaps which occasionally caused

uncertainty when interpreting results. Training in proper slide preparation techniques and how to use a hemocytometer are also necessary.

It is possible that the lower centrifugation speeds tested in later experiments need to run for an extended time to achieve the desired results. Using the larger 15mL tubes may take longer to separate components in because of the longer migration distances involved. Further tests at both 1/20 and 1/4 scale will be needed to determine the centrifugation parameters necessary to achieve purity. An unlikely and unfortunate scenario is that the vacuoles are degrading or bursting during the purification step. A more likely scenario is that if the protoplast isolation and lysis steps are further optimized to produce high-quality, well-stained vacuoles consistently, then only the centrifugation parameters at low speed need to be determined. In this case, we may not have been collecting all the vacuoles available or collecting in the wrong place. The collection step was never truly improved, and a better technique would greatly improve the protocol ending. In addition to spin time, temperature and solution volumes could be adjusted to optimize the process. It is also possible to perform the protoplast isolation and lysis steps at a large scale and then to use many separate microcentrifuge tubes for purification. Although this would require more pipetting, it could potentially be an efficient way to purify the vacuoles, especially if the higher speeds attainable by a microcentrifuge are highly beneficial. Lastly, obtaining Arabidopsis seed and sowing plants in mass (possibly with the help of a collaborator) will provide greater flexibility to choose the best quality material for experiments without having to contend with conserving material or timing experiments as much.

3.5.2. Post-isolation Testing

An intermediate step between isolating vacuoles and using them in transport experiments is to verify the quality of the product obtained. It will be necessary to take aliquots of vacuole-

containing solution and measure their purity, yield, and intactness. Three assays can be used to accomplish this. To measure the purity of the vacuoles, we purchased three rabbit antibodies from Agrisera which bind to proteins localizing to different compartments of the plant cell: an H^+ ATPase for the plasma membrane, BiP for the endoplasmic reticulum (ER), and the epsilon subunit of V-ATPase for the tonoplast. A western blot protocol will be used to detect the presence or absence of these proteins; pure vacuoles should contain V-ATPase but not BiP or H^+ ATPase (Gomez and Chrispeels 1993; Shimaoka et al. 2004). Yield will be determined by measuring total protein content as a proxy, using the BCA protein assay kit by Thermo Fisher. It is less certain how intactness will be confirmed. Readdressing problem of inconsistent staining will likely be part of determining this. In the case that some vacuoles remain unstained because of leakage and pH changing during the isolation protocol, a leakage assay not involving neutral red may be necessary. If the purity, yield, and intactness of the vacuoles all meet predetermined quality standards for transport assay experiments, then the project can proceed into the next phase of testing our hypotheses about monolignol transport.

4. Testing Hypotheses about Monolignol Transport

Isolating a sufficient quantity of purified, intact vacuoles from *Arabidopsis thaliana* leaves will enable us to perform kinetics assays to characterize mechanisms of molecular transport across the tonoplast. To test hypotheses regarding how monolignols and their derivative compounds enter the vacuole, we will introduce these compounds to the vacuoles in solution as potential substrates for tonoplast transporter proteins and attempt to modulate the kinetics of the uptake process by adding various known pharmacological inhibitors of active transporters. By modulating the kinetics of substrate uptake, we can derive information about which classes of active transporters, if any, are acting on the substrate compound. If the uptake of a substrate is inhibited by an inhibitor

of a certain class of transporter proteins, we have evidence that a member of that class may be responsible for transporting that substrate into the vacuole (Figure 3-19). Identification of the specific protein involved would require further investigation. The uptake assays will be halted using a clean-up procedure and the vacuoles will be stored at -80°C until they are prepared for targeted analysis using HPLC/MS.

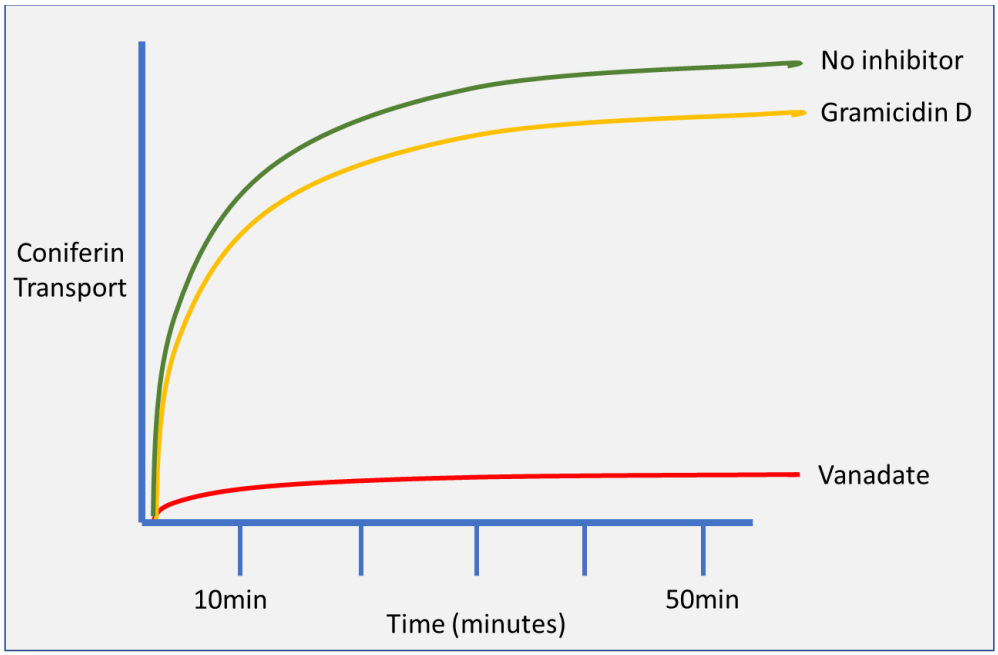


Figure 3-19: A graph of expected results in the case that coniferin is transported into the vacuole by an ABC transporter. Vanadate inhibits ABC transporter activity, so we would expect greatly reduced transport into the vacuole with vanadate present. Gramicidin D inhibits SLC transporters, so we would not expect a drastic reduction in coniferin accumulation with gramicidin D present.

We intend to apply a variety of modulators, each of which inhibits the activity of a known class of transporter proteins, to the reaction mixture. ATP (5mM) will be added by default to all reaction mixtures but can be withheld to inhibit ABC transporters. Orthovanadate is a well-known inhibitor of many transmembrane proteins with ATPase activity, including ABC transporters, and will be used to confirm ABC transporter activity. Gramicidin D, nigericin, and CCCP (carbonyl cyanide m-chlorophenyl hydrazone) are chemicals which disrupt pH gradients or act as an ionophore and can inhibit SLC transporters which depend on chemical gradients for energy.

Beginning with inhibitors of broad transporter classes, we can incrementally collect more specific evidence of which kind of transporter is responsible for substrate transport. For example, cimetidine and pyrimethamine are inhibitors of human MATE (multi-antimicrobial extrusion) proteins, a subclass of the SLC superfamily (SLC47), and any effect from them would be highly suggestive of activity by a related protein. MK571 specifically inhibits ABCC family transporters in mammalian species and would suggest similar information. Again, discovering the specific transporter may not be possible using only this method, but the general family can be discerned with strong evidence.

The most important substrates for us to test are the three glycosylated monolignols coniferin, syringin, and p-coumaryl alcohol glucoside. The three corresponding aglycones coniferyl alcohol, syringyl alcohol, and p-coumaryl alcohol are lower priority substrates, but could provide validation of previous observations that only glycosylated monolignols enter the vacuole (Miao and Liu 2010). Lastly, testing certain compounds with structures similar to the monolignols could provide additional information about how these compounds physically interact with transporter proteins. Such compounds we plan to use include ferulic acid, caffeic acid, and caffeic acid beta-D-glucuronide.

A proposed protocol for the uptake experiments is as follows. In a 1.5mL microcentrifuge tube, prepare 500µL of the reaction mixture consisting of appropriate concentrations of the substrate, modulator, and ATP in vacuole storage buffer (Table 3-3). All monolignol substrates will be tested at 50µM in accordance with Tsuyama et al. 2013 To begin the reaction, add a volume of solution containing the isolated vacuoles equivalent to 80µg of vacuolar protein as determined by the BCA test (Miao and Liu 2010). Immediately place the tube into a Thermomixer at 25°C and gently agitate at 300rpm. At selected time points between 1 and 60 minutes, withdraw 50µL

of the reaction mixture and dilute it in 450µL of ice-cold vacuole storage buffer. Spin down the tube for 5 minutes at 5,000g, discard the supernatant, and add 500µL of fresh ice-cold vacuole buffer (Cools et al. 2020). Repeat the spin down, discard the supernatant, and add only 100µL of fresh ice-cold vacuole storage buffer. The contents of the vacuoles will subsequently be analyzed by HPLC/MS. Freeze the sample at -80°C until ready for HPLC/MS preparation.

Table 3-3: Selected available pharmacological modulators of transporter proteins and appropriate concentrations for their use in uptake experiments. Control experiments may be performed to adjust the concentrations of modulators to optimal concentrations.

Modulator	**Concentration (µM)**	**Reference**
ATP	5mM or none	(Miao and Liu 2010)
Orthovanadate	1mM	(Tsuyama et al. 2013; Väisänen et al. 2020)
Gramicidin D	25µM	(Tsuyama et al. 2013)
Nigericin	2µM	(Miao and Liu 2010)
CCCP	20µM	(Väisänen et al. 2020)
Cimetidine	10µM	(Elsby et al. 2017)
Pyrimethamine	1µM	(Elsby et al. 2017)
MK571	5µM	(Pawarode et al. 2007)

Each pharmacological inhibitor will be a considered as an experimental treatment and compared to a negative control with 5mM ATP and no modulator. The Dunnett method will be used to determine which treatments are significantly different from the control. The extraction run used to obtain vacuoles from a set of *Arabidopsis thaliana* leaves will be used as a blocking variable and we will attempt to include at least one replicate of each treatment in each block. At least three replicates of each treatment will be performed.

It is important to note that investigating the transport of the glycosylated monolignols is just one specific application that this procedure can be applied to. Once we can reliably obtain pure, intact vacuoles from an efficient isolation procedure, the substrates and modulators can be altered to test a variety of hypotheses concerning the transport of many compounds, whether endogenous, environmental, or anthropogenic in origin.

Appendix 1: UCSF-FDA TransPortal – Additional Procedures, Findings, and Discussion Topics

1. UCSF-FDA TransPortal Literature Search

The literature search for the UCSF-FDA Transportal update was performed concurrently with the literature search for the new UC Transportal database, but with certain limitations. The literature on interactions between chemicals and drug transporter proteins was searched using PubMed (https://pubmed.ncbi.nlm.nih.gov), Google Scholar (https://scholar.google.com) and Web of Science (https://www.webofscience.com). Keywords such as "environmental chemical", "pesticide", "inhibition", "substrate" and "IC_{50}", were combined with protein nomenclature terms, such as "ABC transporter", "SLC transporter", "ABCB1", etc. The results were further filtered to only include mammalian ABC and SLC superfamily drug transporters, including human (*Homo sapiens)*, mouse (*Mus musculus*), rat (*Rattus norvegicus*), and monkey (*Chlorocebus aethiops*). Search results were examined for any quantitative parameters descriptive of the kinetics of transporter-chemical interactions including K_m, IC_{50}, and K_i values. Research articles in which any of these values were present were selected for thorough review. Only primary research articles were selected; review articles were omitted. In these articles, only data from *in vitro* assays in which a specific transporter protein was identified as responsible for the observed interaction were incorporated into the database.

2. Data Standardization

All kinetic values and interaction types were recorded according to the published interpretation of the authors. Abbreviations, acronyms, or alternate spellings of compound names between research articles were harmonized to create a concise index of all chemicals in the

database. Certain salts were also merged to reflect the chemical of interest (for example "vincristine sulfate" was counted as "vincristine").

An EC_{50} value is the concentration of a substance at which a response with magnitude halfway between the baseline and maximal response is induced. As such, an EC_{50} value can be assigned to many different observation types, including substrate transport or inhibitory effects. Therefore, any EC_{50} values from the literature were given close review and recorded as an IC_{50} or K_m value only if the investigator's method of analysis caused the reported value to match the definition of an IC_{50} or K_m value (see section 4.2).

3. Inhibitory Versus Substrate Interactions

Inhibitor data makes up just over 80% of the UCSF-FDA Transportal database. Hence, the proteins with the most total interactions also tend to have the most inhibitor data. The five transporters with the most substrate data are SLCO1B1 (n = 74), ABCB1 (n = 71), SLC22A2 (n = 68), SLC22A1 (n = 55), and SLCO1B3 (n = 54). Among the 15 transporters with over 100 total interactions, only four have more than 20% substrate data: SLCO1B1 (25.3%), ABCG2 (28.3%), ABCB11 (33.7%), and SLCO1B3 (36.0%) (Figure S1-1).

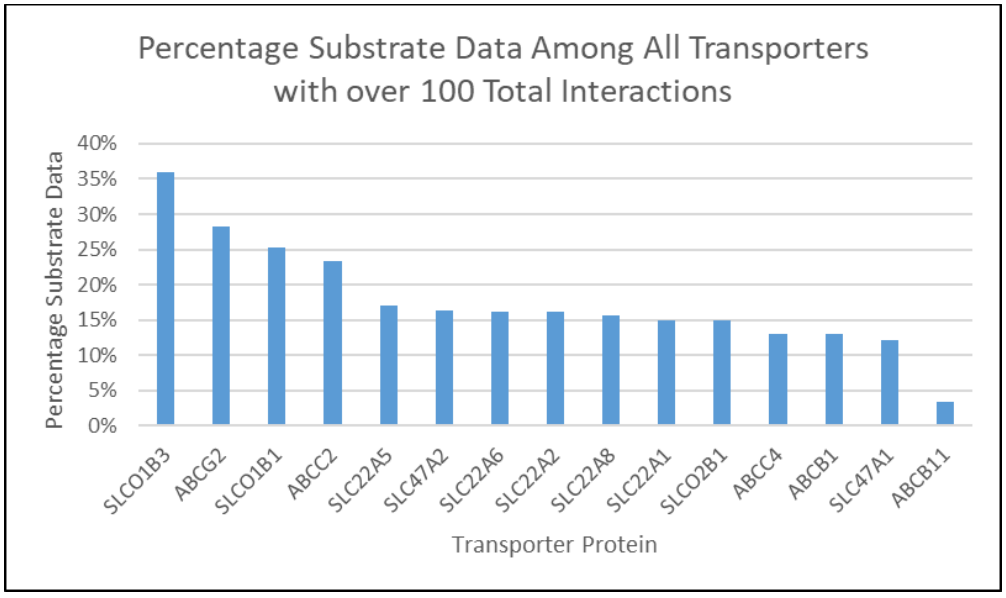


Figure S1-1: The percentage of the total data that is substrate data for all transporters with over 100 total interactions

The difference between the number of substrates versus inhibitors examined per article is striking. The largest number of K_m values measured in any article was 22, followed by 19 and 13. Meanwhile, 35 different articles contributed 20 or more IC_{50} and K_i values, with the top five articles contributing 331, 142, 105, 100 and 80 inhibitor interactions respectively (Figure S1-2). We also found that data produced from assays for inhibitory interactions and substrate interactions tend to be published separately. A majority of the top 20 articles with the most inhibitory interaction and the top 20 articles with the most substrate interactions (Figure X) contained zero interactions of the other kind.

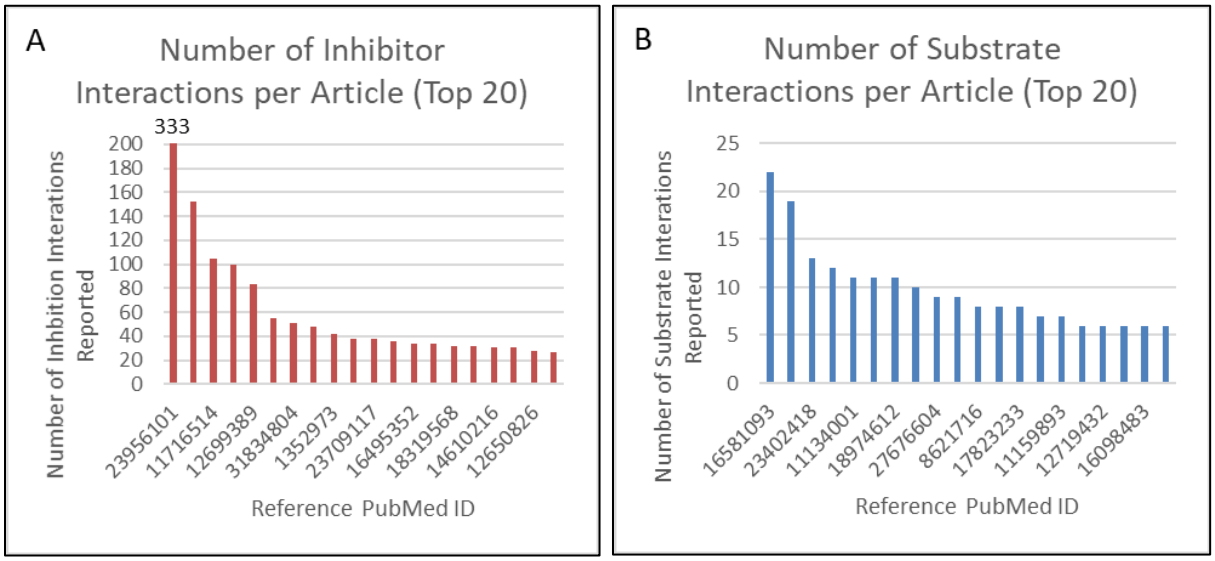


Figure S1-2: A bar chart of the number of inhibitor (A) and substrate (B) interactions provided by the top 20 providers of inhibitor and substrate interactions respectively. Chart A is cut off at 200 interactions for viewability, but the top contributor has 333 interactions.

4. Challenges and Pitfalls in Data Analysis

4.1. Data Monopoly

The top 15 articles, about 2.5% of all articles, containing the most data contribute 1213 interactions, approximately 30% of the total interactions. There are 179 articles which contribute just one interaction each and 101 contributing two interactions each, together making up close to half of the articles (47%) (n = 280) yet contributing just under 10% (n = 381) of the data. There

are four cases in which data from transporters with more than 25 total interactions received over half of their data from one article. These transporters are mouse_abcb1a (receiving 66.0% of its data from one article), ABCC3 (71.0%), ABCC4 (73.8%), and ABCB11 (87.1%).

4.2. EC_{50} Values

A complicating factor in our literature search was the question of whether to accept reported EC_{50} values as IC_{50} and K_m values in the database. EC_{50} values are widely used to describe the magnitude of effects observed in studies but do not have a consistent definition. We observed that across studies, liberal treatment was given to the EC_{50} concept to fit the context of the study at hand. While treating EC_{50} values this way may provide a convenient metric for investigators to report results with, they can be potentially misleading if incorrectly compared with one another. For example, to compare an EC_{50} value measuring the halfway inhibition of chemical transport of a substrate in one article with an EC_{50} measuring halfway cell mortality due to inhibition of chemical transport of a cytotoxic agent in another study could yield a serious misunderstanding of the data. We were careful to strictly adhere to the definitions of an IC_{50} or K_m value when reviewing EC_{50} values and would recommend that researchers consider using terminology with narrower definitions when reporting kinetic data.

5. Supplemental Tables

Table S1-1: The full list of articles referenced in UCSF-FDA TransPortal + UCSD/UCD-NIEHS TICBase. NA = "Not Applicable". DOI = "Digital Object Identifier". Many older references do not have a DOI because they were published well before widespread adoption of the DOI system, which was created in the year 2000.

PubMed ID	DOI	Year of Publication	First Author
1352973	10.1016/0006-2952(92)90149-d	1992	Höllt
1356264	10.1073/pnas.89.18.8472	1992	Ambudkar
1662481	10.1111/j.1440-1673.1991.tb03013.x	1991	Kreel
2646988	10.1016/s0196-6553(89)80008-2	1989	Leibovici
7495840	10.1016/0005-2736(95)00178-7	1995	Ramamoorthy
7514263	NA	1994	Rao
7557095	10.1016/0016-5085(95)90588-x	1995	Kullak-Ublick

7562598	NA	1995	Hirai
7681059	NA	1993	Saeki
7805856	10.1016/0014-5793(94)01282-2	1994	Borrel
7961706	NA	1994	Leier
8100632	10.1023/a:1018972102702	1993	Hunter
8100817	NA	1993	Hunter
8396335	10.1152/ajpgi.1993.265.2.G289	1993	Saito
8621644	10.1074/jbc.271.16.9683	1996	Loe
8621716	10.1074/jbc.271.6.3163	1996	Borgnia
8640791	NA	1996	Jedlitschky
8842452	10.1111/j.1476-5381.1996.tb15612.x	1996	Fricker
8862725	10.1097/00001813-199607000-00012	1996	Tiberghien
8917702	10.1006/taap.1996.0286	1996	Bain
8938553	10.1016/s0168-8278(96)80246-7	1996	Bossuyt
9187257	10.1124/mol.51.6.913	1997	Zhang
9187270	10.1124/mol.51.6.1034	1997	Loe
9260930	10.1089/dna.1997.16.871	1997	Gorboulev
9262363	NA	1997	Ito
9347896	10.1289/ehp.97105812	1997	Bain
9355767	10.1042/bj3270305	1997	Jedlitschky
9380680	10.1073/pnas.94.20.10594	1997	Dey
9458785	10.1152/ajpgi.1998.274.1.G157	1998	Craddock
9530286	10.1021/bi972709x	1998	Lee
9531522	NA	1998	Walle
9637710	10.1172/JCI1909	1998	Döring
9655880	NA	1998	Zhang
9687576	10.1124/mol.54.2.342	1998	Busch
9688051	10.1023/a:1011965707998	1998	Döppenschmitt
9751076	10.1016/s0006-2952(98)00212-3	1998	Shepard
9753615	10.1006/bbrc.1998.9298	1998	Balimane
9794924	10.1002/hep.510280528	1998	Huang
9823324	NA	1998	Chu
9887087	10.1152/ajprenal.1999.276.1.F122	1999	Hosoyamada
9914792	10.1111/j.1349-7006.1998.tb00518.x	1998	Kusunoki
9950961	10.1152/ajprenal.1999.276.2.F295	1999	Lu
10024515	NA	1999	Paulusma
10052994	10.1021/js980132e	1999	Han
10075817	10.1006/abio.1998.3087	1999	Bebawy
10082798	10.1016/s0005-2736(99)00005-x	1999	Okuda
10087037	NA	1999	Guo
10198227	10.1006/bbrc.1999.0475	1999	Wigler
10215651	NA	1999	Yabuuchi

10220572	NA	1999	Cui
10330006	10.1152/ajpgi.1999.276.5.G1153	1999	Niinuma
10385678	10.1124/mol.56.1.1	1999	Gründemann
10421612	NA	1999	Cvetkovic
10421658	10.1002/hep.510300220	1999	Kamisako
10454528	NA	1999	Wu
10462545	10.1124/mol.56.3.570	1999	Cihlar
10525100	NA	1999	Ohashi
10561591	10.1046/j.1432-1327.1999.00885.x	1999	Brandsch
10565843	NA	1999	Kim
10570049	10.1124/mol.56.6.1219	1999	Chen
10604957	NA	2000	Hirohashi
10617675	NA	2000	Lecureur
10636865	10.1074/jbc.275.3.1699	2000	Ganapathy
10660625	10.1074/jbc.275.6.4507	2000	Cha
10681378	NA	2000	Zhang
10688634	NA	2000	Dresser
10703662	10.1681/ASN.V113383	2000	Ho
10725273	10.1038/sj.bjp.0703150	2000	Soldner
10760098	10.1046/j.1523-1755.2000.00007.x	2000	Leier
10773005	NA	2000	Pauli-Magnus
10820137	NA	2000	Choo
10825452	10.1016/s0005-2736(00)00189-9	2000	Wu
10893247	10.1074/jbc.M005463200	2000	Jedlitschky
10901697	NA	2000	Wandel
10929807	10.1006/abio.2000.4633	2000	Cihlar
10947971	NA	2000	Renes
10966924	10.1152/ajprenal.2000.279.3.F449	2000	Wu
10966938	10.1152/ajprenal.2000.279.3.F584	2000	Wagner
10987286	NA	2000	Zeng
10991954	NA	2000	Mulato
10992002	NA	2000	Yamaguchi
10997946	NA	2000	Kawahara
11010964	10.1074/jbc.M005340200	2000	Tamai
11027568	10.1006/bbrc.2000.3554	2000	Wang
11102445	10.1074/jbc.M008251200	2001	Qian
11113639	10.1016/s0928-0987(00)00117-2	2001	Eneroth
11134001	10.1074/jbc.M004968200	2001	Cui
11145223	10.1023/a:1007568811691	2000	Katoh
11159893	10.1053/gast.2001.21176	2001	Kullak-Ublick
11205882	10.1016/s0024-3205(00)00971-1	2000	Takeda
11231118	10.1016/s0928-0987(00)00215-3	2001	Katoh

11306701	10.1124/mol.59.5.1171	2001	Leslie
11306713	10.1124/mol.59.5.1277	2001	Cha
11375950	10.1053/gast.2001.24804	2001	Abe
11375986	10.1074/jbc.M102453200	2001	Leslie
11405287	10.1023/a:1011076217118	2001	Gao
11408360	NA	2001	Störmer
11426832	10.1016/s0014-2999(01)00962-1	2001	Takeda
11437380	10.1006/bbrc.2001.5130	2001	Ozvegy
11447229	10.1074/jbc.M104833200	2001	Chen
11452702	10.1177/00912700122010609	2001	Störmer
11474784	10.1023/a:1011036428972	2001	Wang
11504826	NA	2001	Jang
11504828	NA	2001	Smith
11560874	NA	2001	Shu
11563082	10.1081/NCN-100002341	2001	Cihlar
11581266	10.1074/jbc.M107041200	2001	Zelcer
11585759	NA	2001	Zeng
11602689	NA	2001	Islinger
11606389	NA	2001	Wang
11669456	10.1016/s0024-3205(01)01296-6	2001	Jung
11688982	10.1006/bbrc.2001.5850	2001	Nakatomi
11714740	10.1172/JCI13219	2001	Buyse
11716514	10.1006/bbrc.2001.6000	2001	Wang
11742882	10.1161/hq1201.100262	2001	Love
11743742	10.1021/tx010125x	2001	Wang
11751127	10.1128/AAC.46.1.160-165.2002	2002	Wang
11770010	10.1007/s00210-001-0489-7	2001	Pauli-Magnus
11786185	10.1016/s0196-9781(01)00564-2	2001	Oude Elferink
11815391	10.1038/sj.bjp.0704482	2002	Motojima
11855680	10.1254/jjp.88.69	2002	Babu
11856762	10.1681/ASN.V133595	2002	van Aubel
11861798	10.1124/jpet.300.3.918	2002	Takeda
11880368	10.1074/jbc.M110918200	2002	Iliás
11901101	10.1124/dmd.30.4.457	2002	Guo
11907186	10.1124/jpet.301.1.293	2002	Kimura
11927002	10.1111/j.1349-7006.2002.tb02162.x	2002	Imai
11961113	10.1124/mol.61.5.964	2002	Ekins
11964599	10.1097/00000542-200204000-00019	2002	Wandel
12023506	10.1124/jpet.301.3.797	2002	Enomoto
12031686	10.1016/s0024-3205(02)01680-6	2002	Boulton
12036927	NA	2002	Chen
12063169	10.1016/s0167-4889(02)00187-8	2002	Babu

12065434	10.1124/dmd.30.7.763	2002	Luo
12085361	10.1053/jhep.2002.34133	2002	Vavricka
12110607	10.1038/sj.bjp.0704785	2002	Hayer-Zillgen
12128170	10.1016/s0928-0987(02)00082-9	2002	Takara
12130727	10.1124/jpet.102.034728	2002	Bhardwaj
12130730	10.1124/jpet.102.034330	2002	Takeda
12134945	10.1023/a:1016140429238	2002	Tang
12134946	10.1023/a:1016192413308	2002	Tang
12235265	10.1124/jpet.102.036541	2002	Wacher
12388633	10.1124/jpet.102.037580	2002	Khamdang
12475240	10.1021/bi0205404	2002	Hallén
12490595	10.1124/jpet.102.043026	2003	Tirona
12499648	10.1248/bpb.25.1604	2002	Kakumoto
12523936	10.1042/BJ20021886	2003	Zelcer
12538813	10.1124/jpet.102.041921	2003	Shitara
12568656	10.1042/BJ20030034	2003	Briz
12569305	10.1097/00001813-200302000-00012	2003	Lee
12604693	10.1124/jpet.102.045096	2003	Kumar
12606755	10.1124/mol.63.3.489	2003	Bednarczyk
12636153	10.1023/a:1022359300826	2003	Horie
12650826	10.1016/s0014-2999(03)01381-5	2003	Khamdang
12682043	10.1074/jbc.M212399200	2003	Suzuki
12695538	10.1124/mol.63.5.1094	2003	Reid
12699389	10.1021/jm021012t	2003	Schwab
12700464	10.1097/01.aids.0000060380.78202.b5	2003	Vishnuvardhan
12719432	10.1074/jbc.M301106200	2003	Seward
12731885	10.1021/bi027347u	2003	Paumi
12835412	10.1073/pnas.1033060100	2003	Reid
12867490	10.1124/dmd.31.8.1016	2003	Nunoya
12874005	NA	2003	Chen
12920197	10.1124/mol.64.3.610	2003	Imai
12924948	10.1021/bi034462b	2003	Zhang
12948018	10.1023/a:1025049014674	2003	Troutman
14500392	NA	2003	Volk
14530907	10.1007/s00210-003-0814-4	2003	Fehrenbach
14610216	10.1124/jpet.103.059139	2004	Hasannejad
14610227	10.1124/jpet.103.060194	2004	Nozawa
14660639	10.1074/jbc.M305782200	2004	Ho
14680379	10.1021/tx034101x	2003	Wortelboer
14722317	10.1124/jpet.103.062091	2004	Chu
14729100	10.1016/j.ejphar.2003.10.017	2004	Takeda
15020234	10.1016/j.bbrc.2004.02.063	2004	Terada

15056866	10.1248/bpb.27.559	2004	Watanabe
15107849	10.1038/ng1339	2004	Peltekova
15159445	10.1124/jpet.104.068056	2004	Hirano
15180340	10.1023/b:pham.0000026434.82855.69	2004	Collett
15282265	10.1124/jpet.104.073098	2004	Pinsonneault
15289793	10.1016/j.clpt.2004.03.010	2004	Simonson
15316089	10.1124/jpet.104.072363	2005	Rytting
15359574	10.1023/b:pham.0000036913.90332.b1	2004	Petri
15367706	NA	2004	Yanase
15504935	10.1097/01.ASN.0000143473.64430.AC	2004	Smeets
15548848	10.2133/dmpk.19.369	2004	Uwai
15616150	10.1124/dmd.104.002477	2005	Chen
15618649	10.2133/dmpk.17.23	2002	Horikawa
15640378	10.1124/dmd.104.002337	2005	Satoh
15708966	10.1124/dmd.104.003301	2005	Kobayashi
15795384	10.1073/pnas.0408624102	2005	Gründemann
15817714	10.1165/rcmb.2004-0363OC	2005	Lips
15824923	10.1007/s00280-004-0914-y	2005	Barthomeuf
15832500	10.1021/mp0340136	2004	Su
15863325	10.1016/j.bmcl.2005.03.038	2005	Takahashi
15864504	10.1007/s00424-004-1373-3	2005	Asif
15897250	10.1158/1535-7163.MCT-04-0291	2005	Pratt
15899835	10.1158/0008-5472.CAN-04-2810	2005	Wielinga
15901346	10.1211/0022357055966	2005	Kobayashi
15901796	10.1124/jpet.105.084830	2005	Hirano
15916776	10.1016/j.lfs.2005.04.005	2005	Łania-Pietrzak
15930306	10.1158/0008-5472.CAN-04-1817	2005	Ahmed-Belkacem
15953199	10.1111/j.1365-2885.2005.00655.x	2005	Griffin
15955871	10.1124/mol.105.014019	2005	Hirano
15970799	10.1097/01.fpc.0000170913.73780.5f	2005	Kameyama
16006492	10.1124/jpet.105.088104	2005	Tahara
16014768	10.1124/dmd.105.004622	2005	Shimizu
16046661	10.1124/mol.105.014605	2005	Kopplow
16098483	10.1016/j.bcp.2005.06.024	2005	Ueo
16135657	10.1124/dmd.105.006056	2005	Iwanaga
16141367	10.1124/jpet.105.091223	2005	Bourdet
16184031	10.1097/01.aids.0000183626.74299.77	2005	Woodahl
16210916	10.4161/cbt.4.8.1867	2005	Smith
16225954	10.1016/j.jhep.2005.07.022	2006	Zelcer
16259759	10.1211/jpp.57.10.0009	2005	Fujino
16263091	10.1016/j.bcp.2005.09.011	2005	Müller
16272756	10.2133/dmpk.20.379	2005	Kimura

16282361	10.1152/ajpgi.00354.2005	2006	Rius
16316932	10.1080/00498250500136676	2005	Yamazaki
16330770	10.1073/pnas.0506483102	2005	Otsuka
16337112	10.1016/j.ejps.2005.09.012	2006	Pratt
16384552	10.1016/j.cbi.2005.11.002	2006	Lespine
16394027	10.1152/ajpcell.00622.2005	2006	Biermann
16406207	10.1016/j.ejps.2005.11.011	2006	Keogh
16434549	10.1124/dmd.105.006791	2006	Li
16454695	10.2174/138920006774832550	2006	Tian
16455803	10.1124/dmd.105.008607	2006	Eriksson
16455804	10.1124/dmd.105.008375	2006	Tahara
16455806	10.1124/dmd.105.008615	2006	Rautio
16489126	10.1124/jpet.105.099036	2006	Saito
16490820	10.1161/CIRCULATIONAHA.105.586107	2006	Grube
16495352	10.1093/toxsci/kfj141	2006	Letschert
16501004	10.1124/dmd.105.008888	2006	Nakagomi-Hagihara
16581093	10.1016/j.neuropharm.2006.01.005	2006	Amphoux
16595711	10.1124/dmd.106.009290	2006	Hirano
16608919	10.1124/mol.106.023556	2006	Tamura
16611857	10.1124/dmd.105.009175	2006	Ishiguro
16624871	10.1124/dmd.105.008938	2006	Yamashiro
16627748	10.1124/jpet.106.103390	2006	Liu
16686371	10.1021/mp050063u	2006	Maeda
16697742	10.1053/j.gastro.2006.02.034	2006	Ho
16702441	10.1124/jpet.106.104364	2006	Badagnani
16749865	10.1021/mp0500768	2006	Luo
16807400	10.1681/ASN.2006030205	2006	Masuda
16904803	10.1016/j.fct.2006.07.003	2006	Yoshida
16928358	10.1016/j.bbamem.2006.07.002	2006	Hirano
17005917	10.1124/jpet.106.110379	2007	El-Sheikh
17029589	10.1042/BJ20060632	2007	Kimura
17031644	10.1007/s00280-006-0357-8	2007	Pawarode
17043154	10.1124/jpet.106.112755	2007	Iwanaga
17045309	10.1016/j.neuropharm.2006.07.038	2007	Baltes
17178262	10.1016/j.clpt.2006.09.010	2007	Grube
17192770	10.1038/sj.clpt.6100038	2007	Lau
17202245	10.1093/jac/dkl474	2007	Weiss
17220244	10.1124/dmd.106.011866	2007	Xia
17274666	10.1021/mp060082j	2007	Maeda
17296622	10.1124/dmd.106.014407	2007	Seithel
17301733	10.1038/sj.clpt.6100104	2007	Bailey
17314201	10.1124/jpet.106.116517	2007	Chu

17325024	10.1124/dmd.106.012187	2007	Yu
17374746	10.1124/jpet.106.119073	2007	Leslie
17455113	10.1080/00498250601188808	2007	Nakagomi-Hagihara
17470528	10.1124/dmd.106.012930	2007	Noé
17473959	10.1007/s11095-007-9254-z	2007	Koepsell
17495125	10.1124/jpet.107.123554	2007	Chen
17496208	10.1124/dmd.106.013615	2007	Treiber
17502342	10.1124/dmd.106.013912	2007	Mizuno
17504223	10.2174/138920007780655423	2007	Xia
17509530	10.1016/j.bbrc.2007.04.207	2007	Xie
17509534	10.1016/j.bcp.2007.04.010	2007	Tanihara
17532304	10.1016/j.bcp.2007.04.015	2007	Grigat
17585018	10.1124/jpet.107.125831	2007	Windass
17614008	10.1080/00498250701397705	2007	Umehara
17664327	10.1128/AAC.00671-07	2007	Tong
17674156	10.1007/s11095-007-9401-6	2007	Sato
17682070	10.1124/dmd.107.016352	2007	Mizuno
17700366	10.1097/FPC.0b013e3281c6d08e.	2007	Urban
17709369	10.1124/dmd.107.017723	2007	Shaik
17823233	10.1124/dmd.107.017459	2007	Yamada
17876861	10.1002/bdd.576	2007	Choi
17890094	10.1016/j.bmc.2007.07.024	2007	Müller
17901929	10.1007/s11095-007-9446-6	2008	Hinton
17962372	10.1124/dmd.107.017434	2008	Feng
17965255	10.1530/REP-06-0173	2007	Kobayashi
17971819	10.1038/sj.clpt.6100405	2008	Feng
17977516	10.1016/j.bcp.2007.09.015	2008	Rytting
17977676	10.1016/j.ijpharm.2007.09.022	2008	Hirano
18027988	10.2165/00003088-200746120-00004	2007	Eberl
18058507	10.1080/15257770701506426	2007	Kim
18083034	10.1016/j.bmc.2007.11.057	2008	Müller
18157518	10.1007/s00210-007-0219-x	2008	Nies
18180273	10.1124/dmd.107.018903	2008	Ishiguro
18216183	10.1124/mol.107.043117	2008	Cropp
18236139	10.1007/s11095-007-9492-0	2008	Deng
18314419	10.2337/db07-1515	2008	Bachmakov
18319568	10.1254/jphs.fp0070911	2008	Tachampa
18321482	10.1016/j.ejphar.2008.01.042	2008	Gui
18373647	10.1111/j.1742-7843.2007.00155.x	2008	Nakagawa
18397960	10.1136/ard.2007.086264	2008	Kis
18408565	10.1097/FPC.0b013e3282fb02a3	2008	Deng
18408886	10.1007/s00018-008-8065-7	2008	Todesco

18433974	10.1016/j.tox.2008.03.003	2008	Oosterhuis
18457386	10.1021/jm7015683	2008	Pedersen
18490433	10.1124/dmd.108.020826	2008	Jung
18559608	10.1158/1078-0432.CCR-07-4793	2008	Filipski
18617601	10.1124/dmd.108.021410	2008	Kitamura
18654741	10.1007/s11095-008-9678-0	2008	Li
18678495	10.1016/j.bmc.2008.07.034	2008	Pick
18686197	10.1080/00498250802334409	2008	Umehara
18707884	10.1016/j.bmcl.2008.07.127	2008	Leyers
18762717	10.2133/dmpk.23.293	2008	Ohnishi
18788725	10.1021/jm8003152	2008	Ahlin
18790787	10.1158/1535-7163.MCT-08-0539	2008	Okabe
18824524	10.1124/dmd.108.022418	2009	Knütter
18834354	10.1111/j.1742-7843.2008.00298.x	2008	Wang
18971316	10.1124/dmd.108.024083	2009	Ming
18974612	10.2133/dmpk.23.347	2008	Nakakariya
18981167	10.1124/dmd.108.023929	2009	Grigat
19002438	10.1007/s00210-008-0369-5	2009	Zolk
19041296	10.1016/j.cbi.2008.10.052	2009	Pochini
19076159	10.1111/j.1365-2125.2008.03303.x	2009	Suzuki
19141712	10.1124/jpet.108.146225	2009	Minuesa
19164462	10.1124/jpet.108.147918	2009	Tsuda
19170519	10.1021/jm8013822	2009	Kühnle
19172157	10.1038/tpj.2008.19	2009	Chen
19220985	10.1124/dmd.108.025015	2009	Kano
19250834	10.1016/j.bmc.2009.01.072	2009	Klinkhammer
19251820	10.1124/dmd.108.023762	2009	Zolk
19252303	10.1248/bpb.32.497	2009	Jani
19282394	10.1124/dmd.109.026880	2009	Duan
19357179	10.1152/ajprenal.90754.2008	2009	Lee
19402665	10.1021/jm900253g	2009	Gannon 2nd
19427586	10.1016/j.phrs.2009.03.004	2009	Gui
19427995	10.1016/j.bcp.2009.04.002	2009	Tiwari
19437106	10.1007/s11095-009-9905-3	2009	Diao
19439487	10.1124/dmd.109.026922	2009	Ogihara
19439489	10.1124/dmd.108.026294	2009	Umehara
19493273	10.1111/j.1349-7006.2009.01213.x	2009	Noguchi
19520775	10.1124/dmd.109.027359	2009	Iwai
19520776	10.1124/dmd.108.024778	2009	Kis
19540211	10.1016/j.bcp.2009.06.014	2009	Tanihara
19591196	10.1002/hep.23103	2009	Nies
19616083	10.1016/j.toxlet.2009.07.011	2009	Yang

19631272	10.1016/j.phrs.2009.07.002	2010	Jutabha
19643159	10.1016/j.tox.2009.07.014	2009	Bakhiya
19684012	10.1167/iovs.09-4080	2010	Tachikawa
19721074	10.1128/AAC.01541-08	2009	Hoffmann
19725578	10.1021/jm900194w	2009	Wong
19741038	10.1124/dmd.109.028522	2010	Lan
19779132	10.1124/jpet.109.159756	2010	Shirasaka
19789362	10.1124/jpet.109.159822	2010	Anderson
19833842	10.1124/dmd.109.028142	2010	Minematsu
19833843	10.1124/dmd.109.029454	2010	Ose
19934028	10.1177/0091270009337514	2010	Werner
19936896	10.1007/s10928-009-9139-3	2009	Poirier
20020740	10.1021/mp900206j	2010	Nakamura
20025245	10.1021/mp900174z	2010	Cook
20053795	10.1152/ajprenal.00431.2009	2010	Meyer zu Schwabedissen
20060011	10.1016/j.taap.2009.12.035	2010	Tsirulnikov
20067523	10.1111/j.1742-4658.2009.07528.x	2010	Geissler
20075570	10.2169/internalmedicine.49.2597	2010	Uetake
20102298	10.3109/00498250903509375	2010	Annaert
20159988	10.1124/dmd.109.030791	2010	Pollex
20220747	10.1038/clpt.2009.266	2010	Oswald
20345483	10.1111/j.1349-7006.2010.01539.x	2010	Kawahara
20360303	10.1124/dmd.109.031526	2010	Han
20406852	10.1124/dmd.110.032862	2010	Leonhardt
20448812	10.2174/1875397301004010001	2010	Gui
20507927	10.1124/jpet.110.166314	2010	Kis
20520284	10.1097/JCP.0b013e3181d2ef42	2010	Lee
20567254	10.1038/tpj.2010.54	2011	Ahlin
20574995	10.1002/jps.22262	2011	Jani
20599802	10.1016/j.bcp.2010.06.025	2010	Hemauer
20639259	10.1093/toxsci/kfq219	2010	Yang
20670210	10.2174/187231210792928279	2010	Yang
20831193	10.1021/mp100226q	2010	Diao
20921968	10.1038/tpj.2010.75	2012	Tzvetkov
21124314	10.1038/clpt.2010.232	2011	Imamura
21189339	10.1128/AAC.01527-10	2011	Zembruski
21212936	10.1007/s00210-010-0590-x	2011	Wenge
21217360	10.1097/FPC.0b013e3283402efb	2010	Siccardi
21252289	10.1158/1535-7163.MCT-10-0731	2011	Minematsu
21341745	10.1021/jm101421d	2011	Broccatelli
21351087	10.1002/ijc.26000	2012	Tang

21354800	10.1016/j.bmc.2010.12.043	2011	Pick
21372390	10.1248/bpb.34.389	2011	Yamaguchi
21430235	10.1124/dmd.110.034991	2011	König
21437911	10.1002/jps.22548	2011	Yasujima
21440623	10.1016/j.ejps.2011.03.006	2011	Elsby
21456052	10.1002/bdd.749	2011	Choi
21505084	10.1177/0091270011400414	2012	Cutler
21518836	10.1128/AAC.01835-10	2011	Müller
21543413	10.1152/ajprenal.00735.2010	2011	Astorga
21562485	10.1038/clpt.2011.56	2011	Tzvetkov
21570282	10.1016/j.bmcl.2011.04.094	2011	Puentes
21599003	10.1021/jm2001629	2011	Kido
21628496	10.1124/dmd.111.039370	2011	Li
21641380	10.1016/j.ajpath.2011.02.020	2011	Grube
21646436	10.1124/dmd.110.035865	2011	Cheng
21719246	10.1016/j.biopha.2011.04.031	2011	Svoboda
21726197	10.1042/BJ20110544	2011	Pochini
21779389	10.1371/journal.pone.0022163	2011	Nies
21783720	10.1016/j.etap.2006.04.002	2006	Pivčević
21861202	10.1007/s11095-011-0564-9	2011	Karlgren
21864659	10.1016/j.ijpharm.2011.08.009	2011	Vadlapatla
21865262	10.1152/ajprenal.00169.2011	2011	Kaufhold
22004608	10.1016/j.ejphar.2011.10.004	2011	Wright
22021325	10.1124/mol.111.074823	2012	Visentin
22072731	10.1124/jpet.111.184986	2012	Ito
22112208	10.1021/jm201305y	2012	Pellicani
22206629	10.1016/j.bbamem.2011.12.014	2012	Pochini
22223530	10.1158/1078-0432.CCR-11-2503	2012	Ciarimboli
22273603	10.1016/j.peptides.2012.01.007	2012	Zhu
22300367	10.1111/j.1476-5381.2012.01887.x	2012	Kusuhara
22355035	10.1124/dmd.111.041616	2012	Wang
22389472	10.1158/1535-7163.MCT-11-0980	2012	Hu
22415520	10.1002/bdd.1783	2012	Choi
22419765	10.1124/jpet.112.191577	2012	Astorga
22452412	10.1021/jm201705f	2012	Jabeen
22534868	10.1038/clpt.2011.351	2012	Yoshida
22541068	10.1021/jm300212s	2012	Karlgren
22711709	10.1158/1078-0432.CCR-12-0761	2012	M de Graan
22771883	10.1148/radiol.12112061	2012	Nassif
22775210	10.1111/j.2042-7158.2012.01498.x	2012	Choi
22822035	10.1124/dmd.112.047183	2012	Gao
22847220	10.1002/bdd.1801	2012	Nagasaka

22902721	10.1016/j.bcp.2012.07.032	2012	Kunze
22913740	10.1111/j.1472-8206.2012.01071.x	2014	Jouan
22973893	10.1021/mp300365t	2012	Duan
23020787	10.3109/00498254.2012.720740	2013	Jin
23073734	10.1124/dmd.112.048470	2013	Ishiguro
23121773	10.1111/bcpt.12031	2013	Klatt
23132334	10.1124/dmd.112.048918	2013	Reese
23169446	10.1002/jbt.21458	2013	Bircsak
23241029	10.1021/jm301302s	2013	Wittwer
23248200	10.1124/dmd.112.049023	2013	Steeg
23284953	10.1371/journal.pone.0052247	2012	Schmidt-Lauber
23293300	10.1124/dmd.112.049668	2013	Chu
23340295	10.1158/1078-0432.CCR-12-3306	2013	Zimmerman
23402418	10.1021/mp300562q	2013	Wei
23428312	10.1016/j.ijantimicag.2013.01.004	2013	Weiss
23582785	10.1016/j.biopha.2013.02.003	2013	Sogame
23652407	10.1097/FPC.0b013e3283620c3b	2013	Tamraz
23709117	10.1124/jpet.113.203257	2013	Belzer
23729661	10.1124/dmd.113.051193	2013	Reyner
23763587	10.1021/mp400113d	2013	Thévenod
23770354	10.1016/j.tox.2013.06.001	2013	Chiba
23794501	10.1002/jps.23653	2013	Shirasaka
23831208	10.1016/j.tox.2013.06.009	2013	Tu
23835420	10.1016/j.bcp.2013.06.019	2013	Tzvetkov
23856525	10.1016/j.tox.2013.07.004	2013	Mandíková
23886114	10.3109/00498254.2013.820006	2013	Chiou
23956101	10.1093/toxsci/kft176	2013	Morgan
23967177	10.1371/journal.pone.0071266	2013	Bhullar
23984907	10.1021/jm400966v	2013	Hendrickx
24001450	10.1016/j.taap.2013.08.024	2013	Li
24014645	10.1093/toxsci/kft198	2013	Dankers
24026623	10.1124/dmd.113.053215	2013	Shen
24036158	10.1016/j.trsl.2013.08.003	2013	El-Sheikh
24042472	10.2133/dmpk.dmpk-13-rg-042	2014	Kurata
24122511	10.1002/jps.23737	2013	Kikuchi
24150606	10.2337/db13-1005	2014	Meyer zu Schwabedissen
24184213	10.1016/j.bmc.2013.10.007	2013	Juvale
24246570	10.1016/j.pnpbp.2013.11.005	2014	Sun
24266811	10.1111/jnc.12618	2014	Brown
24295974	10.1128/AAC.02049-13	2014	Rizk
24338900	10.1002/jps.23805	2014	Huo

24375866	10.1002/etc.2493	2014	Georgantzopoulou
24398510	10.1038/bjc.2013.811	2014	Hu
24440960	10.1124/dmd.113.054767	2014	Takeuchi
24646860	10.1038/ki.2014.66	2014	Lepist
24688079	10.1124/dmd.113.055095	2014	Boxberger
24692216	10.1124/dmd.113.055194	2014	Wang
24754247	10.1021/mp400699p	2014	Gozalpour
24807167	10.1515/dmdi-2014-0014	2014	Khurana
24867984	10.1128/AAC.02724-14	2014	Furihata
24961373	10.1073/pnas.1314939111	2014	Chen
25155823	10.1208/s12248-014-9649-9	2014	Dos Santos Pereira
25165131	10.1124/dmd.114.059097	2014	Johnston
25268938	10.1016/j.toxlet.2014.09.020	2015	Mazur
25310383	10.1021/jm500862r	2014	Fu
25414411	10.1124/dmd.114.059105	2015	Izumi
25466967	10.1021/mp500532x	2015	Bexten
25477469	10.1152/ajprenal.00467.2014	2015	Hagos
25521244	10.3390/pharmaceutics6040632	2014	Weiss
25587128	10.1124/dmd.114.062174	2015	Lee
25739790	10.1248/bpb.b14-00740	2015	Ohya
25855895	10.1021/acs.jmedchem.5b00188	2015	Köhler
25914645	10.3389/fphar.2015.00078	2015	Moss
26037524	10.1002/bdd.1961	2015	Chapy
26327616	10.1371/journal.pone.0136451	2015	Hacker
26330539	10.1124/jpet.115.227546	2015	Visentin
26340566	10.3109/00498254.2015.1081993	2016	Reese
26374172	10.1124/dmd.115.066175	2015	Yin
26659468	10.1002/cpt.317	2016	Matthaei
26668209	10.1124/dmd.115.067744	2016	Vermeer
26696140	10.1021/acs.molpharmaceut.5b00664	2016	Izumi
26700956	10.1124/dmd.115.066795	2016	Elsby
26702643	10.1021/acs.molpharmaceut.5b00733	2016	Misaka
26774038	10.1016/j.ejmech.2015.12.010	2016	Obreque-Balboa
26927160	10.3390/pharmaceutics8010005	2016	Weiss
26951201	10.1007/s40261-016-0386-y	2016	Shen
26976869	10.1128/AAC.02765-15	2016	Parvez
26995013	10.1016/j.taap.2016.03.008	2016	Miyata
27152359	10.1126/sciadv.1600001	2016	Nicklisch
27178732	10.1016/j.phrs.2016.05.012	2016	Sauzay
27271370	10.1124/dmd.115.068163	2016	Lechner
27321165	10.1111/jcpt.12408	2016	Ayalasomayajula
27324234	10.1111/bph.13537	2016	Li

27350110	10.1111/jphp.12574	2016	Futatsugi
27393949	10.1016/j.ejmech.2016.06.039	2016	Baiceanu
27418674	10.1124/mol.116.105056	2016	Martínez-Guerrero
27458210	10.1128/AAC.01335-16	2016	Gupta
27503646	10.1128/AAC.00986-16	2016	Yang
27504015	10.1124/dmd.116.069807	2016	Hubeny
27550354	10.1128/AAC.01151-16	2016	Parvez
27645247	10.1128/AAC.01471-16	2016	Te Brake
27676604	10.1515/hsz-2016-0236	2016	Chen
27684210	10.1080/00498254.2016.1241449	2016	McCormick
27737931	10.1124/dmd.116.072397	2017	Johnson
27989701	10.1016/j.tiv.2016.12.009	2016	Bruyere
28093031	10.1080/00498254.2017.1282647	2018	Shams
28099443	10.1371/journal.pone.0169480	2017	Chedik
28209616	10.1158/0008-5472.CAN-16-2548	2017	Drenberg
28223391	10.1128/AAC.02392-16	2017	Parvez
28230985	10.1021/acs.jmedchem.6b01317	2017	Chen
28281384	10.1080/00498254.2017.1295171	2018	Hanna
28283500	10.1124/dmd.116.074609	2017	Zhang
28320730	10.1124/dmd.116.073932	2017	Ellens
28373111	10.1016/j.xphs.2017.03.022	2017	Pahwa
28428365	10.1124/dmd.116.074708	2017	Wagner
28455521	10.1038/s41598-017-01438-4	2017	Harrach
28456731	10.1016/j.xphs.2017.04.044	2017	Tsuruya
28499878	10.1016/j.xphs.2017.05.006	2017	Takano
28535976	10.1016/j.xphs.2017.05.011	2017	Wang
28562037	10.1021/acs.jcim.6b00720	2017	Khuri
28615288	10.1124/jpet.117.242552	2017	Severance
28630284	10.1124/jpet.117.241406	2017	Burckhardt
28737453	10.1080/00498254.2017.1357088	2018	Gnerre
28749581	10.1111/cts.12486	2017	Posada
28801980	10.1002/bdd.2091	2017	Narumi
28841513	10.1016/j.ejmech.2017.08.020	2017	Krapf
28921988	10.1021/acs.molpharmaceut.7b00563	2017	Sjöstedt
28926871	10.1002/bdd.2104	2017	Li
28971610	10.1002/prp2.357	2017	Elsby
29061131	10.1186/s12936-017-2062-y	2017	van der Velden
29236753	10.1371/journal.pone.0189521	2017	Meyer
29277663	10.1016/j.ijpharm.2017.12.037	2018	Wen
29307856	10.1016/j.ejps.2018.01.002	2018	Lu
29448871	10.1080/00498254.2018.1442030	2019	Chedik
29498478	10.1111/bcpt.12992	2018	Chen

29748863	10.1007/s00228-018-2477-6	2018	Pei
29882324	10.1002/cpt.1128	2019	Matthaei
29884691	10.1124/mol.117.111443	2018	Sandoval
29888219	10.3389/fchem.2018.00180	2018	Hucke
29915999	10.1007/s11095-018-2445-y	2018	Xiao
30012768	10.1128/AAC.00512-18	2018	Parvez
30114293	10.1371/journal.pone.0202706	2018	Ceckova
30361780	10.1007/s11095-018-2526-y	2018	Zhu
30449032	10.1002/cpt.1296	2019	Chappell
30656943	10.1021/acs.molpharmaceut.8b00779	2019	Deutsch
30706983	10.1111/cts.12623	2019	McFeely
30745395	10.1128/AAC.02492-18	2019	Bleasby
30891606	10.1093/jac/dkz101	2019	Chan
31034908	10.1016/j.xphs.2019.04.023	2019	Kikuchi
31047942	10.1016/j.xphs.2019.04.019	2019	Farasyn
31322418	10.1152/ajprenal.00141.2019	2019	Yin
31341258	10.1038/s41401-019-0283-z	2020	Liao
31371478	10.1124/jpet.119.259341	2019	Taniguchi
31437515	10.1016/j.toxlet.2019.08.013	2019	Jiang
31506301	10.1124/dmd.119.086918	2019	Cheong
31514980	10.1016/S1875-5364(19)30070-6	2019	Chen
31542894	10.1002/jcph.1523	2020	Abebe
31571146	10.1007/s13318-019-00577-5	2019	Zurth
31834804	10.1021/acs.molpharmaceut.9b00897	2020	Unger
31990564	10.1021/acs.molpharmaceut.9b00658	2020	Zou
32129697	10.1080/00498254.2020.1737759	2020	Liao
32334013	10.1016/j.lfs.2020.117696	2020	Sarkar
32344570	10.3390/pharmaceutics12040390	2020	Antonescu
32421406	10.1080/00498254.2020.1771473	2020	Guéniche
32571913	10.1073/pnas.1920483117	2020	Zou
32661908	10.1007/s13318-020-00634-4	2020	Bajraktari-Sylejmani
32910949	10.1016/j.xphs.2020.09.004	2021	Zou
33759449	10.1002/cpt.2236	2021	Yee
35089508	10.1007/s11095-022-03171-8	2022	Chen

Table S1-2: Cumulative statistics for all transporter-chemical interaction quantitative kinetic values in UCSF-FDA TransPortal + UCSD/UCD-NIEHS TICBase

Statistic	**K_i**	**IC_{50}**	**K_m**	
Minimum	0.0067	0.00002	0.0076	
5th percentile	0.38	0.4	0.532	
20th percentile	2.5	2.1	3.58	
40th percentile	12.1	6.99	14.1	
Median	22.7	12	23.5	
60th percentile	42	20.1	44.5	
80th percentile	230	61.9	257	
95th percentile	2520	343	1580	
Maximum	8620	14000	8590	
Mean	370.55	119.98	299.66	**Sums**
Hard Values	566	2780	742	**4004**
Bounded Values	7	2211	4	**2222**
Approx. Values	0	12	0	**12**
ND Values	0	16	42	**58**
Sums	**573**	**5019**	**788**	**6296**

Interactions with both a K_i and an IC_{50} value: 84

Table S1-3: A ranked list of transporters in UCSF-FDA Transportal by total interactions, inhibitory interaction, and substrate interactions.

Rank	**Transporter**	**Total Interactions**	**Transporter**	**Inhibitors**	**Transporter**	**Substrates**
1	ABCB1	549	ABCB1	478	ABCB1	71
2	SLC22A2	421	SLC22A2	353	SLC22A2	68
3	SLC22A1	366	SLC22A1	311	SLC22A1	55
4	SLCO1B1	293	SLCO1B1	219	SLCO1B1	74
5	SLC22A8	262	SLC22A8	221	SLC22A8	41
6	SLC22A6	235	SLC22A6	197	SLC22A6	38
7	SLC47A1	180	SLC47A1	158	SLC47A1	22
8	ABCB11	178	ABCB11	172	ABCB11	6
9	ABCG2	166	ABCG2	119	ABCG2	47
10	SLCO1B3	150	SLCO1B3	96	SLCO1B3	54
11	ABCC4	145	ABCC4	126	ABCC4	19
12	SLC47A2	129	SLC47A2	108	SLC47A2	21
13	SLCO2B1	127	SLCO2B1	108	SLCO2B1	19
14	SLC22A5	123	SLC22A5	102	SLC22A5	21
15	ABCC2	103	ABCC2	79	ABCC2	24
16	SLC22A3	62	SLC22A3	55	SLC22A3	7
17	ABCC3	62	ABCC3	46	ABCC3	16

18	SLC22A11	52	SLC22A11	39	SLC22A11	13
19	mouse Abcb1a	47	mouse Abcb1a	47	mouse Abcb1a	0
20	SLC22A4	45	SLC22A4	32	SLC22A4	13
21	ABCC5	41	ABCC5	31	ABCC5	10
22	SLC22A7	38	SLC22A7	27	SLC22A7	11
23	ABCC1	36	ABCC1	18	ABCC1	18
24	SLC15A1	22	SLC15A1	10	SLC15A1	12
25	SLC10A1	21	SLC10A1	12	SLC10A1	9
26	grivet ABCB1	21	grivet ABCB1	21	grivet ABCB1	0
27	SLCO1A2	17	SLCO1A2	7	SLCO1A2	10
28	mouse Abcb1b	14	mouse Abcb1b	14	mouse Abcb1b	0
29	SLC15A2	14	SLC15A2	8	SLC15A2	6
30	SLC10A2	14	SLC10A2	7	SLC10A2	7
31	rat Slc22a1	12	rat Slc22a1	8	rat Slc22a1	4
32	rat Slc22a2	12	rat Slc22a2	7	rat Slc22a2	5
33	SLC22A12	11	SLC22A12	8	SLC22A12	3
34	rat Slc22a3	10	rat Slc22a3	5	rat Slc22a3	5
35	rat Slco1a1	7	rat Slco1a1	5	rat Slco1a1	2
36	rat Slc47a1	5	rat Slc47a1	4	rat Slc47a1	1
37	mouse Slco1a4	3	mouse Slco1a4	0	mouse Slco1a4	3
38	OSTalpha	2	OSTalpha	1	OSTalpha	1
39	OSTbeta	2	OSTbeta	1	OSTbeta	1
40	ABCC6	2	ABCC6	0	ABCC6	2
41	rat Slc7a10	1	rat Slc7a10	1	rat Slc7a10	0
42	SLC7A10	1	SLC7A10	1	SLC7A10	0
43	mouse Abcc2	1	mouse Abcc2	0	mouse Abcc2	1
44	rat Slc22a6	1	rat Slc22a6	0	rat Slc22a6	1
45	rat Slc22a8	1	rat Slc22a8	0	rat Slc22a8	1
	SUM:	**4004**	**SUM:**	**3262**	**SUM:**	**742**

Table S1-4: FDA model inhibitors (https://www.fda.gov/drugs/drug-interactions-labeling/drug-development-and-drug-interactions-table-substrates-inhibitors-and-inducers) and the ranges of their IC_{50} values in the database. NA = "Not Applicable".

Transporter	**FDA Model Inhibitor**	**IC_{50} Range (µM)**
ABCB1	Cyclosporine	.1 - 9.3
	Elacridar (gf120918)	.027 - .2

	Ketoconazole	1.2 - 53.4
	Quinidine	1 - 340
	Reserpine	0.5 - 6.1
	Ritonavir	3.8 - 28.2
	Tacrolimus	NA
	Valspodar (PSC833)	.11 - 3.2
	Verapamil	0.2 - 446.5
	Zosuquidar	0.024 - 0.1
ABCG2	Elacridar	0.31
	Fumetrimorgan C	0.25 - 0.47
	Ko134	NA
	Ko143	0.01 - 0.4
	Novobiocin	1.4
	Sulfasalazine	0.61 - 2.9
SLCO1B1	Cyclosporine	0.05 - 3.5
	Estradiol-17β-glucuronide	NA
	Estrone-3-sulfate	0.06 - 0.79
	Rifampicin	0.24 - 120
	Rifamycin SV	0.23
SLCO1B3	Cyclosporine	.06 - 1.3
	Estradiol-17β-glucuronide	NA
	Estrone-3-sulfate	7.1
	Rifampicin	.8 - 2.6
	Rifamycin SV	NA
SLC22A6	Benzylpenicillin	NA
	Probenecid	3.9 - 17
SLC22A8	Benzylpenicillin	137
	Probenecid	1.9 - 24.6
SLC47A1	Cimetidine	1.2 - 5.7
	Pyrimethamine	0.07
SLC47A2	Cimetidine	5.47 - 39
	Pyrimethamine	.01 - .059
SLC22A2	Cimetidine	25.4 - 373

Table S1-5: FDA model substrates (https://www.fda.gov/drugs/drug-interactions-labeling/drug-development-and-drug-interactions-table-substrates-inhibitors-and-inducers) and the ranges of their K_m values in the database. NA = "Not Applicable".

Transporter	**FDA Model Substrate**	**K_m Range (µM)**
ABCB1	Digoxin	73 - 181
	Fexofenadine	150
	Loperamide	11.4
	Quinidine	18.2

	Talinolol	NA
	Vinblastine	.8 - 253
ABCBG2	2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP)	NA
	Coumestrol	NA
	Daidzein	NA
	Dantrolene	NA
	Estrone-3-sulfate	6.8 - 16.6
	Genistein	NA
	Prazosin	NA
	Sulfasalazine	.4 - .7
SLCO1B1	CCK-8	NA
	Estradiol-17β-glucuronide	2.5 - 10
	Estrone-3-sulfate	.0872 - 12.5
	Pitavastatin	.429 - 6.7
	Pravastatin	27 - 85.7
	Rosuvastatin	.8 - 9.31
	Telmisartan	NA
SLCO1B3	CCK-8	3.82 - 16.5
	Estradiol-17β-glucuronide	15.8 - 24.6
	Estrone-3-sulfate	58
	Pitavastatin	3.25 - 3.8
	Pravastatin	NA
	Rosuvastatin	9.8 - 14.2
	Telmisartan	0.81
SLC22A6	Adefovir	23.8 - 30
	Para-aminohippurate	3.9 - 28
	Cidofovir	30 - 58
	Tenofovir	33.8
SLC22A8	Benzylpenicillin	NA
	Estrone-3-sulfate	2.18 - 21.2
	Methotrexate	10.9 - 76.6
	Pravastatin	27.2
SLC47A1	Metformin	202 - 780
	1-methyl-4-phenylpyridinium (MPP+)	NA
	Tetraethylammonium	220 - 380
SLC47A2	Metformin	1050 - 1980

	1-methyl-4-phenylpyridinium (MPP+)	NA
	Tetraethylammonium	760 - 830
SLC22A2	Metformin	235 - 3356
	1-methyl-4-phenylpyridinium (MPP+)	4.6 - 5.1
	Tetraethylammonium	33.8 - 76

Table S1-6: A description of cell/membrane environments used in assays use to produce 97.5% of interaction data in the updated database.

Cell/membrane Environment	Total Count	Description/Origin
HEK293	1530	**H**uman **E**mbryonic **K**idney cells
CHO	451	**C**hinese **H**amster **O**vary cells
Sf9	433	***S**podoptera **f**rugiperda* (fall armyworm) ovary cells
MDCK	302	**M**adin-**D**arby **C**anine **K**idney cells
S2	220	**S**chneider **2** cells – derived from *Drosphilia melanogaster* embryos
Caco-2	153	"**Ca**ncer **Co**li" – human colorectal adenocarcinoma cells
Oocytes	150	Immature ovum cells of *Xenopus laevis* (African clawed frog)
LLC-PK1	145	**L**ewis **L**ung **C**ancer **P**orcine **K**idney **C**ells – taken from porcine kidney epithelial tissue
NIH/3T3	132	Embryonic mouse fibroblast cells
HeLa	122	"**He**nrietta **La**cks" – cervical cancer cells
COS	51	**C**V-1 **O**rigin **S**V40 cells – fibroblast like cells from *Chlorocebus aethiops* (monkey) kidney
PDS-protein	31	**P**urified, **D**etergent-**S**olubilized protein
A2780	26	Human ovarian cancer cells
CEM	22	Human acute lymphoblastic leukemia cells
F4-6P	21	Murine leukemia cells derived from the Friend virus
BeWo	16	Human placental choriocarcinoma cells
HRPE	13	**H**uman **R**etinal **P**igment **E**pithelial cells
B16/F10	11	Murine melanoma cells
K562	9	Human myelogenous leukemia cells
V79	9	Genetically transformed lung cells from Chinese hamster
K cell	6	Proprietary vesicular transporter assay kit under Predivez™ trademark
P388	6	Murine lymphoma cells
High Five	5	*Trichoplusia ni* ovary cells (insect cells)
OK	5	**O**possum **K**idney cells from proximal tubule
L5178	4	Murine leukemia cells
MCF-7	4	(**M**ichigan **C**ancer **F**oundation) Human breast cancer cells

Mes-SA	3	Human epithelial uterus cancer
KCL22	3	Human chronic myeloid leukemia cells
HepG2	3	Human **Hep**atocellular carcinoma cells
HL-60	2	**H**uman **L**eukemia (acute myeloid) cells
Huh-7	2	**Hu**man **h**epatocellular carcinoma, highly susceptible to hepatitis C
JAR	2	Human placental choriocarcinoma cells
HCT-15	2	Colon adenocarcinoma cells
PC-6	2	Human lung small cell carcinoma cells

Appendix 2: UC Transportal – Additional Design Rationale and Areas for Improvement

1. Rationale for Interaction Type Field Values

When deciding how to define the parameters that would describe transporter-chemical interactions, we identified three main ways to classify interactions: (1) the site of the interaction on the transporter, (2) the effect of the interaction on the rate of transport of a reporter, and (3) whether the interacting chemical is transported. These would eventually become the three data fields that make up the Interaction Type category. For each of these classifications, the potential values were identified, shown in Table S2-1.

Table S2-1: Three independent ways to classify transporter-chemical interactions and their possible values.

	Interaction Site (1)	**Effect on Reporter (2)**	**Transported? (3)**
1	Allosteric	Speed up	Yes
2	Orthosteric	Slow down	No
3	No interaction	No change	

The number of possible combinations by choosing one value from each classification is 3 x 3 x 2 = 18. Designations for transporter-chemical interactions fitting all 18 combinations were considered but did not produce a system to our liking (Table 2). The designations, while accurate, did not have consistent names across similar interaction types, were not concise, and certain combinations were "nonsense" (#13, 15, and 17, Table 2).

Table S2-2: Eighteen technical classifications of transporter-chemical interaction. The position each "Combination" digit corresponds to the columns of Table S2-1, and values of the digits correspond to the rows. For example, "121" classifies an interaction in which the chemical interacts allosterically, slows down the transport of a reporter, and is itself transported by the transporter protein.

#	**Combination**	**Technical Classification**
1	111	Allosteric stimulator and substrate
2	112	Allosteric stimulator
3	121	Allosteric inhibitor and substrate
4	122	Noncompetitive inhibitor
5	131	Weak interactor and substrate
6	132	Allosteric modulator

7	211	Orthosteric stimulator and substrate
8	212	Orthosteric stimulator
9	221	Orthosteric inhibitor and substrate
10	222	Orthosteric inhibitor
11	231	Noncompetitive substrate
12	232	Orthosteric modulator
13	311	Nonsense
14	312	Inducer
15	321	Nonsense
16	322	Repressor
17	331	Nonsense
18	332	"Weak" interaction

Use of terms such as “modulator” and “activator” were considered but later discarded because of the potential ambiguity in their meaning across different fields of study (such as genetics, enzymology, and synthetic chemistry). Terminology regarding competitive binding in enzymology was also considered for use, however the definitions of these terms in enzymology (see Table S2-3) seemed too specific for our application. Ultimately, we decided to break up the description the interaction type into the three categories and use simple nomenclature for each of them. “Competitive” and “Noncompetitive” are used to describe an interaction at the protein binding site or elsewhere (synonymous with allosteric and Orthosteric) but their definitions do not correspond with those in enzymology. “Inhibitor”, “Stimulator”, and “Weak Interactor” are used to define the core interaction type, the database being branded as the “Kinetic Transportal”, and “Substrate” and “Nonsubstrate” are used to indicate the transport of the interacting chemical.

Table S2-3: Different types of competitive binding between chemicals and proteins as used in enzymology.

Competition Type	**Description as Used in Enzymology**
Competitive	Binds only if the substrate is not bound
Non-competitive	Binds equally well whether substrate is there or not
Mixed	Binds preferentially to the free enzyme or the ES complex
Uncompetitive	Binds only to the enzyme-substrate complex

The process of defining these parameters helped to clarify the scope of interactions Kinetic Transportal would include. It became clear that direct physical interactions between chemicals and transporters were to be the main focus of the project, and that interactions mediated through altered gene expression or other biological signals are of another category. Thus, the Inducer and Repressor designations (Table S2-2), while valid, have been saved until the envisioned Gene Expression Transportal can be created.

2. Level of Control over Data Field Values

Additionally, due to the importance of correctly defining acceptable values in the Interaction Type category fields, it is the only category for which all values were fully defined before any data curation occurred and its values are completely controlled (i.e., only the predefined values are accepted). Values in most other data fields are also controlled, but to varying degrees. The development of rules concerning acceptable values in other data fields occurred gradually as the database was populated with new data from an array of primary literature sources and adjustments to incorporate them were made. In this manner, many details of the database design are based on precedent and may continue to be adjusted in the future. Such rules up to the present have been described in detail in the curation instructions for Kinetic Transportal.

Data fields that have metadata attached to them may be said to be "semi-controlled". The metadata is incredibly helpful for defining values accurately and consistently across time for fields which are projected to add new values as the database expands. Not all fields need this, however, because the definitions of their values are clear (i.e., PubMed ID's, chemical concentrations), hence some fields are "uncontrolled". The "level of control" exerted on a field can be defined as how much restriction there is on allowing new values into the data field. The Interaction Type category fields and MXR/MDR field have completely predefined values and are therefore

"completely controlled" (this should be taken with a grain of salt, however). The Assay Type and Expression fields are tightly controlled because new values should be added with hesitancy and their existing values cover most scenarios that could be encountered. The Primary Parameter category fields are loosely controlled since new values are expected regularly, and the metadata keeps the values well-defined.

3. Table of Primary Parameter Reference Databases

Table S2-4: Reference databases for transporter proteins, chemicals, and organisms.

Proteins	**Database for Reference**	**Description**
NCBI Gene	https://www.ncbi.nlm.nih.gov/gene	The primary reference for all protein identification
MycoCosm	https://mycocosm.jgi.doe.gov/mycocosm/home	A database of fungal genes and proteins, including a specific category for transporters
Chemicals	**Database for Reference**	**Description**
PubChem	https://pubchem.ncbi.nlm.nih.gov/	The primary reference for all chemical identification
CompTox	https://comptox.epa.gov/dashboard/	An EPA chemical database, has a subpage specifically for alternative identifiers
DrugBank	https://go.drugbank.com/	A database focusing on human drugs / pharmaceuticals
Wikipedia	https://en.wikipedia.org/	Good for seeing a collection of disparate data, often picks up what major databases miss, check for references
PPDB	http://sitem.herts.ac.uk/aeru/ppdb/en/index.htm	A database of pesticide compounds by the University of Hertfordshire in England, is quite extensive
EU Pesticides DB	https://ec.europa.eu/food/plants/pesticides/eu-pesticides-database_en	A pesticide database by the EU, focuses on MRLs (maximus residue levels) and specific commodities
Organisms	**Database for Reference**	**Description**
NCBI Taxonomy	https://www.ncbi.nlm.nih.gov/taxonomy	"The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet."
ITIS	https://www.itis.gov/	Integrated Taxonomic Information System

4. Class Searches and Chemical Use Classifications

Searching the databases by transporter, chemical and organism classifications is a feature that will be implemented in the future. Classifying transporter proteins by superfamily or more specific groupings is straightforward and classifying organisms by their accepted taxonomic grouping (mammal, fish, plant etc.) is also usually straightforward. There is no clearly accepted best method of classifying chemicals, however, and the number of ways they can be grouped is practically without limit. Therefore, we settled on giving each chemical one of a manageable number of "intended use" or "class" designations: Nanomaterial, Inorganic, Pharmaceutical, Food/herb/supplement, Pesticide, Surfactant, Flame retardant, Endogenous (human), Biotoxin, Byproduct/metabolite, Scientific use, Plastics and plasticizers, or Other natural product. These classifications are still under development and may be subject to change.

The appropriate category can be determined by searching for information about the chemical in databases listed in the "*Primary Parameter References*" spreadsheet (section 3). Each classification will be given a detailed definition. Some chemicals may plausibly fall into more than one classification. In such a case, a hierarchy of which classification to choose will be referred to (currently ordered from first to last above). A rule of thumb to avoid this situation is to consider the main source of exposure or intended usage of the chemical. For example, paclitaxel was originally extracted from yew trees but is synthesized for pharmaceutical use, and exposure from yew tree is not significant to humans. So, paclitaxel should be considered a pharmaceutical and not a natural product even without a designation hierarchy.

5. Potential Areas of Improvement for Kinetic Transportal

The primary drawback of Kinetic Transportal coincides with one of its main strengths: since we have aimed to provide a very comprehensive characterization of transporter-chemical

interactions including a large amount of individual data (22 data fields) per interaction, the data curation process takes a substantial amount of time as well as some practice. It can be somewhat tedious to continuously look through an article for the necessary data, which is almost always distributed between the methods and materials, results, and figures, and are not always displayed or written clearly. This may make recruitment and retention of volunteers to continue adding information to the database (of which there is no shortage) a challenge.

It is also easy to overlook one or more steps in the curation process, such as standardizing the protein nomenclature in the primary data table or adding an entry to the assay metadata. Therefore, the Verification spreadsheet was added to the working curation document: it can catch errors that are easy to miss. For data fields with a very limited number of allowed entries (such as the Interaction Type fields, Expression field and Assay Type field), data filters or data validation can be used to easily spot errors. However, these are examples of technical errors which can be spotted and fixed. In contrast, errors which pose a more serious concern are when data from articles are misinterpreted or entered into the data fields incorrectly. The likelihood of this occurring is in part dependent on the expertise of the curator, but the specificity of the database's subject matter and certain aspects of the design may also increase the likelihood of erroneous or ambiguous data. Notes about this have been included in the curators' instructions and include topics such as identifying reporters and measurement methods for different assay types, making sure that IC_{50} and EC_{50} values are interpreted correctly, and most importantly, determining the interaction type correctly.

One notable point of potential misinterpretation in transporter efflux/uptake studies concerns the difference between pre-incubations and concurrent incubations of cells with reporter substances and chemicals of interest. Most assays incubate cells concurrently with both a reporter

molecule (typically a transporter substrate) and the chemical of interest, which may or may not modulate transporter activity. This method is standard for measuring direct interactions with transporters or direct modulation of transport activity. Pre-incubations, however, introduce the chemical of interest prior to the beginning of the assay, giving time for a change in gene expression to occur and introduce the possibility of gene induction or repression as a mechanism of transport activity modulation. Such data does not belong in the Kinetic Transportal database but could be erroneously considered an inhibition or stimulation interaction between the transporter and chemical of interest. Details about incubation times are easily overlooked, and without comment from the author an incorrect interpretation can easily be made by readers. One example of this is the study by Pessatti et al. of lead influence on MXR activity in *Perna perna* (brown mussel). Specimens were pre-exposed to lead nitrate and later gill tissue was collected for efflux experiments with rhodamine B. Rhodamine B efflux was stimulated in the tissues from individuals exposed to lead nitrate compared to the control group. The author explicitly stated that their interpretation of this data was that the stimulation was due to an increase in expression of the MXR mechanism, however without such commentary a less experienced curator could have easily interpreted this as direct stimulation of transporters involved in MXR activity by lead nitrate.

The design of the Kinetic Transportal also does not bring attention to interactions in which the reporter molecule may be of primary interest to users. The reporter field is “uncontrolled”, so its contents are not officially standardized like the Chemical field and are not in the metadata. A partial solution to this problem would be to add the option to include the reporter field as part of the chemical search when searching Kinetic Transportal. This is already possible because of filtering features in the data table plug-in, but it is not an intuitive way to search interaction data.

Standardizing the reporter values (adding some form of metadata for field control) could also help with future data analysis.

Additionally, the Chemical Concentration field can be somewhat misleading because we report the total range of concentrations over which a chemical was tested in an assay. This is necessary when reporting weak interactors but may be misleading for inhibitors because users may interpret the range as the concentrations at which inhibition activity was observed even though this distinction is made on the How to Use page. Many concentration ranges are estimates because they are only indicated by figures.

In the future, it could prove beneficial to include more kinds of quantitative data such as efflux ratios between cell types, apical-basolateral flux ratios and more. However, there are many possible parameters to extract and many of them would need context to provide meaningful information. Others, such as the turnover number (K_{cat}), are reported extremely rarely in the literature and could not justify having their own data field. An alternate design scheme that could potentially allow many kinds of quantitative data to be included in would be to change the quantitative data field category to consist of three fields: Numeric Value, Numeric Value Type, and Numeric Unit. A Numeric Value Type field could support any designation for quantitative kinetic measurements taken from the literature and a Numeric Unit field would mitigate any problems introduced by unusual units of measurement likely to be encountered. Such a scheme would not necessarily resolve the problem of context that some measurements need though, and it could increase the risk of including irrelevant or misleading quantitative data.

6. Potential Areas of Improvement for the UC Transportal Website

Although the official approval of data can only be performed by administrators and safeguards such as column counting are in place against mistaken uploads, additional security

measures could further protect the integrity of data in Kinetic Transportal and future databases. Checking field headers to match expected values is one way to ensure field have not been misplaced and would further encourage using the official templates for data upload. Currently, it is also possible to upload data with novel ID numbers and circumvent the quality assurance process. Disallowing novel ID numbers in the Import Approved Data function would not affect the normal upload process or ability to edit existing entries but would prevent this work-around. There is also no front-end function for deleting entries yet. Entries can be edited and overridden with blank data, but this is not a maintainable solution long-term and would use up ID numbers. Therefore, a deletion function could be an improvement as well.

Minor improvements also should be made to the web tables in tables in the future. Currently, the numeric fields of the search results table cannot be sorted correctly because the datatype is set to “string”. We did not set the datatype to “float” or “integer” because they set the number of decimal places to a constant value, and we need the decimal place to adjust to the data. The New Entries table could also use some adjustments to make it easier to scroll through and edit as administrators review. When changing the status of new entries to “Sent Back”, a larger input element for writing notes to the contributor will aid communication so it is clear what changes or additions are needed before it can receive approval.

Appendix 3: Plant Experiments – Making Solutions, Experiments with Norway Spruce and Alfalfa, and a Practical Guide to Growing Arabidopsis

1. Preparation and Handling Tips for Vacuole Isolation Solutions

The 1M mannitol solution will develop filament-like crystals over time if left at the temperature of the lab for a few days. These crystals are extremely difficult to redissolve. The maximum solubility of mannitol at 25°C is 216g/L, which is not much higher than 1M, equivalent to 186.17 g/L. It is quite possible that temperature fluctuations in the lab, particularly cooler temperatures at night, caused the mannitol to begin forming crystals. To solve this problem, simply store the mannitol solution in an incubator at 25°C or more. This permanently prevented the problem from occurring.

Dissolving Ficoll powder can be very difficult. Some of the initial protocols we used called for 30% w/v Ficoll stock solution to prepare the gradient layers, however, I changed this to 20% Ficoll to make the solution easier to make. Ficoll will form large sticky clumps that do not break up easily when added all at once to water. It can be necessary to break up these clumps with a metal spatula. Heating aids the dissolution process, however, Ficoll will degrade if heated too much (Robert et al. 2007). The easiest way we discovered to make Ficoll solution is to weigh out the Ficoll and place it in a closed microcentrifuge tube. Add the liquid components to the proper volume, then vortex briefly to remove clumps and hydrate the powder. Then, place the closed tube in a microtube shaker or on a microtiter plate and let it vibrate for 1-2 hours. This method worked very well for making 2mL solutions or smaller and may be scalable to work with 5mL or 15mL tubes as well.

Instead of purchasing neutral red solution, it can be prepared from dry powder very easily. For our experiments we used a 3.3g/L solution. The appropriate weight can be measured out and

dissolved in the .2M pH 7.5 phosphate buffer solution used in the vacuole isolation protocol. The solution will be dark red; caution should be because the powder can stain surfaces easily.

2. Testing Species other than Arabidopsis

Considering our original research question, to examine the mechanism of vacuolar transport for glycosylated monolignols, two additional plant species, Norway spruce (*Picea abies*) and alfalfa (*Medicago sativa*), were also explored for the potential to isolate vacuoles from their mesophyll tissue. Norway spruce was chosen because it is a conifer, meaning it is very distantly related to *Arabidopsis thaliana*, and it was the first conifer to have its genome sequenced (Nystedt et al. 2013). By comparing such distantly related species, it may be possible test hypotheses about the evolutionary divergence or conservation of vacuolar transport mechanisms. Alfalfa was chosen because it is an economically important forage crop of the family Fabaceae, or the legume family. Like *Arabidopsis*, a member of the Brassica family, legumes are angiosperms but the two are not closely related. A number of other economically important crops also occur in the legume family including many species of beans, peas, lentils, peanuts, clovers, and vetches. Alfalfa could also be an abundant source of leaf tissue for large scale experiments because of its nature as a perennial forage crop which can be cut multiple times in a growing season.

The starting point for vacuole isolation protocol development for Norway spruce and alfalfa was the same as for *Arabidopsis thaliana*: the protocols of Carter et al. 2004, Robert et al. 2007, and Zouhar 2017.

Two Norway spruce seedlings were obtained from the Johnsteen Company in McKinleyville, CA. Upon arrival, the seedings were 18 and 23 inches tall. Upon arrival they were planted in two 4” square pots 4.5” tall, however this pot size was insufficient to completely cover the roots, so they were transplanted in 8” round pots 6” tall soon afterwards. The spruce seedlings

were too tall for the growth chamber shelves and grew noticeably during their six months in the growth chamber, so they were positioned on the edge of the shelf and allowed to extend beyond the shelf limit (Figure S3-1).

Figure S3-1: The growth chamber shelf at CEF for growing all our plants. The spruce trees had to be placed on the edge of the shelf.

To collect the appropriate amount of needle tissue for the vacuole isolation protocol, we had to determine the average mass of spruce needles from the seedlings and which ones to use. As the spruce trees grew, we found that younger and thinner needles were located lower on the tree and farther out on the branches, while the older and thicker needles were those higher on the tree, either on or closest to the trunk (Figure S3-2). Thick needles from near the top of the trunk of our seedlings weighed on average 8.2mg (n = 20 needles), while the thinner needles from near the end of the lowest branches weighed on average 2.0mg (n = 25 needles). At first, we used thicker needles because less were required for the tissue needed (about 100mg) and they are much easier to cut than thin needles.

Figure S3-2: Norway spruce needles - Left: Thick, older needles located high on the trunk, Right: thin, younger needles located near the end of lower branches.

Spruce needles were sliced with a scalpel much like Arabidopsis rosette leaves and incubated in the same enzyme digest solution (see table 3-2) for four hours. No release of protoplasts from spruce needles was observed even when the experiment was repeated, and the spruce needles were allowed to incubate for up to 44 hours. Several ideas were proposed to solve this problem.

First, we supposed that the wax coating of the cuticle was preventing the enzyme solution from accessing most of the tissue despite the relatively thin slicing. Three pretreatments were attempted to remove or weaken the waxy coating from the needles: acetone, 1N sodium hydroxide, or 2N hydrogen chloride for two hours with shaking, after which the needles were washed with water and examined. The acid and base pretreatments were too strong and caused degradation of the pine needles (Figure S3-3), turning them a brown color where the solution was taken up at the base. The epidermis of the of the pine needles in each pretreatment remained glossy and crisp, so the pretreatment was likely ineffective. A five-minute pretreatment in basic solution was tried before a subsequent incubation in enzyme solution and the base still disrupted the tissue too much, turning the enzyme solution brown after shaking for some time.

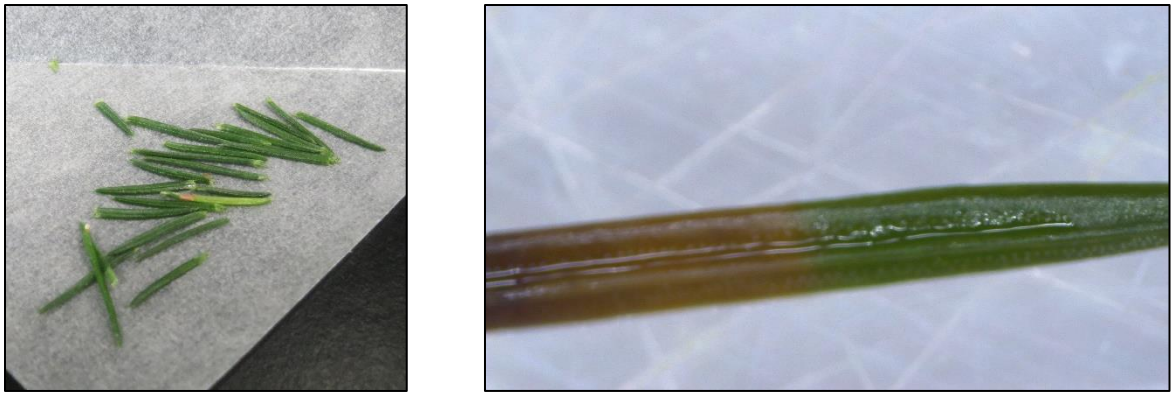


Figure S3-3: Norway spruce needles – Left: about 100mg of spruce needles, Right: a spruce needle with discoloration from acid pretreatment. The darker discoloration is near the base of the needle.

Second, we decided to try slicing the spruce needles longitudinally (the long way) instead of creating cross-sections. Regardless of the waxy cuticle, it would be ideal to bypass the tough epidermis altogether to and expose as much of the mesophyll tissue as possible. Close examination of the inner structure of spruce needles using a dissection microscope revealed that the tissue is arranged in stacked cross-sectional compartments and slicing longitudinally exposed much more mesophyll surface area than slicing in cross-sections, a finding supported in Marco 1939. This slicing technique was much more difficult, however, especially with longer, thinner needles. Ultimately, this method did not produce protoplasts either.

Third and lastly, we tried using new growth from the spruce trees that was much softer, less rigid and light green compared to older needles (Figure S3-4). These needles did not yield any protoplasts either. Only a milky-colored cloudy solution was ever obtained from incubating Norway spruce needles in the enzyme digest solution. The reason why Norway spruce needles are so resistant to cellulase, and other enzyme activity may ironically be due to a higher lignin content than in Arabidopsis leaves. Lignin is well-known to inhibit cellulase activity in biofuel production by physically impeding access to polysaccharides and promoting non-productive enzyme binding (Zeng et al. 2014). Another explanation could lie in the shape of the cells: much of the tissue in spruce needles is made of cells that are covered in protrusions or semi-interlocked, which may

prevent the release of free protoplasts or make it unfeasible for them to release without rupturing soon afterwards (Marco 1939). Our observations suggest that protoplasts were not release at all because no chloroplasts were observed in solution.

Figure S3-4: New growth on Norway spruce seedlings.

Alfalfa (*Medicago sativa*) seed of variety CUF101 was kindly provided by Dr. Charlie Brummer. CUF101 is a nondormant variety "…developed as a joint effort by the University of California, the US Department of Agriculture, and California farmers. It was first released in 1976." (Schlosser 2018 Oct 23). Alfalfa seeds were inoculated with appropriate soil rhizobia, sown individually in moist soil about .25 to .5 inches deep, and covered with soil. We used the same pots as for Arabidopsis but with only one or two plants in each pot.

The alfalfa plants did not grow as robustly as expected. Although the seeds germinated within two days, subsequent health problems occurred. The plants developed curled leaves with some minor brown spots and did not establish a solid crown at their base to support upward growth. Cutting back at about 4 inches high did promote lateral shoot growth, however growth remained slow (Figure S3-5). CUF101 is a nondormant variety developed in California, so the temperature of the growth chamber was probably too low and the humidity too high for optimal growth. Adding soil rhizobia did not produce a visible effect on plant health.

During vacuole isolation protocol development, alfalfa was treated identically to *Arabidopsis thaliana*. Incubation for four hours in the enzyme digest solution did produce protoplasts which were able to be lysed and produce visible, stained, intact vacuoles much like Arabidopsis (Figure S3-6), however, the yield was far too low to be of use in further experiments. Increasing the amount of enzyme in the solution did increase protoplast yield, but not enough to justify continued experiments. Using opened trifoliate leaves versus the unifoliate leaves or unopened trifoliate leaves produced no discernable difference in yield.

Figure S3-5: Alfalfa plants, CUF101 variety – Left: 24 days old, Right: 38 days old with some damage visible.

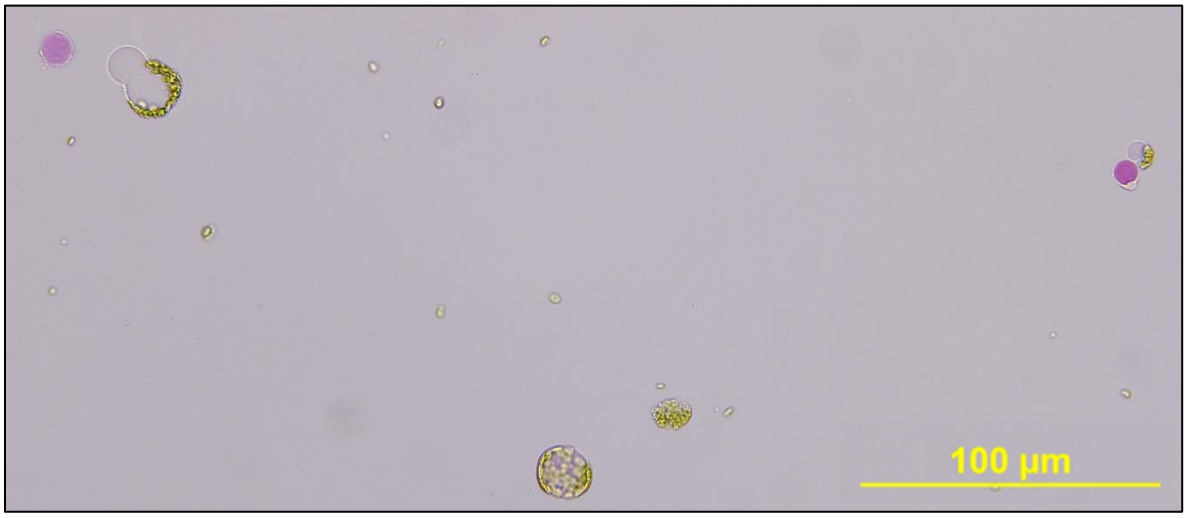


Figure S3-6: Stained vacuoles and debris isolated from *Medicago sativa* (alfalfa).

3. Recommendations for Growing Arabidopsis

When growing *Arabidopsis thaliana*, there are several steps and maintenance tasks to consider. Table S3-1 is a list of recommendations from this project pertaining to different aspects of growing *Arabidopsis thaliana*.

Table S3-1: Recommendations for how to approach several aspects of growing *Arabidopsis thaliana*.

Aspect	Decision-Making Factors / Recommendation
Container size	Arabidopsis can be grown in any size pot, however, larger pot sizes can be watered less frequently. The amount of space available, the number of plants being grown, their purpose, and whether they will be transported are all factors to consider when choosing container size.
Watering frequency	Younger plants do not need to be watered as frequently as mature plants. When the inflorescence and siliques appear, the plants will need much more water than before, potentially everyday depending on container size. For young plants, every other to every three days should be okay. The soil should never dry out completely, however, if algae begin to grow on the top of the soil, water less frequently.
Watering method	Subirrigation was used exclusively and is a convenient method to water many pots at once, even for different container sizes. When plants are at the four-leaf stage or younger, moisten the top of the soil with a spray bottle. To know how long to leave the containers in water, feel their weight by hand. Dry containers are very light and fully watered containers are considerably heavier.
Seed storage	Seeds should be stored in a cool, dry environment. To synchronize germination times, it is recommended to chill at 4°C them for 2-4 days before sowing, although this is not necessary. Seeds can be stored in the refrigerator permanently for convenience if desired.
Soil preparation	Make sure the soil is moist before filling the container. Dry soil will not absorb water by the subirrigation method and is even resistant to water sinking in through the top. If soil is dry, spread it out in a standard 1020 tray and mix it with water until it is moist. The soil should not be anywhere close to dripping wet; soil that is too wet when added will compact too much and not drain sufficiently, leading to algae growth and poor root health.
Sowing method	Seeds of *Arabidopsis thaliana* are exceptionally small; the pipette method described in chapter 3, section 2 is recommended for sowing individual seeds.
Plants per container	Having fewer plants per container allows plants to grow larger and healthier if all other growth conditions are maintained well. However, space constraints can make it inconvenient or infeasible to reduce the number to just one or two. Growing more than four plants in a 4 inch square pot is not recommended.

4. Illustrations of Arabidopsis Growth

The following is an illustrated guide of the growth of *Arabidopsis thaliana*. All timings here are approximate because of variation between plants and germination times for seeds.

Table S3-2: An illustration of various *Arabidopsis thaliana* (Columbia ecotype) plants at different stages of growth. Time is measured from the date seeds were sown. The phenotypes displayed are generally healthy examples with normal germination times.

<table>
<tr><td colspan="2">By day 4, most plants have emerged</td></tr>
<tr><td> 
(Day 4)</td><td> 
(Day 4)</td></tr>
<tr><td colspan="2">By day 6-7, cotelydons grow and become more visible</td></tr>
<tr><td> 
(Day 6, individually sown)</td><td> 
(Day 7, mass sown)</td></tr>
<tr><td colspan="2">By day 9, the four-leaf stage appears</td></tr>
<tr><td> 
(Day 8, 4-leaf stage emerging)</td><td> 
(Day 9, 4-leaf stage clearly visible)</td></tr>
</table>

By day 15, the six-leaf stage appears

(Day 14, 6-leaf stage)

(Day 15)

By day 18, plants are starting to form their rosette

(Day 18)

(Day 18)

In these plants, 22 days old, we can clearly see that variation between the pots is more prominent than the variation within them. This underscores the importance of good soil preparation when sowing seeds. At this stage, the plants will continue growing outward and adding more leaves.

(Day 22)

In these plants, 26 days old, the variation was mainly due to germination time. Since variation is germination time can be significant, measuring plant age from the first appearance of the cotyledons instead of sowing may be better in some applications.

(Day 26)

By day 30-32, the leaves take on an elongated shape.

(Day 28)

(Day 29)

(Day 31)

(Day 32)

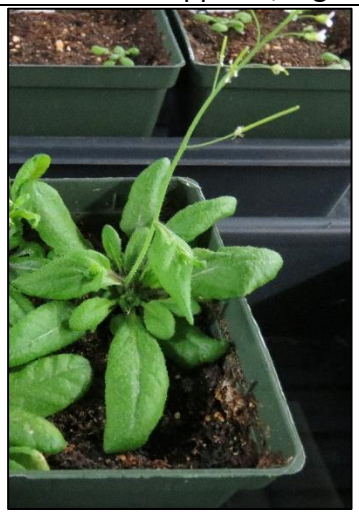
Around day 32-36, the inflorescence starts to grow.

(Day 32)

(Day 34)

After the inflorescence appears, it grows rapidly.

(Day 39)



(Day 39)

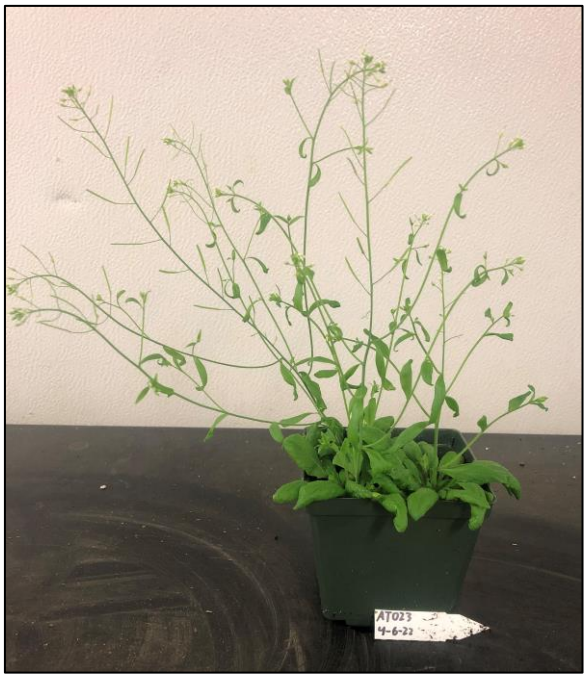
(Day 40)

By day 40, the inflorescences produce flowers, lateral branches, and eventually siliques for producing seed. By day 50, the plants are fully grown and will soon begin to scenesce.

(Day 40)

(Day 43)

(Day 46)



(Day 47)

Table S3-3: Examples of atypical or diseased plant growth.

(Day 14, plants germinated very late)

(Day 26, same plants as to the left)

(Day 35, unhealthy plants due to poor soil drainage and very wet conditions)

(Day 45, an unhealthy plant with an inflorescence)

(Day 42 – a particularly large plant, also slow to produce an inflorescence)

(Day 51, same plant as to the left)

References

1. 3M Company. 2013. Peridex™ (CHLORHEXIDINE GLUCONATE 0.12%) ORAL RINSE. Food and Drug Administration Report No.: 3246401. https://www.accessdata.fda.gov/drugsatfda_docs/label/2013/019028s020lbl.pdf.

2. Alejandro S, Lee Yuree, Tohge T, Sudre D, Osorio S, Park J, Bovet L, Lee Youngsook, Geldner N, Fernie AR, et al. 2012. AtABCG29 Is a Monolignol Transporter Involved in Lignin Biosynthesis. Curr Biol. 22(13):1207–1212. doi:10.1016/j.cub.2012.04.064.

3. Almén MS, Nordström KJ, Fredriksson R, Schiöth HB. 2009. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. BMC Biol. 7(1):50. doi:10.1186/1741-7007-7-50.

4. Arabidopsis Biological Resource Center. 2002. Handling Arabidopsis Plants and Seeds. http://www.biosci.ohio-state.edu/~plantbio/Facilities/abrc/handling.htm.

5. Barros J, Serk H, Granlund I, Pesquet E. 2015. The cell biology of lignification in higher plants. Ann Bot. 115(7):1053–1074. doi:10.1093/aob/mcv046.

6. Behrens CE, Smith KE, Iancu CV, Choe J, Dean JV. 2019. Transport of Anthocyanins and other Flavonoids by the Arabidopsis ATP-Binding Cassette Transporter AtABCC2. Sci Rep. 9(1):437. doi:10.1038/s41598-018-37504-8.

7. Bhuiyan NH, Selvaraj G, Wei Y, King J. 2009. Role of lignification in plant defense. Plant Signal Behav. 4(2):158–159. doi:10.4161/psb.4.2.7688.

8. Bröer S, Fairweather SJ. 2018. Amino Acid Transport Across the Mammalian Intestine. In: Terjung R, editor. Comprehensive Physiology. 1st ed. Wiley. p. 343–373. [accessed 2022 Jun 29]. https://onlinelibrary.wiley.com/doi/10.1002/cphy.c170041.

9. Carter C, Pan S, Zouhar J, Avila EL, Girke T, Raikhel NV. 2004. The Vegetative Vacuole Proteome of *Arabidopsis thaliana* Reveals Predicted and Unexpected Proteins[W]. Plant Cell. 16(12):3285–3303. doi:10.1105/tpc.104.027078.

10. CDPH, OEHHA, DTSC. 2022 May 10. Chemicals Biomonitored in California | Biomonitoring California. Chem Biomonitored Calif. [accessed 2022 May 10]. https://biomonitoring.ca.gov/chemicals/chemicals-biomonitored-california.

11. Cheng Y, Martinez-Guerrero LJ, Wright SH, Kuester RK, Hooth MJ, Sipes IG. 2011. Characterization of the Inhibitory Effects of *N* -Butylpyridinium Chloride and Structurally Related Ionic Liquids on Organic Cation Transporters 1/2 and Human Toxic Extrusion Transporters 1/2-K In Vitro and In Vivo. Drug Metab Dispos. 39(9):1755–1761. doi:10.1124/dmd.110.035865.

12. Cheung KWK, Groen BD, Spaans E, Borselen MD, Bruijn ACJM, Simons-Oosterhuis Y, Tibboel D, Samsom JN, Verdijk RM, Smeets B, et al. 2019. A Comprehensive Analysis of Ontogeny of Renal Drug Transporters: mRNA Analyses, Quantitative Proteomics, and Localization. Clin Pharmacol Ther. 106(5):1083–1092. doi:10.1002/cpt.1516.

13. Cho E, Montgomery RB, Mostaghel EA. 2014. Minireview: SLCO and ABC Transporters: A Role for Steroid Transport in Prostate Cancer Progression. Endocrinology. 155(11):4124–4132. doi:10.1210/en.2014-1337.

14. Cools M, Lissoir S, Bodo E, Ulloa-Calzonzin J, DeLuna A, Georis I, André B. 2020. Nitrogen coordinated import and export of arginine across the yeast vacuolar membrane. PLOS Genet. 16(8):e1008966. doi:10.1371/journal.pgen.1008966.

15. Cunha V, Burkhardt-Medicke K, Wellner P, Santos MM, Moradas-Ferreira P, Luckenbach T, Ferreira M. 2017. Effects of pharmaceuticals and personal care products (PPCPs) on multixenobiotic resistance (MXR) related efflux transporter activity in zebrafish (Danio rerio) embryos. Ecotoxicol Environ Saf. 136:14–23. doi:10.1016/j.ecoenv.2016.10.022.

16. Dean M, Annilo T. 2005. EVOLUTION OF THE ATP-BINDING CASSETTE (ABC) TRANSPORTER SUPERFAMILY IN VERTEBRATES. Annu Rev Genomics Hum Genet. 6(1):123–142. doi:10.1146/annurev.genom.6.080604.162122.

17. Dima O, Morreel K, Vanholme B, Kim H, Ralph J, Boerjan W. 2015. Small Glycosylated Lignin Oligomers Are Stored in Arabidopsis Leaf Vacuoles. Plant Cell. 27(3):695–710. doi:10.1105/tpc.114.134643.

18. Do THT, Martinoia E, Lee Y, Hwang J-U. 2021. 2021 update on ATP-binding cassette (ABC) transporters: how they meet the needs of plants. Plant Physiol. 187(4):1876–1892. doi:10.1093/plphys/kiab193.

19. Eddy R, Hahn D. 2012. 101 Ways to Try to Grow Arabidopsis: Protocol Summary. https://docs.lib.purdue.edu/pmag/7/.

20. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT. 2017. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Res. 45(D1):D320–D324. doi:10.1093/nar/gkw1068.

21. Elsby R, Chidlaw S, Outteridge S, Pickering S, Radcliffe A, Sullivan R, Jones H, Butler P. 2017. Mechanistic in vitro studies confirm that inhibition of the renal apical efflux transporter multidrug and toxin extrusion (MATE) 1, and not altered absorption, underlies the increased metformin exposure observed in clinical interactions with cimetidine, t. Pharmacol Res Perspect. 5(5):e00357. doi:10.1002/prp2.357.

22. Epel D, Luckenbach T, Stevenson CN, MacManus-Spencer LA, Hamdoun A, Smital and T. 2008. Efflux Transporters: Newly Appreciated Roles in Protection against Pollutants. Environ Sci Technol. 42(11):3914–3920. doi:10.1021/es087187v.

23. Etxeberria E, Pozueta-Romero J, Gonzalez P. 2012. In and out of the plant storage vacuole. Plant Sci. 190:52–61. doi:10.1016/j.plantsci.2012.03.010.

24. Fardel O, Kolasa E, Le Vee M. 2012. Environmental chemicals as substrates, inhibitors or inducers of drug transporters: implication for toxicokinetics, toxicity and pharmacokinetics. Expert Opin Drug Metab Toxicol. 8(1):29–46. doi:10.1517/17425255.2012.637918.

25. Fedosova NU, Habeck M, Nissen P. 2021. Structure and Function of Na,K-ATPase—The Sodium-Potassium Pump. In: Terjung R, editor. Comprehensive Physiology. 1st ed. Wiley. p. 2659–2679. [accessed 2022 Jun 29]. https://onlinelibrary.wiley.com/doi/10.1002/cphy.c200018.

26. Fischer WJ, Altheimer S, Cattori V, Meier PJ, Dietrich DR, Hagenbuch B. 2005. Organic anion transporting polypeptides expressed in liver and brain mediate uptake of microcystin. Toxicol Appl Pharmacol. 203(3):257–263. doi:10.1016/j.taap.2004.08.012.

27. Forster IC, Hernando N, Biber J, Murer H. 2013. Phosphate transporters of the SLC20 and SLC34 families. Mol Aspects Med. 34(2–3):386–395. doi:10.1016/j.mam.2012.07.007.

28. Frei M. 2013. Lignin: Characterization of a Multifaceted Crop Component. Sci World J. 2013:1–25. doi:10.1155/2013/436517.

29. Gao X, Liao X, Wu R, Liu T, Wang H, Lu L. 2017. Purified isolation of vacuoles from Sedum alfredii leaf-derived protoplasts. J Zhejiang Univ-Sci B. 18(1):85–88. doi:10.1631/jzus.B1600138.

30. Gaxiola RA, Palmgren MG, Schumacher K. 2007. Plant proton pumps. FEBS Lett. 581(12):2204–2214. doi:10.1016/j.febslet.2007.03.050.

31. Giacomini KM, Balimane PV, Cho SK, Eadon M, Edeki T, Hillgren KM, Huang S-M, Sugiyama Y, Weitz D, Wen Y, et al. 2013. International Transporter Consortium Commentary on Clinically Important Transporter Polymorphisms. Clin Pharmacol Ther. 94(1):23–26. doi:10.1038/clpt.2013.12.

32. Giacomini KM, Huang S-M. 2013. Transporters in Drug Development and Clinical Pharmacology. Clin Pharmacol Ther. 94(1):3–9. doi:10.1038/clpt.2013.86.

33. Giacomini KM, Sugiyama Y. 2017. Membrane Transporters and Drug Response. In: Brunton LL, Hilal-Dandan R, Knollmann BC, editors. Goodman & Gilman's: The Pharmacological Basis of Therapeutics, 13e. New York, NY: McGraw-Hill Education. [accessed 2022 Aug 18]. accessmedicine.mhmedical.com/content.aspx?aid=1162533244.

34. Gomez L, Chrispeels MJ. 1993 Sep 1. Tonoplast and Soluble Vacuolar Proteins Are Targeted by Different Mechanisms. Plant Cell.:1113–1124. doi:10.1105/tpc.5.9.1113.

35. Gottesman MM, Fojo T, Bates SE. 2002. Multidrug resistance in cancer: role of ATP–dependent transporters. Nat Rev Cancer. 2(1):48–58. doi:10.1038/nrc706.

36. Guragain YN, Herrera AI, Vadlani PV, Prakash O. 2015. Lignins of bioenergy crops: a review? Nat Prod Commun. 10(1):201–208.

37. Halwachs S, Wassermann L, Honscha W. 2014. A novel MDCKII in vitro model for assessing ABCG2-drug interactions and regulation of ABCG2 transport activity in the caprine mammary gland by environmental pollutants and pesticides. Toxicol In Vitro. 28(3):432–441. doi:10.1016/j.tiv.2013.12.015.

38. Higgins CF. 2001. ABC transporters: physiology, structure and mechanism – an overview. Res Microbiol. 152(3–4):205–210. doi:10.1016/S0923-2508(01)01193-7.

39. Hites RA, Foran JA, Carpenter DO, Hamilton MC, Knuth BA, Schwager SJ. 2004. Global Assessment of Organic Contaminants in Farmed Salmon. Science. 303(5655):226–229. doi:10.1126/science.1091447.

40. Hu Z, Yang X, Ho PCL, Chan SY, Heng PWS, Chan E, Duan W, Koh HL, Zhou S. 2005. Herb-Drug Interactions: A Literature Review. Drugs. 65(9):1239–1282. doi:10.2165/00003495-200565090-00005.

41. Hwang J-U, Song W-Y, Hong D, Ko D, Yamaoka Y, Jang S, Yim S, Lee E, Khare D, Kim K, et al. 2016. Plant ABC Transporters Enable Many Unique Aspects of a Terrestrial Plant's Lifestyle. Mol Plant. 9(3):338–355. doi:10.1016/j.molp.2016.02.003.

42. Jaquinod M, Villiers F, Kieffer-Jaquinod S, Hugouvieux V, Bruley C, Garin J, Bourguignon J. 2007. A Proteomics Dissection of Arabidopsis thaliana Vacuoles Isolated from Cell Culture. Mol Cell Proteomics. 6(3):394–412. doi:10.1074/mcp.M600250-MCP200.

43. Jones PD, Mankidy R, Newsted J. 2014. Tetrabromobisphenol A. In: Encyclopedia of Toxicology. Elsevier. p. 490–492. [accessed 2022 May 10]. https://linkinghub.elsevier.com/retrieve/pii/B9780123864543004358.

44. Karpiński TM, Szkaradkiewicz AK. 2015. Chlorhexidine--pharmaco-biological activity and application. Eur Rev Med Pharmacol Sci. 19(7):1321–1326.

45. Kodavanti PRS, Loganathan BG. 2014. Polychlorinated biphenyls, polybrominated biphenyls, and brominated flame retardants. In: Biomarkers in Toxicology. Elsevier. p. 433–450. [accessed 2022 May 10]. https://linkinghub.elsevier.com/retrieve/pii/B9780124046306000257.

46. Koziolek M, Alcaro S, Augustijns P, Basit AW, Grimm M, Hens B, Hoad CL, Jedamzik P, Madla CM, Maliepaard M, et al. 2019. The mechanisms of pharmacokinetic food-drug interactions – A perspective from the UNGAP group. Eur J Pharm Sci. 134:31–59. doi:10.1016/j.ejps.2019.04.003.

47. Le Roy J, Huss B, Creach A, Hawkins S, Neutelings G. 2016. Glycosylation Is a Major Regulator of Phenylpropanoid Availability and Biological Activity in Plants. Front Plant Sci. 7. doi:10.3389/fpls.2016.00735. [accessed 2022 Apr 18]. http://journal.frontiersin.org/Article/10.3389/fpls.2016.00735/abstract.

48. Lefèvre F, Baijot A, Boutry M. 2015. Plant ABC transporters: time for biochemistry? Biochem Soc Trans. 43(5):931–936. doi:10.1042/BST20150108.

49. Li Y, Lu J, W. Paxton J. 2012. The Role of ABC and SLC Transporters in the Pharmacokinetics of Dietary and Herbal Phytochemicals and their Interactions with Xenobiotics. Curr Drug Metab. 13(5):624–639. doi:10.2174/1389200211209050624.

50. Li Y, Revalde J, Paxton JW. 2017. The effects of dietary and herbal phytochemicals on drug transporters. Adv Drug Deliv Rev. 116:45–62. doi:10.1016/j.addr.2016.09.004.

51. Liu C-J. 2012. Deciphering the Enigma of Lignification: Precursor Transport, Oxidation, and the Topochemistry of Lignin Assembly. Mol Plant. 5(2):304–317. doi:10.1093/mp/ssr121.

52. Liu Q, Luo L, Zheng L. 2018. Lignins: Biosynthesis and Biological Functions in Plants. Int J Mol Sci. 19(2):335. doi:10.3390/ijms19020335.

53. Lourenço A, Pereira H. 2018. Compositional Variability of Lignin in Biomass. In: Poletto M, editor. Lignin - Trends and Applications. InTech. [accessed 2022 May 7]. http://www.intechopen.com/books/lignin-trends-and-applications/compositional-variability-of-lignin-in-biomass.

54. Luckenbach T, Epel D. 2008. ABCB- and ABCC-type transporters confer multixenobiotic resistance and form an environment-tissue barrier in bivalve gills. Am J Physiol-Regul Integr Comp Physiol. 294(6):R1919–R1929. doi:10.1152/ajpregu.00563.2007.

55. Maciaszczyk-Dziubinska E, Wawrzycka D, Wysocki R. 2012. Arsenic and Antimony Transporters in Eukaryotes. Int J Mol Sci. 13(3):3527–3548. doi:10.3390/ijms13033527.

56. Marchetti S, Mazzanti R, Beijnen JH, Schellens JHM. 2007. Concise Review: Clinical Relevance of Drug–Drug and Herb–Drug Interactions Mediated by the ABC Transporter ABCB1 (MDR1, P-glycoprotein). The Oncologist. 12(8):927–941. doi:10.1634/theoncologist.12-8-927.

57. Marco HF. 1939. The Anatomy of Spruce Needles. J Agric Res. 58(5):357–368.

58. Martinoia E. 2018. Vacuolar Transporters – Companions on a Longtime Journey. Plant Physiol. 176(2):1384–1407. doi:10.1104/pp.17.01481.

59. Mastrantonio V, Ferrari M, Negri A, Sturmo T, Favia G, Porretta D, Epis S, Urbanelli S. 2019. Insecticide Exposure Triggers a Modulated Expression of ABC Transporter Genes in Larvae of Anopheles gambiae s.s. Insects. 10(3):66. doi:10.3390/insects10030066.

60. Miao Y-C, Liu C-J. 2010. ATP-binding cassette-like transporters are involved in the transport of lignin precursors across plasma and vacuolar membranes. Proc Natl Acad Sci. 107(52):22728–22733. doi:10.1073/pnas.1007747108.

61. Montalbetti N, Simonin A, Kovacs G, Hediger MA. 2013. Mammalian iron transporters: Families SLC11 and SLC40. Mol Aspects Med. 34(2–3):270–287. doi:10.1016/j.mam.2013.01.002.

62. Morgan RE, van Staden CJ, Chen Y, Kalyanaraman N, Kalanzi J, Dunn RT, Afshari CA, Hamadeh HK. 2013. A Multifactorial Approach to Hepatobiliary Transporter Assessment Enables Improved Therapeutic Compound Development. Toxicol Sci. 136(1):216–241. doi:10.1093/toxsci/kft176.

63. Morrissey KM, Wen CC, Johns SJ, Zhang L, Huang S-M, Giacomini KM. 2012. The UCSF-FDA TransPortal: A Public Drug Transporter Database. Clin Pharmacol Ther. 92(5):545–546. doi:10.1038/clpt.2012.44.

64. Nagy R, Grob H, Weder B, Green P, Klein M, Frelet-Barrand A, Schjoerring JK, Brearley C, Martinoia E. 2009. The Arabidopsis ATP-binding Cassette Protein AtMRP5/AtABCC5 Is a High Affinity Inositol Hexakisphosphate Transporter Involved in Guard Cell Signaling and Phytate Storage. J Biol Chem. 284(48):33614–33622. doi:10.1074/jbc.M109.030247.

65. Nicklisch SCT, Bonito LT, Sandin S, Hamdoun A. 2017a. Geographic Differences in Persistent Organic Pollutant Levels of Yellowfin Tuna. Environ Health Perspect. 125(6):067014. doi:10.1289/EHP518.

66. Nicklisch SCT, Bonito LT, Sandin S, Hamdoun A. 2017b. Mercury levels of yellowfin tuna (Thunnus albacares) are associated with capture location. Environ Pollut. 229:87–93. doi:10.1016/j.envpol.2017.05.070.

67. Nicklisch SCT, Hamdoun A. 2020. Disruption of small molecule transporter systems by Transporter-Interfering Chemicals (TICs). FEBS Lett. 594(23):4158–4185. doi:10.1002/1873-3468.14005.

68. Nicklisch SCT, Pouv AK, Rees SD, McGrath AP, Chang G, Hamdoun A. 2021. Transporter-interfering chemicals inhibit P-glycoprotein of yellowfin tuna (Thunnus albacares). Comp Biochem Physiol Part C Toxicol Pharmacol. 248:109101. doi:10.1016/j.cbpc.2021.109101.

69. Nicklisch SCT, Rees SD, McGrath AP, Gökirmak T, Bonito LT, Vermeer LM, Cregger C, Loewen G, Sandin S, Chang G, et al. 2016. Global marine pollutants inhibit P-glycoprotein: Environmental levels, inhibitory effects, and cocrystal structure. Sci Adv. 2(4):e1600001. doi:10.1126/sciadv.1600001.

70. Nigam SK. 2015. What do drug transporters really do? Nat Rev Drug Discov. 14(1):29–44. doi:10.1038/nrd4461.

71. Nigam SK. 2018. The SLC22 Transporter Family: A Paradigm for the Impact of Drug Transporters on Metabolic Pathways, Signaling, and Disease. Annu Rev Pharmacol Toxicol. 58(1):663–687. doi:10.1146/annurev-pharmtox-010617-052713.

72. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, et al. 2013. The Norway spruce genome sequence and conifer genome evolution. Nature. 497(7451):579–584. doi:10.1038/nature12211.

73. Palmeira A, Rodrigues F, Sousa E, Pinto M, Vasconcelos MH, Fernandes MX. 2011. New Uses for Old Drugs: Pharmacophore-Based Screening for the Discovery of P-Glycoprotein Inhibitors: Pharmacophore-Based Screening for the Discovery of P-Glycoprotein Inhibitors. Chem Biol Drug Des. 78(1):57–72. doi:10.1111/j.1747-0285.2011.01089.x.

74. Pawarode A, Shukla S, Minderman H, Fricke SM, Pinder EM, O'Loughlin KL, Ambudkar SV, Baer MR. 2007. Differential effects of the immunosuppressive agents cyclosporin A, tacrolimus and sirolimus on drug transport by multidrug resistance proteins. Cancer Chemother Pharmacol. 60(2):179–188. doi:10.1007/s00280-006-0357-8.

75. Perkins M, Smith RA, Samuels L. 2019. The transport of monomers during lignification in plants: anything goes but how? Curr Opin Biotechnol. 56:69–74. doi:10.1016/j.copbio.2018.09.011.

76. Pessatti ML, Resgalla Jr. C, REIS Fo. RW, Kuehn J, Salomão LC, Fontana JD. 2002. Variability of filtration and food assimilation rates, respiratory activity and multixenobiotic resistance (MXR) mechanism in the mussel Perna perna under lead influence. Braz J Biol. 62(4a):651–656. doi:10.1590/S1519-69842002000400013.

77. Robert S, Zouhar J, Carter C, Raikhel N. 2007. Isolation of intact vacuoles from Arabidopsis rosette leaf–derived protoplasts. Nat Protoc. 2(2):259–262. doi:10.1038/nprot.2007.26.

78. Robey RW, Pluchino KM, Hall MD, Fojo AT, Bates SE, Gottesman MM. 2018. Revisiting the role of ABC transporters in multidrug-resistant cancer. Nat Rev Cancer. 18(7):452–464. doi:10.1038/s41568-018-0005-8.

79. Romersi RF, Nicklisch SCT. 2022. Interactions of Environmental Chemicals and Natural Products With ABC and SLC Transporters in the Digestive System of Aquatic Organisms. Front Physiol. 12:767766. doi:10.3389/fphys.2021.767766.

80. Sahoo S, Aurich MK, Jonsson JJ, Thiele I. 2014. Membrane transporters in a human genome-scale metabolic knowledgebase and their implications for disease. Front Physiol. 5. doi:10.3389/fphys.2014.00091. [accessed 2022 Aug 18]. http://journal.frontiersin.org/article/10.3389/fphys.2014.00091/abstract.

81. Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, Lam KJK, Tian N, Russum S, Wang J, et al. 2021. The Transporter Classification Database (TCDB): 2021 update. Nucleic Acids Res. 49(D1):D461–D467. doi:10.1093/nar/gkaa1004.

82. Schecter A, Colacino J, Patel K, Kannan K, Yun SH, Haffner D, Harris TR, Birnbaum L. 2010. Polybrominated diphenyl ether levels in foodstuffs collected from three locations from the United States. Toxicol Appl Pharmacol. 243(2):217–224. doi:10.1016/j.taap.2009.10.004.

83. Schecter A, Cramer P, Boggess K, Stanley J, Päpke O, Olson J, Silver A, Schmitz M. 2001. Intake Of Dioxins and Related Compounds from Food in the U.S. Population. J Toxicol Environ Health A. 63(1):1–18. doi:10.1080/152873901750128326.

84. Schecter A, Haffner D, Colacino J, Patel K, Päpke O, Opel M, Birnbaum L. 2010. Polybrominated Diphenyl Ethers (PBDEs) and Hexabromocyclodecane (HBCD) in Composite U.S. Food Samples. Environ Health Perspect. 118(3):357–362. doi:10.1289/ehp.0901345.

85. Schlessinger A, Yee SW, Sali A, Giacomini KM. 2013. SLC Classification: An Update. Clin Pharmacol Ther. 94(1):19–23. doi:10.1038/clpt.2013.73.

86. Schlosser M "Kitty." 2018 Oct 23. Alfalfa - CUF 101. UC Davis Found Seed Program. [accessed 2022 Jul 4]. https://fsp.ucdavis.edu/seed-catalog/alfalfa-varieties/cuf-101.

87. Seong KM, Sun W, Clark JM, Pittendrigh BR. 2016. Splice form variant and amino acid changes in MDR49 confers DDT resistance in transgenic Drosophila. Sci Rep. 6(1):23355. doi:10.1038/srep23355.

88. Shapiro AB, Ling V. 1994. ATPase activity of purified and reconstituted P-glycoprotein from Chinese hamster ovary cells. J Biol Chem. 269(5):3745–3754. doi:10.1016/S0021-9258(17)41923-5.

89. Shen J, Zeng Y, Zhuang X, Sun L, Yao X, Pimpl P, Jiang L. 2013. Organelle pH in the Arabidopsis Endomembrane System. Mol Plant. 6(5):1419–1437. doi:10.1093/mp/sst079.

90. Shimaoka T, Ohnishi M, Sazuka T, Mitsuhashi N, Hara-Nishimura I, Shimazaki K-I, Maeshima M, Yokota A, Tomizawa K-I, Mimura T. 2004. Isolation of Intact Vacuoles and Proteomic Analysis of Tonoplast from Suspension-Cultured Cells of Arabidopsis thaliana. Plant Cell Physiol. 45(6):672–683. doi:10.1093/pcp/pch099.

91. Shitan N, Yazaki K. 2020. Dynamism of vacuoles toward survival strategy in plants. Biochim Biophys Acta BBA - Biomembr. 1862(12):183127. doi:10.1016/j.bbamem.2019.183127.

92. Shukla S, Abel B, Chufan EE, Ambudkar SV. 2017. Effects of a detergent micelle environment on P-glycoprotein (ABCB1)-ligand interactions. J Biol Chem. 292(17):7066–7076. doi:10.1074/jbc.M116.771634.

93. Sibout R, Höfte H. 2012. Plant Cell Biology: The ABC of Monolignol Transport. Curr Biol. 22(13):R533–R535. doi:10.1016/j.cub.2012.05.005.

94. Sjöstedt N, Deng F, Rauvala O, Tepponen T, Kidron H. 2017. Interaction of Food Additives with Intestinal Efflux Transporters. Mol Pharm. 14(11):3824–3833. doi:10.1021/acs.molpharmaceut.7b00563.

95. Smital T, Luckenbach T, Sauerborn R, Hamdoun AM, Vega RL, Epel D. 2004. Emerging contaminants—pesticides, PPCPs, microbial degradation products and natural substances as inhibitors of multixenobiotic defense in aquatic organisms. Mutat Res Mol Mech Mutagen. 552(1–2):101–117. doi:10.1016/j.mrfmmm.2004.06.006.

96. Song W-Y, Mendoza-Cózatl DG, Lee Y, Schroeder JI, Ahn S-N, Lee H-S, Wicker T, Martinoia E. 2014. Phytochelatin-metal(loid) transport into vacuoles shows different substrate preferences in barley and *Arabidopsis*: Phytochelatin-metal(loid) vacuolar transport in barley. Plant Cell Environ. 37(5):1192–1201. doi:10.1111/pce.12227.

97. STRING Consortium 2022. STRING: functional protein association networks. STRING. [accessed 2022 Jul 3]. https://string-db.org/.

98. Susann, Biddulph O. 1959. The Circulatory System of Plants. Sci Am. 200(2):44–59. doi:10.1038/scientificamerican0259-44.

99. The International Transporter Consortium, Giacomini KM, Huang S-M, Tweedie DJ, Benet LZ, Brouwer KLR, Chu X, Dahlin A, Evers R, Fischer V, et al. 2010. Membrane transporters in drug development. Nat Rev Drug Discov. 9(3):215–236. doi:10.1038/nrd3028.

100. Thorens B, Mueckler M. 2010. Glucose transporters in the 21st Century. Am J Physiol-Endocrinol Metab. 298(2):E141–E145. doi:10.1152/ajpendo.00712.2009.

101. Tikkanen A, Pierrot E, Deng F, Sánchez VB, Hagström M, Koenderink JB, Kidron H. 2020. Food Additives as Inhibitors of Intestinal Drug Transporter OATP2B1. Mol Pharm. 17(10):3748–3758. doi:10.1021/acs.molpharmaceut.0c00507.

102. Tirona RG, Leake BF, Wolkoff AW, Kim RB. 2003. Human Organic Anion Transporting Polypeptide-C (SLC21A6) Is a Major Determinant of Rifampin-Mediated Pregnane X Receptor Activation. J Pharmacol Exp Ther. 304(1):223–228. doi:10.1124/jpet.102.043026.

103. Tobimatsu Y, Schuetz M. 2019. Lignin polymerization: how do plants manage the chemistry so well? Curr Opin Biotechnol. 56:75–81. doi:10.1016/j.copbio.2018.10.001.

104. Tsuyama T, Kawai R, Shitan N, Matoh T, Sugiyama J, Yoshinaga A, Takabe K, Fujita M, Yazaki K. 2013. Proton-Dependent Coniferin Transport, a Common Major Transport Event in Differentiating Xylem Tissue of Woody Plants. Plant Physiol. 162(2):918–926. doi:10.1104/pp.113.214957.

105. Tsuyama T, Matsushita Y, Fukushima K, Takabe K, Yazaki K, Kamei I. 2019. Proton Gradient-Dependent Transport of p-Glucocoumaryl Alcohol in Differentiating Xylem of Woody Plants. Sci Rep. 9(1):8900. doi:10.1038/s41598-019-45394-7.

106. UNEP. 2019. Listing of POPs in the Stockholm Convention. POPs List Stockh Conv. [accessed 2022 May 27]. http://chm.pops.int/TheConvention/ThePOPs/ListingofPOPs/tabid/2509/Default.aspx.

107. US EPA. 2016 Jan 3. Pentachlorophenol. [accessed 2022 May 27]. https://www.epa.gov/ingredients-used-pesticide-products/pentachlorophenol.

108. Väisänen E, Takahashi J, Obudulu O, Bygdell J, Karhunen P, Blokhina O, Laitinen T, Teeri TH, Wingsle G, Fagerstedt KV, et al. 2020. Hunting monolignol transporters: membrane proteomics and biochemical transport assays with membrane vesicles of Norway spruce. Zhao Q, editor. J Exp Bot. 71(20):6379–6395. doi:10.1093/jxb/eraa368.

109. Wang N, Westerterp M. 2020. ABC Transporters, Cholesterol Efflux, and Implications for Cardiovascular Diseases. In: Jiang X-C, editor. Lipid Transfer in Lipoprotein Metabolism and Cardiovascular Disease. Vol. 1276. Singapore: Springer Singapore. (Advances in Experimental Medicine and Biology). p. 67–83. [accessed 2022 Jun 29]. https://link.springer.com/10.1007/978-981-15-6082-8_6.

110. Welker C, Balasubramanian V, Petti C, Rai K, DeBolt S, Mendu V. 2015. Engineering Plant Biomass Lignin Content and Composition for Biofuels and Bioproducts. Energies. 8(8):7654–7676. doi:10.3390/en8087654.

111. Xiong J, Feng J, Yuan D, Zhou J, Miao W. 2015. Tracing the structural evolution of eukaryotic ATP binding cassette transporter superfamily. Sci Rep. 5(1):16724. doi:10.1038/srep16724.

112. Xu J, Peng H, Ting Zhang J. 2007. Human Multidrug Transporter ABCG2, a Target for Sensitizing Drug Resistance in Cancer Chemotherapy. Curr Med Chem. 14(6):689–701. doi:10.2174/092986707780059580.

113. Yang X, Chen W, Song X, Ma X, Cotto-Rivera RO, Kain W, Chu H, Chen Y-R, Fei Z, Wang P. 2019. Mutation of ABC transporter ABCA2 confers resistance to Bt toxin Cry2Ab in Trichoplusia ni. Insect Biochem Mol Biol. 112:103209. doi:10.1016/j.ibmb.2019.103209.

114. Yoo S-D, Cho Y-H, Sheen J. 2007. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat Protoc. 2(7):1565–1572. doi:10.1038/nprot.2007.199.

115. Zeng Y, Zhao S, Yang S, Ding S-Y. 2014. Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels. Curr Opin Biotechnol. 27:38–45. doi:10.1016/j.copbio.2013.09.008.

116. Zhang Yong, Zhang Yuping, Sun K, Meng Z, Chen L. 2019. The SLC transporter in nutrient and metabolic sensing, regulation, and drug development. Liu F, editor. J Mol Cell Biol. 11(1):1–13. doi:10.1093/jmcb/mjy052.

117. Zou L, Stecula A, Gupta A, Prasad B, Chien H-C, Yee SW, Wang L, Unadkat JD, Stahl SH, Fenner KS, et al. 2018. Molecular Mechanisms for Species Differences in Organic Anion Transporter 1, OAT1: Implications for Renal Drug Toxicity. Mol Pharmacol. 94(1):689–699. doi:10.1124/mol.117.111153.

118. Zouhar J. 2017. Isolation of Vacuoles and the Tonoplast. In: Taylor NL, Millar AH, editors. Isolation of Plant Organelles and Structures. Vol. 1511. New York, NY: Springer New York. (Methods in Molecular Biology). p. 113–118. [accessed 2022 Apr 6]. http://link.springer.com/10.1007/978-1-4939-6533-5_9.