Environmental and Molecular Mutagenesis 53:478–487 (2012)

Research Article

Predicted Toxicity of the Biofuel Candidate 2,5-Dimethylfuran in Environmental and Biological Systems

Jimmy Phuong, Simon Kim, Reuben Thomas, and Luoping Zhang*

Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, California

Although not mutagenic by Ames test, 2,5-dimethylfuran (DMF), a leading biofuel candidate, was found to induce chromosomal damage in cultured murine cells, suggesting that it may be genotoxic. We sought to prioritize the environmental and biological impacts of using DMF as a combustible biofuel. First, we assessed DMF and its combustion intermediates for potential persistence, bioaccumulation, and aquatic toxicity (PBT) using PBT profiler. Our findings predict DMF to have moderate-level aquatic toxicity; however, a greater subset of the combustion intermediates is predicted to have moderate- and high-level aquatic toxicity with bioaccumulation and persistence concerns. Second, we assessed the biological impact of DMF by testing for statistically significant chemical–disease associations. No direct associations for DMF were found; however, indirect associations were identified from two structurally similar analogs. Curated associations between furfuryl alcohol to kidney neoplasm and adenoma, and significant inferred associations between furan to lung neoplasm, drug-induced liver injury, and experimentally induced liver cirrhosis were found, based on 21 furan–gene interactions. Nine of 49 DMF combustion intermediates analyzed, including benzene and 1,3-butadiene, were found to have associations with 26 tumors and systemic diseases. Although inadequate for a stand-alone risk assessment, our data suggest that DMF combustion intermediates pose a much broader range of hazards than DMF itself, and that both should be further investigated. Environ. Mol. Mutagen. 53:478–487, 2012.

Key words: 2,5-dimethylfuran; biofuel; computational toxicology; air toxics; systems biology

INTRODUCTION

2,5-Dimethylfuran (DMF) is among many first-generation biofuel candidates that have received renewed interest because of increased economic demand for new sources of energy [Schmidt and Dauenhauer, 2007; Luque et al., 2008]. However, the environmental impact, human health risks, and overall sustainability of these candidate biofuels are concerns to be addressed [McKone et al., 2011].

DMF has the potential to be mass produced from plant cellulose faster and with higher energy content than ethanol, through an improved method for reacting biomass [Román-Leshkov et al., 2007; Schmidt and Dauenhauer, 2007; Luque et al., 2008]. Compared with ethanol, DMF has 40% higher energy content per molecule, a higher proposed research octane number for better engine performance, and physical properties that allow for more versatile means of extraction from water and for product transportation. Furthermore, the currently proposed production method for DMF has a net energy balance of 2.2 units of energy returned per unit invested in the process, which is potentially a twofold increase over ethanol [Román-Leshkov et al., 2007; Schmidt and Dauenhauer, 2007; Luque et al., 2008].

Despite the benefits of using DMF as a biofuel, there are concerns about its downstream environmental and human health impacts [Luque et al., 2008; Centers for Disease Control and Prevention, 2009; McKone et al., 2011]. DMF is one of many volatile organic compounds (VOCs) in cigarette smoke and coffee vapor, and it has been detected within exhaled air, systemic blood, and excreted urine of active and passive tobacco smokers

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Institute for Environmental Health Sciences (NIEHS); Grant Number R01ES017452; Grant sponsors: Northern California Center for Occupational and Environmental Health, Council for Education and Research on Toxics.

*Correspondence to: Luoping Zhang, Division of Environmental Health Sciences, School of Public Health, University of California, 237 Hildebrand Hall, Berkeley, CA 94720-7356, USA. E-mail: luoping@berkeley.edu

Received 3 February 2012; provisionally accepted 9 April 2012; and in final form 11 April 2012

DOI 10.1002/em.21702

Published online 25 June 2012 in Wiley Online Library (wileyonlinelibrary.com).

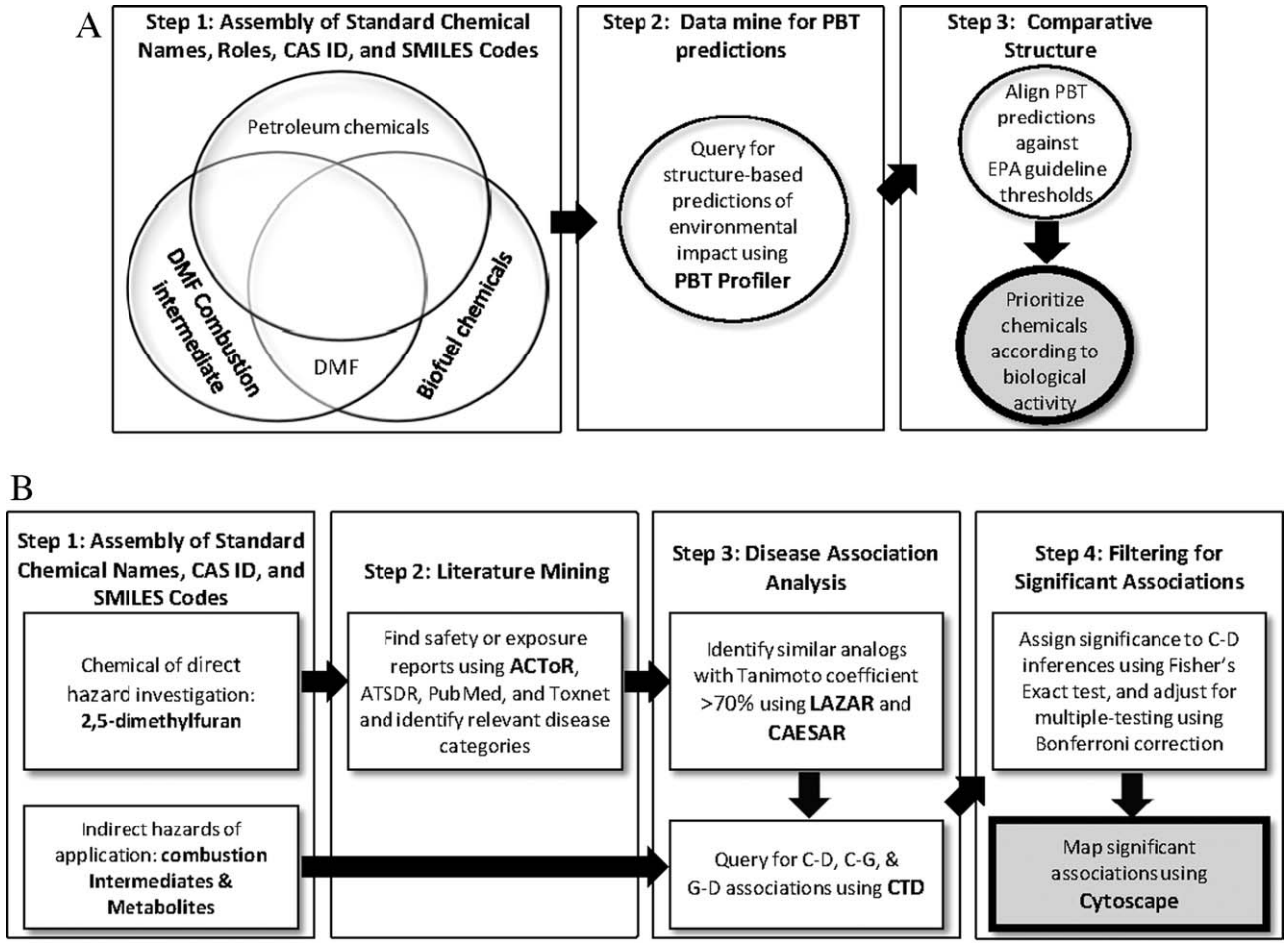

Fig. 1. The general workflow for predicting the biological and environmental effects of DMF. C-D (chemical–disease); C-G (chemical–gene); and G-D (gene–disease).

[Perbellini et al., 2003; Centers for Disease Control and Prevention, 2009; Alonso et al., 2010]. In addition, DMF is one of the metabolites excreted in the urine of subjects exposed to hexane [Iwata et al., 1983, 1984; Centers for Disease Control and Prevention, 2009]. It has not been determined, however, whether DMF or its derivatives play a role in smoking-related diseases or in hexane neurotoxicity.

Currently, there is limited evidence of DMF environmental and biological toxicity. DMF was tested in 119 PubChem Bioassays and determined only to cause acute baseline narcosis in the EPA Fathead Minnow bioassay, with a lethal concentration at 50% of 71.1 mg/l [National Center for Biotechnology Information, 2011]. DMF was nonmutagenic in the Ames bacterial tests [Zeiger et al., 1992]; however, it was recently found to cause chromosome aberrations in murine cell cultures, suggesting that it may have genotoxic action [Fromowitz et al., 2012].

Given the limited toxicity data on DMF and the furan chemical class, the aim of this study was to use computational toxicology approaches to form toxicity predictions for DMF and its combustion intermediates described by Wu et al. [2009]. We used quantitative structure–activity relationship (QSAR) models to predict persistence, bioaccumulation, and aquatic toxicity (PBT) of DMF and its combustion intermediates and to identify structurally similar analogs. We also tested, DMF, its combustion intermediates and structurally similar analogs for significant chemical–disease associations using curated chemical data from the Comparative Toxicogenomics Database (CTD). Thus, our approach facilitates the identification of plausible immediate and downstream environmental and biological hazards and could be applied in the preliminary life-cycle assessment of other emerging, new green chemical candidates and environmental contaminants.

MATERIALS AND METHODS

The Computational Approach

The goal of this study is to evaluate DMF and its downstream derivatives with regards to potential effects: (1) cancer, (2) noncancer toxicities, and (3) environmental impacts.

The computational approach for examining cancer and noncancer toxicities consists of four steps (Fig. 1a). In step 1, the direct hazard for investigation (DMF) and the indirect downstream chemicals were identi-

fied with Chemical Abstract Service Registry Number (CASRN), simplified molecular-input line-entry system (SMILES) code, and alternative names. In step 2, a preliminary text search using chemical and literature knowledgebases such as the Aggregate Computational Toxicology Resource (ACToR), ToxNet, and PubMed was conducted to identify existing diseases of concern associated with current modes of DMF exposure. In step 3, we used QSAR cheminformatics, Lazy Structure–Activity Relationships (LAZAR), and the Computer-Assisted Evaluation of Industrial Chemical Substances According to Regulations (CAESAR) to generate quantitative predictions and identify structurally similar analogs. The Computational Toxicogenomics Database (April 7, 2011 update) was then applied to identify curated chemical–gene–disease associations. In step 4, the Fisher's exact test was used to assign statistical significance and the relevant associations were mapped using Cytoscape.

The computational approach for examining environmental impacts consists of three steps (Fig. 1b). In step 1, all chemicals were categorized with relation to their role as a petroleum gasoline constituent, biofuel chemical, and/or DMF combustion intermediate. In step 2, each group of chemicals was screened for PBT concerns generated from structure-based predictions using PBT profiler. In step 3, the PBT profiler predictions are aligned with the United States Environmental Protection Agency (US EPA) guidelines for PBT. These values were then used to prioritize chemicals according to their predicted biological activity.

Associated Chemicals of Concern

A list of chemicals related to DMF in its application as a biofuel was compiled for comparison. The list included other biofuel candidates [Luque et al., 2008], major in-use gasoline constituents [Fujita et al., 2011], and combustion intermediates identified from DMF combustion flame studies [Lifshitz et al., 1998; Wu et al., 2009; Simmie and Metcalfe, 2011]. The input lines (CASRN, SMILES code, and synonyms), which allow the software to recognize each chemical structure, were retrieved from PubChem and the National Institute for Standards and Technology (Supporting Information Table 1). Chemicals with a CASRN or SMILES code were deemed computable, and chemicals that did not satisfy these criteria were excluded from the computational approach.

Fifty-seven computable compounds were identified, consisting of six biofuel candidates, seven petroleum gasoline compounds, and 49 DMF combustion intermediates (Supporting Information Table 2). Some chemicals showed overlap across multiple categories, but were treated in separate categories because of differences in presumed physical phase (i.e., petroleum gasoline constituents in liquid phase may also occur as DMF combustion intermediates in vapor phase). Radical compounds were excluded from this study because they were not computable or because the software misread the SMILES codes (Supporting Information Table 2).

Assessment of Environmental Impacts

All 57 computable chemicals were analyzed for potential PBT concerns using QSAR predictions from PBT profiler. The Toxic Substance Control Act (TSCA) Section 5 of 1990 and the Emergency Planning and Community Right-to-Know Act (EPCRA) Section 313 of 1986 were used for the classification thresholds for environmental persistence and bioaccumulation [U.S. Environmental Protection Agency, 1999a,b].

TSCA guidelines are used as the current standard for US EPA determination of persistent, bioaccumulative, and toxic chemicals. According to TSCA, a persistent compound has a half-life in soil, sediment, and water of >60 days, whereas a very persistent compound has a half-life of >180 days. Persistence classification is considered viable only if the respective predicted relative distribution of the chemical in the respective medium is at least 10%. For instance, a persistent chemical may have a predicted distribution in water at 35% of its original concentration and 75 days is its predicted half-life in water. A bioaccumulative compound is defined as having a bioconcentration factor (BCF) $>$ 1,000, meaning that the chemical is found within the organism at more than 1,000 times the concentration found in the surrounding environment. Aquatic toxicity classifications are defined by predicted chronic exposure value (ChV) for acute toxicity, which is approximated from QSAR predictions for dose exposure, and deemed low if >10 mg/l, moderate if between 10 and 0.1 mg/l, or high if <0.1 mg/l.

The EPCRA criteria, a list of persistence criteria formerly used by the US EPA, were incorporated to identify potentially air-persistent chemicals because the TSCA guidelines do not set parameters for air persistence. These chemicals are classified as having a predicted relative distribution in air greater than 10% and a half-life in air greater than 2 days. The EPCRA water, soil, and sediment persistence parameters are similar to those of the TSCA guidelines.

Associations With Cancer and Noncancer Toxicities

Given the limited data on DMF's biological activity, structurally similar analogs were used as proxies for DMF to predict possible cancer effects and noncancer toxicities. Analogs similar to DMF with at least a 70% Tanimoto coefficient are assumed to have similar modes of action and, hence, similar gene and disease associations. The similar analogs were retrieved on February 11, 2011 from LAZAR [Helma, 2006] and CAESAR [Fjodorova et al., 2010], which are QSAR applications constructed and validated with priority chemicals of toxicological concern. To simplify the analysis, we chose use this analog approach for DMF but not for its combustion intermediates.

Using the downloadable flat files from CTD [Mount Desert Island Biological Laboratory, 2011] accessed on April 15, 2011, we identified curated gene and disease associations with DMF and its associated chemicals in the mammalian models: *Homo sapiens* (human), *Rattus norvegicus* (Norway rat), and *Mus musculus* (house mouse). Each curated chemical–gene and gene–disease association was then compared to identify overlapping genes and chemical–disease associations. Within our identification, gene interaction types (e.g., increase or decrease in mRNA expression) are not taken into account but are provided within Supporting Information Table 3. As described in the following section, the overlapping genes can be used to statistically characterize significant chemical–disease associations.

Statistical Analysis of Chemical–Disease Associations

Using CTD's Batch Query tool, inferred chemical–disease associations and overlapping gene interactions were gathered to bridge the data gap in toxicity information. The 2×2 Fisher's exact test was used to evaluate potential chemical–disease associations based on the curated chemical–gene and gene–disease overlaps. Two null hypotheses were tested, each requiring separate adjustment for multiple testing: direct hazard from DMF (through similar analogs) and indirect hazard from DMF's combustion intermediates.

- Null hypothesis 1 (for the DMF data gap analysis): ''there is no association between DMF and any of the curated diseases.''
- Null hypothesis 2 (for the data gap analysis of DMF combustion intermediates): ''there is no association between the combustion intermediates of DMF and any of the curated diseases.''

The Fisher's exact tests were conducted using a published Excel spreadsheet from the *Handbook of Biological Statistics* [McDonald, 2009]. To evaluate significance, these tests use the total number of curated gene interactions within CTD per chemical and per disease, the total number of overlapping genes between each chemical and disease, and the total size of the gene pool, 20,811, which was the number of ''genes with curated data'' reported in CTD.

The resulting *P*-values from the Fisher's exact tests were adjusted for multiple testing using the Bonferroni correction. This was done by multiplying the *P*-value by the total number of chemical–disease association evaluated, which is given as the product of the total number of chemicals considered (i.e., two for null hypothesis 1 and 15 for null hypothesis 2) by

the total number of disease endpoints with gene interactions possible (i.e., 2203 for both hypotheses). Adjusted *P*-values were then used for assigning significance according to family-wise error thresholds at 1, 5, 10, and 20%.

Cytoscape Mapping

The inferred chemical–disease associations, stratified according to the assigned significance (Supporting Information Tables 3 and 4), and the curated chemical–disease associations were mapped in Cytoscape to denote shared and common connections between chemicals, genes, and disease endpoints [Smoot et al., 2011].

RESULTS

Comparison of the Environmental Impact of DMF, Biofuels, and Petroleum Chemicals

According to TSCA criteria, PBT profiler predictions for DMF suggest moderate-level aquatic toxicity, which is greater than that of the other biofuels but less than that of most petroleum gasoline constituents (Fig. 2a). With the exception of DMF, all biofuel compounds have low-level predicted aquatic toxicity. In contrast, the major petroleum gasoline constituents have moderate-level aquatic toxicity predictions, aside from methyl *tert*-butyl ether (MTBE). No bioaccumulation concerns were predicted among the petroleum gasoline or biofuel compounds.

The potential for persistence was detected for one biofuel and six gasoline chemicals (Supporting Information Table 2). The TSCA and EPCRA criteria suggest that 1,3,5-trimethylbenzene, benzene, and naphthalene are persistent in water, soil, or sediment. In addition, the EPCRA results indicate that dimethylether, ethylbenzene, MTBE, and toluene may be persistent in ambient air.

Environmental Impact Predictions for DMF Combustion Intermediates

Among the 49 computable DMF combustion intermediates, 23 compounds were predicted to pose greater PBT impacts than DMF (Fig. 2b), wherein 15 of these 23 compounds may pose PBT hazards to the environment. Ethenyl oxirane, phenol, and 1-naphthalenol are predicted to have high-level aquatic toxicity, whereas 37 others have moderate levels of aquatic toxicity. Benzene and naphthalene are predicted to be moderately toxic to aquatic life and persistent in soil, whereas five other combustion intermediates (i.e., acetone, cyclopropane, ethylbenzene, propyne, and toluene) are classified as persistent in air. None of the DMF combustion intermediates were classified as having bioaccumulative. However, 2-methylnaphthalene, 3-methylindene, *E*-1-phenyl-1-butene, *E*-1-propenylbenzene, and 2,5-cyclohexadiene-1,4-dimethylene had the highest predicted BCF values (BCF $>$ 100) accompanied by moderate-level aquatic toxicity. The data for these five chemicals suggest that they may have the highest potential to bioconcentrate within living organisms.

Chemical–Disease Associations for DMF and Its Structural Analogs

DMF, by itself, does not have any curated disease associations within CTD nor are there available biological studies indicating gene or disease association. However, using LAZAR and CAESAR, we identified nine analogs structurally similar to DMF, of which three chemicals—furan, furfuryl alcohol, and pyralene—have curated data within CTD (Fig. 3). A direct association between furfuryl alcohol and adenoma and kidney neoplasm was found. Out of 167 disease associations inferred in our analysis, only three associations from furan (lung neoplasm, experimentally induced liver cirrhosis, and drug-induced liver injury), and none of the associations from pyralene, were found to be significant at the 1% family-wise error threshold. These three diseases were inferred from 21 furan–gene interactions (Supporting Information Table 3). An association with Pasteurellaceae infection was also inferred from furan but found to be significant at the 20% error threshold; thus, this association was determined to be an equivocal finding. The adjusted *P*-value for each inferred disease is reported in Table I.

Predicted Disease Associations for DMF Combustion Intermediates

Of the 49 computable combustion intermediates analyzed, predicted disease associations were identified for only nine chemicals (i.e., 1,3-butadiene, acetaldehyde, benzene, benzofuran, ethylbenzene, naphthalene, phenol, styrene, and toluene), with direct evidence for seven chemicals and significantly inferred associations for eight chemicals (Fig. 4).

A total of 76 chemical–disease associations were found to be significant at the 1% family-wise error rate threshold, shown with red edge lines in Figure 4, whereas 37 additional inferred chemical–disease associations were found significant up to the 20% family-wise error threshold (Supporting Information Table 4). Sixty-one direct-evidence associations are displayed as black and blue edge lines. Black edge lines represent a correlation with disease etiology; blue edges represent a correlated role of therapeutic effect against the disease. The central ring shown in Figure 4 represents diseases that are independently and significantly associated with multiple chemicals (e.g., lung neoplasm is associated with 1,3-butadiene, benzene, and acetaldehyde). Only phenol was found to have direct evidence of therapeutic association.

DISCUSSION

Environmental Impact of DMF Relative to Other Biofuels

Our findings suggest that DMF has a higher environmental impact compared with the other biofuels analyzed.

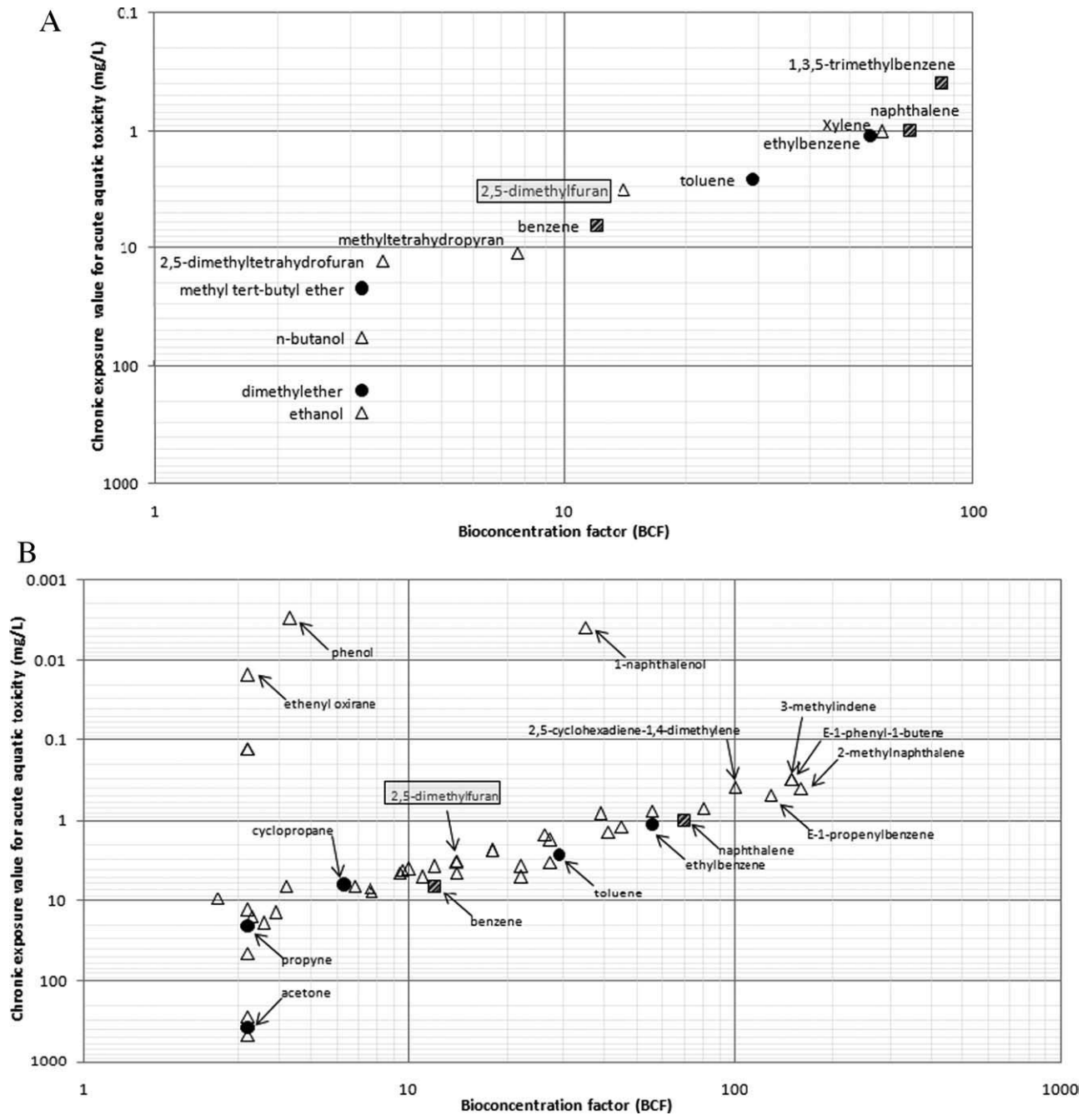

Fig. 2. (**a**) PBT profiler predictions for six biofuel and seven petroleum gasoline chemicals. (**b**) PBT profiler predictions for 49 DMF combustion intermediates. According to TSCA Section 5, a bioaccumulative compound is categorized as having a BCF $>$ 1,000, and a chemical of low-level aquatic toxicity has a ChV $>$ 10 mg/l, moderate-level aquatic toxicity has a ChV between 10 and 0.1 mg/l, and high-level aquatic toxicity has a ChV $<$ 0.1 mg/l. ▨ indicates concerns of environmental persistence according to the TSCA and EPCRA criteria. • indicates concerns of environmental persistence according to only the EPCRA criteria for air persistence. △ indicates no notable persistence concerns by TSCA or EPCRA criteria.

It has the highest relative distribution in water among the biofuels and has a predicted BCF comparable to benzene (Fig. 2a and Supporting Information Table 2). Thus, DMF may be as hazardous to aquatic life as some of the current gasoline constituents.

Environmental Impact of DMF Combustion Intermediates

We analyzed 49 combustion intermediates from the list reported by Wu et al. [2009] that were computable in our study. A recent DMF decomposition mechanism from Simmie and Metcalfe [2011] confirmed the identity of most of

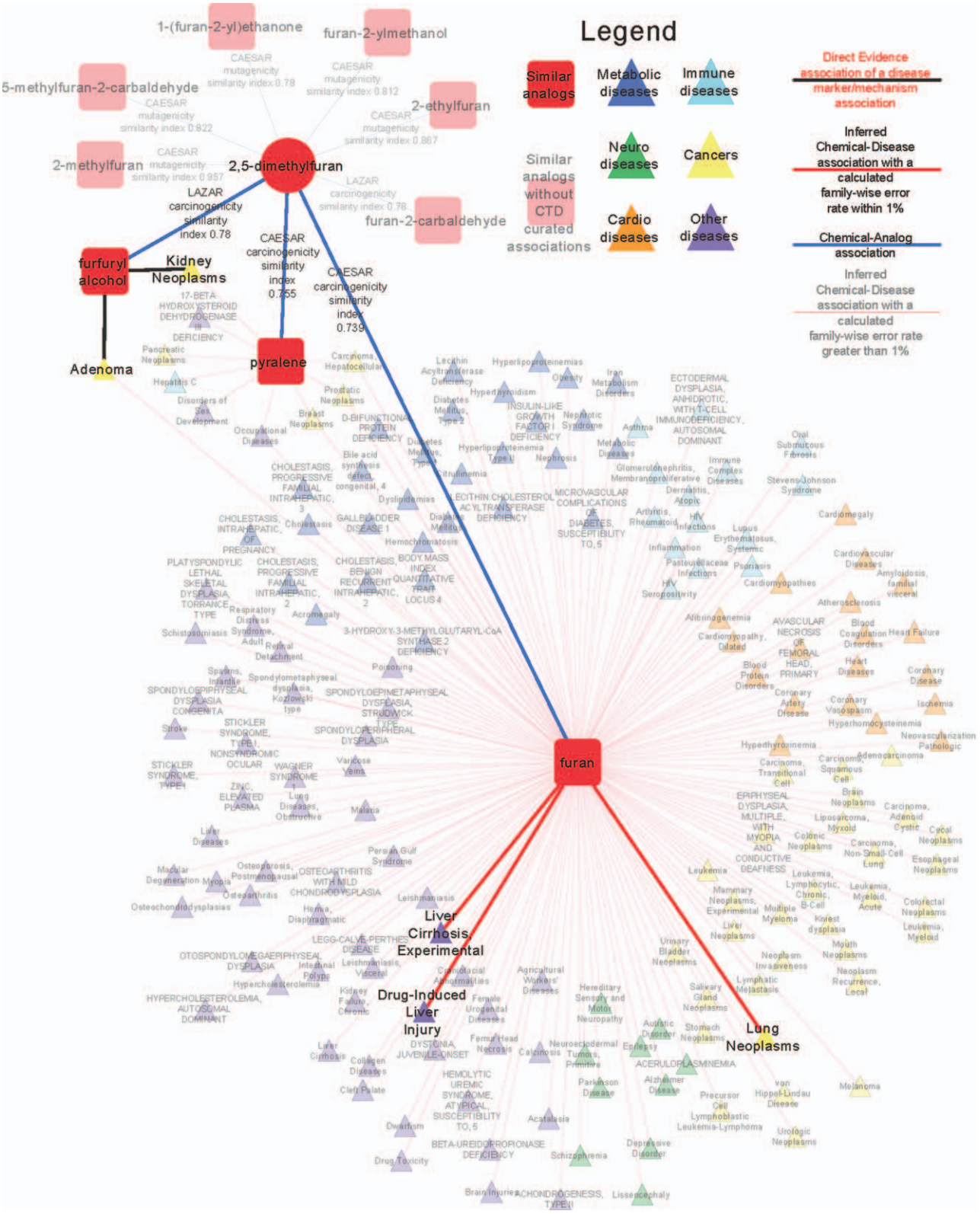

Fig. 3. The DMF data gap analysis for possible disease associations. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

TABLE I. Disease and Gene Identification for DMF Using Its Similar Analogs

Similar analog (total gene interactions)	Disease endpoint (total gene interactions)	$-\text{Log}_{10}$(adjusted *P*-value)	Overlapping genes				No. of gene
Furan (71)	1% family-wise error threshold						
	Drug-induced liver injury (48)	5.64	ALB	AMBP	C3	CXCL10	7
			CYP2E1	ENO1	PON1		
	Lung neoplasms (120)	4.18	A2M	APOC3	CCND1	CCNG1	8
			CRP	CYP2E1	GSTP2	PON1	
	Liver cirrhosis, experimental (121)	4.15	AMACR	CES3	CYP2C	HSD17B2	8
			IGF1	MT1A	OBP3	TIMP1	
	20% family-wise error threshold						
	Pasteurellaceae infections (3)	0.82	CP	CRP			2
Furfuryl alcohol	$\text{Adenoma}_{\text{M}}$						
	$\text{Kidney neoplasms}_{\text{M}}$						

Direct evidence type ''M'' indicates a correlated marker/mechanism association and between the chemical and disease as reported from the CTD as of the April 7, 2011 update. The *P*-values are calculated using the Fisher's exact test and then adjusted for multiple testing: the Fisher's exact test-calculated *P*-value * 2 (total no. of chemicals considered) * 2,203 (total no. of diseases with gene interactions possible), where the two chemicals considered are furan and pyralene. The adjusted *P*-values were used to assign significance according to the 1, 5, 10, and 20% family-wise error threshold.

the combustion species from Wu et al. [2009] but disputed a few misidentified chemicals. These disputes may be attributed to experimental differences and detection sensitivity because Simmie and Metcalfe [2011] proposed a decomposition reaction that also includes secondary reactions that have yet to be confirmed. On the basis of both reports, we have included the majority of the currently accepted combustion species for DMF in our study. Our findings revealed numerous possible PBT effects from individual combustion intermediates. In reality, however, these combustion intermediates are generated in differing proportions as a vaporized mixture [Lifshitz et al., 1998], the effects of which cannot be analyzed by the current method.

The burning conditions (such as a closed-engine turbine or open campfire) and an engine's capability to complete combustion will influence the spectrum of combustion species. Some DMF combustion intermediates are high molecular weight and reminiscent of prominent VOCs from petroleum gasoline combustion and tobacco smoke. Although the relative amounts and concentrations of each species generated are unknown, a recent study suggests that DMF, when applied in a dual-injection spark ignition engine, may be combusted more efficiently for high energy yield and more thoroughly into small-molecular-weight products [Daniel et al., 2010; Zhong et al., 2010]. Consequently, in this application, high-molecular-mass emission compounds may not be as prevalent [Daniel et al., 2010; McKone et al., 2011]. On the other hand, under nonideal conditions such as poor engine performance, a fuel spill in a hot environment, or a fuel tank explosion, combustion products may vary and all stable combustion intermediates may be released into the environment.

Environmental Impact of Reactive Free Radicals

In addition to the 49 DMF combustion intermediates discussed above [Wu et al., 2009], reactive free radicals are present in the combustion mixtures (Supporting Information Table 1). Free radicals are generally assumed to be highly transient and super-reactive in nature. However, such radicals could probably form environmentally persistent free radicals (EPFRs), possibly by reacting with environmental metals and particulate matter [Truong et al., 2010]. EPFRs are stabilized radicals that are not subjected to normal environmental oxidation and degradation [Aschmann et al., 2011]. Because of software limitations, reactive radicals could not be analyzed in this study, but further investigations of their health and environmental impacts are necessary for the comprehensive toxicity analysis of DMF's life cycle as a biofuel.

Environmental Impacts of Other Emissions Products

The chemicals of concern shown within Supporting Information Table 1 may not represent the full spectrum of emissions products. Although it is well reported that fuel combustion typically yields small-molecular-weight chemicals such as carbon monoxide, carbon dioxide, oxides of nitrogen, and formaldehyde [Fujita et al., 2011], only some of these compounds were detected in DMF thermal decomposition [Lifshitz et al., 1998] and in the evaluation of DMF within direct-injection spark-ignition engines [Daniel et al., 2010; Zhong et al., 2010]. Chemicals such as formaldehyde may be downstream tailpipe products of secondary reactions with methyl radical, ethenyl radical, and cyclopropane with environmental oxygen [Klein and Schoen, 1958].

DMF and Its Analog-Related Disease Predictions

Although there is no evidence of a direct association between DMF and genes or diseases, analysis of the structurally similar analogs furan and furfuryl alcohol has illuminated a variety of potential disease associations. A

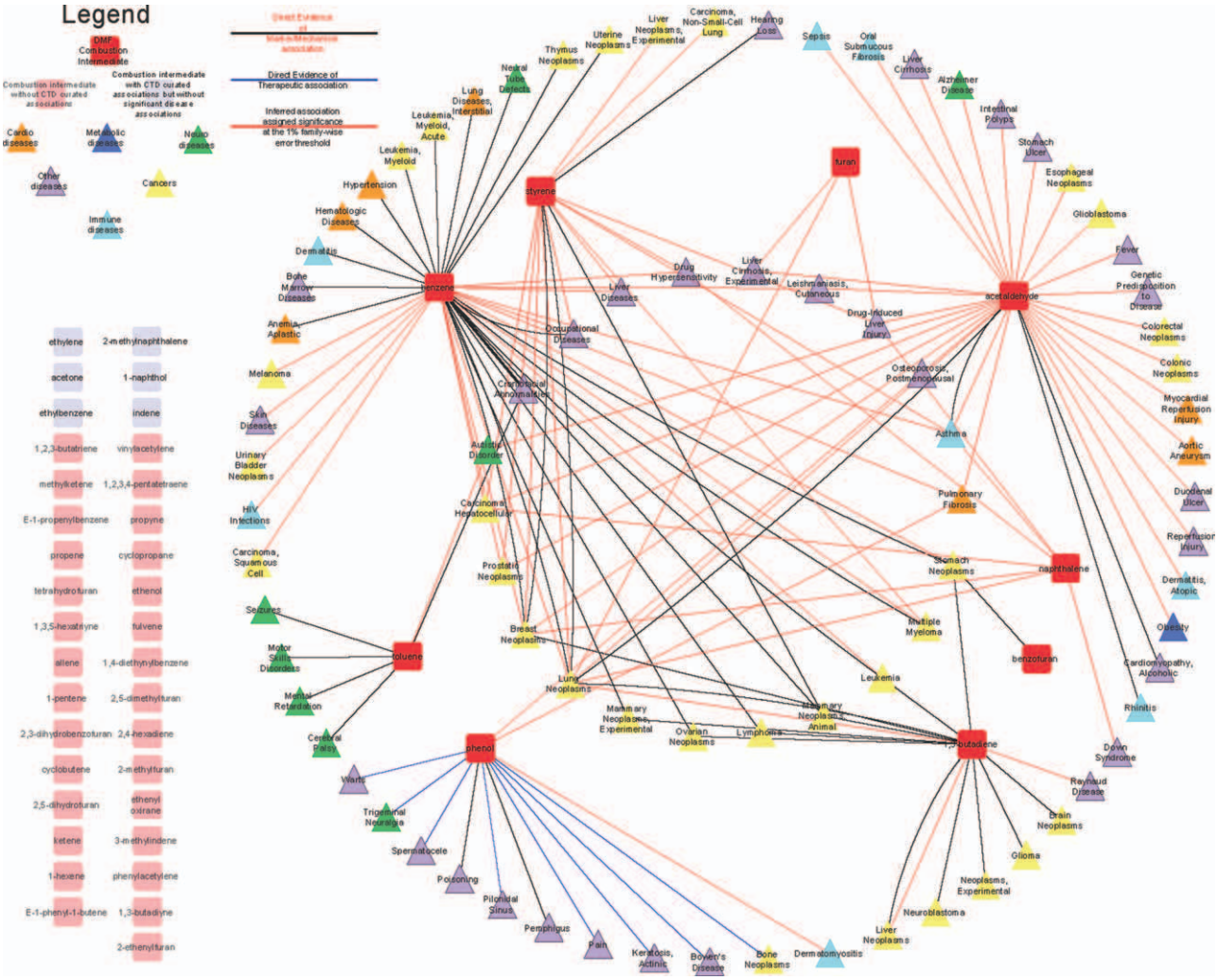

Fig. 4. All disease predictions for DMF combustion intermediates to the 1% family-wise error threshold. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

comparison of DMF's systemic circulation with that of its similar analogs further supports these predictions as probable DMF exposure-related disease associations.

DMF has been detected within human airways, systemic blood, and excreted urine from first- and second-hand tobacco smokers [Perbellini et al., 2003], suggesting that it reaches disease-relevant organ systems. However, a ventilation study in which mongrel dogs were acutely exposed to cigarette smoke showed that DMF has a relatively lower retention rate within lung airways than 2-methylfuran and even lower than furan [Egle and Gochberg, 1979]. This difference reflects that DMF will likely be absorbed into systemic circulation, whereas airway-lingering species like furan may generate lung disease predictions. Data have not been reported on the respiratory response during chronic exposure to DMF, 2-methylfuran, and furan, but it would be informative to see how mammalian physiology responds to these chemicals over an extended period of time. A study conducted by Fromowitz et al. [2012] using a mouse bone marrow micronucleus assay suggests that DMF may be genotoxic to bone marrow without metabolic activation. An increased micronucleus frequency in the peripheral blood lymphocytes of humans has been shown to be predictive of cancer [Bonassi et al., 2006].

An analysis of the structural analogs identified potentially relevant genes and pathways. For instance, furan is a known hepatotoxicant; hence, many of the identified gene interactions are involved in liver-related diseases [Huang et al., 2004]. The significant disease associations from furan are supported by 21 furan–genes interactions, as annotated within Supporting Information Table 3. Each of these individual genes may be of interest for further research onto DMF's modes of action. In addition, a subchronic study found that $p53^{+/-}$ transgenic male mice exposed to furfuryl alcohol by inhalation developed kidney neoplasms, whereas male rats developed nose adenoma [Spalding et al., 2000]. An increase in micronuclei

resulting from DMF exposure, as mentioned above [Fromowitz et al., 2012], could result in altered expression of tumor suppression genes and other genes involved in carcinogenesis.

The high degree of structural similarity (>70%) between DMF and furan, furfuryl alcohol, and pyralene suggests that DMF and these analogs may have similar modes of action. Thus, findings in this study suggest that several target organs, genes, and cellular pathways associated with DMF analogs may be relevant to DMF toxicity. Even so, the proposed associations should still be critically analyzed as discrepancies may be found.

Disease Predictions for DMF Combustion Intermediates

As discussed above, the vapor resulting from DMF combustion is a mixture of compounds. The central ring in Figure 4 represents diseases that may have increased risk factors due to statistically significant associations from multiple combustion intermediates. Being that 40 out of 49 combustion intermediates do not have toxicity or gene data, it is recommended that each of these chemicals should be further investigated.

Advantages of Our Computational Approach

Although various QSAR tools can identify similar analogs, both LAZAR and CAESAR use search algorithms that have been published and validated to find similar chemicals within their training sets for endpoints of concern. Thus, the outputs from LAZAR and CAESAR are more likely to have toxicological data and associations [Helma, 2006; Fjodorova et al., 2010]. CTD provides the gene interactions necessary to generate comprehensive hypotheses and guide testing for potential marker/mechanism association, therapeutic associations, etc. In addition, PBT profiler, a widely used prediction tool, can quantify and predict a chemical's fate in the environment, and these predictions can be evaluated to prioritize chemical hazards.

These four software tools—LAZAR, CAESAR, CTD, and PBT profiler—in coordination with database resources (e.g., ACToR, PubChem, and ToxNet), constitute an approach to perform early quantitative and qualitative assessment for chemicals that have limited data. As updates to the training sets and database records are performed, former predictions may shift, but this also provides a sustainable scheme to advance chemical screening as new chemical toxicity data are discovered and incorporated. Although in vivo studies are currently the most widely accepted form of toxicological data for a particular chemical, extrapolating readily available data from similar chemical analogs and predicted downstream toxicities can help reduce the need for expensive animal studies to gain preliminary data for a chemical's life-cycle assessment.

Limitations of the Computational Approach

Although CTD provides the gene interactions necessary for forming toxicity predictions, the gene interaction data may be biased toward the chemicals and diseases that are more commonly investigated. Rarer diseases may not be as well represented; therefore, the gene interaction information may be scarce or unavailable, making those diseases more difficult to investigate.

Two noticeable issues arise from the PBT profiler QSAR predictions for aquatic toxicity: (1) the timeframe for chronic toxicity exposure is not clarified and (2) the fish species used in this model are not clearly described so the values may be subject to species-specific variation. Although the PBT profiler predictions cannot be directly translated into experimental use, the predicted values can assist in qualitative chemical-structure-based comparisons—useful for chemical prioritization. For instance, chemicals with high BCFs are more prone to retain within living organism and could be subject to biomagnification, producing problems for species higher in food chain.

CONCLUSION

Using a computational toxicology approach to assess the life-cycle impact of DMF, we identified potential adverse biological and environmental impacts of DMF, particularly for its combustion intermediates. Multiple potential associations with disease were predicted. Analysis of structural analogs of DMF revealed 21 genes that are altered by furan and may be potential targets of DMF and play roles in its biological effects. Although this approach is inadequate as a stand-alone risk assessment and there are still data gaps for many of the chemicals analyzed, our data suggest that DMF combustion intermediates pose a much broader range of hazards than DMF itself, and that both should be further investigated.

ACKNOWLEDGMENTS

The authors thank Prof. Martyn Smith, Dr. Cliona McHale, and Ms. Sarah Daniels for their critical comments, and Mr. Benjamin King for insightful peer editing. Their special gratitude goes to Prof. Dale Johnson and Mr. Brandon Gaytán for their early involvement in this project that originated from a Practical Toxicology class assignment, taught by Prof. Zhang at UC Berkeley and supported by the US and NorCal Society of Toxicology.

This work was partially supported by the Northern California Center for Occupational and Environmental Health, and a grant (R01ES017452) from the National Institute for Environmental Health Sciences (NIEHS) to Dr. Zhang. Jimmy Phuong, a previous UC Berkeley undergraduate, was supported in part by the Council for Education and Research on Toxics.

REFERENCES

Alonso M, Godayol A, Antico E, Sanchez JM. 2010. Assessment of environmental tobacco smoke contamination in public premises: Significance of 2,5-dimethylfuran as an effective marker. Environ Sci Technol 44:8289–8294.

Aschmann SM, Nishino N, Arey J, Atkinson R. 2011. Kinetics of the reactions of OH radicals with 2-and 3-methylfuran, 2,3-and 2,5-dimethylfuran, and E-and Z-3-hexene-2, 5-dione, and products of OH+ 2, 5-dimethylfuran. Environ Sci Technol 45:1859–1865.

Bonassi S, Znaor A, Ceppi M, Lando C, Chang WP, Holland N, Kirsch-Volders M, Zeiger E, Ban S, Barale R. 2006. An increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of cancer in humans. Carcinogenesis 28:625–631.

CDC (Centers for Disease Control and Prevention). 2009. Fourth National Report on Human Exposure to Environmental Chemicals. Available at:http://www.cdc.gov/exposurereport/.

Daniel R, Tian G, Xu H, Wyszynski ML, Wu X, Huang Z. 2010. Effect of spark timing and load on a DISI engine fuelled with 2, 5-dimethylfuran. Fuel 90:449–458.

Egle JL, Gochberg BJ. 1979. Retention of inhaled 2-methylfuran and 2, 5-dimethylfuran. Am Ind Hyg Assoc J 40:866–869.

Fjodorova N, Vračko M, Novič M, Roncaglioni A, Benfenati E. 2010. New public QSAR model for carcinogenicity. Chem Cent J 4:1–15.

Fromowitz M, Shuga J, Wlassowsky AY, Zhang L, Smith MT. 2012. Studies on the genotoxicity of 2,5-dimethylfuran, a potential biofuel. Environ Mol Mutagen 53.

Fujita EM, Campbell D, Zielinska B, Arnott W, Chow J. 2011. Concentrations of air toxics in motor vehicle-dominated environments. Res Rep Health Eff Inst 156:3–77.

Helma C. 2006. Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Mol Divers 10:147–158.

Huang Q, Jin X, Gaillard ET, Knight BL, Pack FD, Stoltz JH, Jayadev S, Blanchard KT. 2004. Gene expression profiling reveals multiple toxicity endpoints induced by hepatotoxicants. Mutat Res 549:147–167.

Iwata M, Takeuchi Y, Hisanaga N, Ono Y. 1983. Changes of n-hexane metabolites in urine of rats exposed to various concentrations of n-hexane and to its mixture with toluene or MEK. Int Arch Occup Environ Health 53:1–8.

Iwata M, Takeuchi Y, Hisanaga N, Ono Y. 1984. Changes of n-hexane neurotoxicity and its urinary metabolites by long-term co-exposure with MEK or toluene. Int Arch Occup Environ Health 54:273–281.

Klein R, Schoen LJ. 1958. Role of formaldehyde in combustion. In:Literature of the Combustion of Petroleum, Vol. 20. Washington, DC:ACS Publications. pp 8–68.

Lifshitz A, Tamburu C, Shashua R. 1998. Thermal decomposition of 2, 5-dimethylfuran. Experimental results and computer modeling. J Phys Chem A 102:10655–10670.

Luque R, Herrero-Davila L, Campelo JM, Clark JH, Hidalgo JM, Luna D, Marinas JM, Romero AA. 2008. Biofuels: A technological perspective. Energy Environ Sci 1:542–564.

McDonald JH. 2009. Fisher's Exact Test of Independence, 2nd ed. Baltimore, MD:Sparky House Publishing. pp 70–75.

McKone T, Nazaroff W, Berck P, Auffhammer M, Lipman T, Torn M, Masanet E, Lobscheid A, Santero N, Mishra U. 2011. Grand challenges for life-cycle assessment of biofuels. Environ Sci Technol 45:1751–1756.

Mount Desert Island Biological Laboratory. 2011. CTD. Comparative Toxicogenomics Database. Available at:http://ctd.mdibl.org/.

NCBI (National Center for Biotechnology Information). 2011. PubChem. PubChem Compound Database. Available at:http://pubchem.ncbi.nlm.nih.gov/.

Perbellini L, Princivalle A, Cerpelloni M, Pasini F, Brugnone F. 2003. Comparison of breath, blood and urine concentrations in the biomonitoring of environmental exposure to 1, 3-butadiene, 2, 5-dimethylfuran, and benzene. Int Arch Occup Environ Health 76:461–466.

Román-Leshkov Y, Barrett CJ, Liu ZY, Dumesic JA. 2007. Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates. Nature 447:982–985.

Schmidt LD, Dauenhauer PJ. 2007. Chemical engineering: Hybrid routes to biofuels. Nature 447:914–915.

Simmie JM, Metcalfe WK. 2011. Ab initio study of the decomposition of 2, 5-dimethyl furan. J Phys Chem A 115:8877–8888.

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. 2011. Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics 27:431–432.

Spalding JW, French JE, Stasiewicz S, Furedi-Machacek M, Conner F, Tice RR, Tennant RW. 2000. Responses of transgenic mouse lines p53+/− and Tg● AC to agents tested in conventional carcinogenicity bioassays. Toxicol Sci 53:213–223.

Truong H, Lomnicki S, Dellinger B. 2010. Potential for misidentification of environmentally persistent free radicals as molecular pollutants in particulate matter. Environ Sci Technol 44:1933–1939.

U.S. Environmental Protection Agency. 1999a. Category for Persistent, Bioaccumulative, and Toxic New Chemical Substances.U.S. EPA Federal Register. Available at:http://www.epa.gov/fedrgstr/EPA-TOX/1999/November/Day-04/t28888.htm.

U.S. Environmental Protection Agency. 1999b. Persistent Bioaccumulative Toxic (PBT) Chemicals; Lowering of Reporting Thresholds for Certain PBT Chemicals; Community Right-to-Know Toxic Chemical Reporting. U.S. EPA Federal Register. Available at:http://www.epa.gov/fedrgstr/EPA-WASTE/1999/October/Day-29/f28169.htm.

Wu X, Huang Z, Yuan T, Zhang K, Wei L. 2009. Identification of combustion intermediates in a low-pressure premixed laminar 2, 5-dimethylfuran/oxygen/argon flame with tunable synchrotron photoionization. Combustion Flame 156:1365–1376.

Zeiger E, Anderson B, Haworth S, Lawlor T, Mortelmans K. 1992. Salmonella mutagenicity tests. V. Results from the testing of 311 chemicals. Environ Mol Mutagen 19:2–141.

Zhong S, Daniel R, Xu H, Zhang J, Turner D, Wyszynski ML, Richards P. 2010. Combustion and emissions of 2, 5-dimethylfuran in a direct-injection spark-ignition engine. Energy Fuels 24:2891–2899.

Accepted by—
R. Snyder