

Introducing Membrane Transport Energy into the Design of Sustainable Chemicals against Cytotoxicity

Longzhu Q. Shen,^{§,†} Ji Nan,^{§,‡,¶} and Benzhuo Lu^{*,‡,¶}

[†]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, United Kingdom

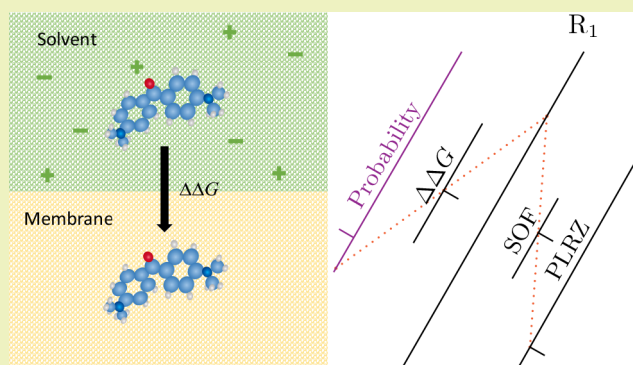
[‡]LSEC, National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

[¶]University of Chinese Academy of Sciences, 19 Yuquan Road, Beijing 100049, China

S Supporting Information

ABSTRACT: Adverse outcomes associated with chemical use and spill have raised concerns over its sustainability. The traditional toxicity testing, using laboratory animals as the downstream safety check, is becoming less viable due to the economical and ethical liabilities. Green chemistry proposed an alternative strategy to attain chemical sustainability: designing chemicals to maximize their intrinsic sustainability and thus minimize their hazardous risk. Following decades of progress, there is still a need to develop new metrics to quantify the level of chemical sustainability. In this report, we developed a new double functional tool capable of estimating the sustainability probability of a chemical and designing new chemicals to meet a desired sustainability probability. This tool was built upon the Naive Bayesian algorithm with the design variables stemming from three sources. Molecular softness and polarizability were derived from density functional theory (DFT), and membrane transport free energy was computed using our in-house developed finite element algorithm. Model validation against the cytotoxicity measured in the U.S. EPA Toxicity ForeCaster (ToxCast) database (tested up to 100 μ M) yielded a score of 0.82 for the area under the curve (AUC) of the receiver operating characteristic (ROC). On the basis of this model, we constructed the assessment tool with the dual capabilities of prediction and design.

KEYWORDS: DFT, Cytotoxicity, Membrane, Finite element, Poisson–Boltzmann, Geometry, Sustainability



INTRODUCTION

Chemicals and chemical products have a broad impact on the global economy and aspects of human health. It has been realized that unintended biological and environmental consequences can be associated with certain chemical use and disposal. The conventional approach to checking chemical safety is to test the toxic effects using laboratory animals such as rats, mice or dogs. These studies are expensive (millions of dollars per chemical) and require the use of large numbers of animals. As a result, many chemicals are put on the market with little or no toxicity testing.¹ It has been estimated that approximately 83% of chemicals in commerce lack safety data.²

Green chemistry³ presented an alternative proposal toward chemical sustainability. It introduced the idea of assessing safety at the initial stage of molecular design so that the redesigned materials that make up the basis of our society and economy can be intrinsically sustainable.⁴ After 25 years of development of green chemistry, progress has been made toward achieving chemical sustainability. However, a strong need is still being recognized⁵ to develop new sustainable metrics to quantify sustainability for chemicals in both assessment^{6–8} and

design.^{9–12} In this report, we have developed a new sustainability assessment tool that can estimate the likelihood of a chemical of incurring cytotoxicity and that can guide safer chemicals design to reduce their probabilities of causing cytotoxicity.

There are multiple reasons behind the choice of cytotoxicity as a meaningful toxicity end point. We present three rationales for this. First, cytotoxicity usually stands at the front line of the safety tests in the early stage of drug development.¹³ Second, cytotoxicity can provide information for inferring *in vivo* toxicity.^{14–16} Third, cytotoxic effects have been recognized for their value in discovering/developing chemotherapeutic agents.^{17–19} Therefore, it is meaningful to assess sustainability for commercial chemicals in terms of cytotoxicity. To evaluate the likelihood of a chemical to incur cytotoxicity, we applied a dichotomous method to split this study's chemical data set into two classes ("active" vs "inactive" with regards to the

Received: September 28, 2017

Revised: December 2, 2017

Published: December 22, 2017

experimental testing concentration range, up to 100 μM). The binary classification decision was based on the outcomes of *in vitro* high throughput screening (HTS) cellular assays. HTS was described by the U.S. National Research Council (NRC) in their report on “Toxicity Testing in the 21st Century”²⁰ (“Tox21”) as a paradigm-shift toxicity testing strategy. The U.S. National Toxicological Program (NTP), the U.S. Environmental Protection Agency (EPA) and the NIH National Center for Advancing Translational Sciences (NCATS) collaboratively forged a research partnership to implement the “Tox21” vision. This “Tox21” partnership uses HTS methods to test thousands of chemicals in a wide variety of cells, pathways and technologies, relevant to many aspects of chemical toxicity.^{21–26}

In developing our sustainability assessment tool, we paid special attention to the choice of the design variables that went into the statistical model. A novel aspect of this research was the introduction of the membrane transport energy as a new effective metric in the predictive model and we developed an in-house algorithm to compute this quantity. We deliberately used this energy to capture the physical cost of transporting a chemical into the cellular membrane. This physical action is particularly important because the ability of chemicals permeating through cell membranes has long been linked to molecular bioavailability and toxicity.^{14,27–30} Minimizing the ability of chemicals to cross biological interfaces has been a fundamental challenge for chemists in designing chemicals for reduced hazardousness.³¹ In combination with two other variables, molecular softness and polarizability, we established a model with clear physical descriptions of chemicals perturbing the biological system at the electronic and thermodynamic levels. We foresee that the meaningfulness and usefulness of this new assessment tool will make it a valuable addition to the sustainability metrics.

METHODS

Data Source and Selection. We applied three criteria to select a subset of chemicals in the U.S. EPA Toxicity ForeCaster (ToxCast) program Phase I and II chemical library.³²

1. Molecular weight <1000. The bioavailability for chemicals with molecular weight greater than 1000 is negligible.³³
2. Single compound with a definite structure, excluding geometrical and optical isomers and mixtures. Isomeric or compositional mixtures incur toxicity via complex interactions. Our model was not designed to treat this group of chemicals and thus we excluded them from consideration.
3. Containing no metal elements. The mechanisms by which metal-containing chemicals cause toxicity are far more complex than organic molecules. We thus defer the treatment of metal complexes to a later stage.

After the three-layer filter, we obtained about 1000 chemicals and subsequently categorized them into two classes, “active” and “inactive” (with respect to the experimental concentration range, up to 100 μM) based on a combinatorial score derived from the 37 cytotoxicity related assays available in the ToxCast database.³⁴ Detailed assay descriptions have been published in our previous work.¹¹ Cytotoxicity at concentrations above 100 μM would not be observed from the current laboratory results. Therefore, we are strictly modeling “cytotoxicity below 100 μM ”. The resultant two classes of chemicals were in balance with a membership ratio of 0.96. All chemical data were then evenly split into a training and a testing set in order to perform the cross-validation and external evaluation. The members in the training and testing sets were equally likely drawn from each class. The selected chemicals were first desalted using the open source chemistry toolbox OpenBabel.³⁵ Afterward, 3D structures with the

lowest energy of them chemicals were generated using ChemAxon Marvin calculator plugins.³⁶

Generation of Design Variables. We employed molecular softness (SOF), polarizability (PLRZ) and membrane transport free energy ($\Delta\Delta G$) to quantitatively infer the perturbation brought by a chemical to the exposing cells. SOF and PLRZ³⁷ were derived from quantum mechanical computations at the DFT level. Boese and Martin’s τ -dependent hybrid functional³⁸ and basis set 6-31+G(d), implemented in Gaussian 09 rev. D.01,³⁹ were used to fully optimize the molecular geometries. Vertical (ionization potential) IP and electron affinity (EA) were calculated in the absence of any solvent. SOF was then calculated using the following formula.⁴⁰

$$\text{SOF} = 1/(\text{IP} - \text{EA}) \quad (1)$$

$\Delta\Delta G$ was computed using an in-house developed finite element algorithm⁴¹ that included biological membranes in the implicit solvation model.⁴² A schematic representation of the system is shown in Figure 1. The solvated biomolecular system occupies a

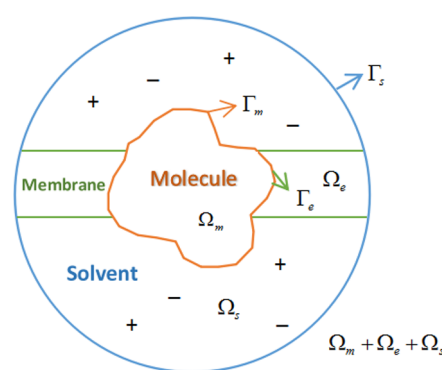


Figure 1. Implicit solvation model including biological membranes.

domain Ω enclosed by a smooth boundary Γ_s . The solute (chemical) region is represented by Ω_m , the solvent region by Ω_s and the membrane region by Ω_e . Γ_m is the dielectric interface between the chemical and the solvent, and Γ_e is the dielectric interface between the chemical and the cellular membrane. It is worth noting that almost all the chemicals in this study were immersed in the membrane (bilayer thickness 30 Å). ϵ_s , ϵ_m and ϵ_e are the relative permittivities in Ω_m , Ω_s and Ω_e , respectively. Given the above setting, the nonlinear Poisson–Boltzmann equation⁴³ in Ω takes the form of eq 2

$$-\nabla \cdot \epsilon \nabla \phi - \lambda \sum_{i=1}^K c_i e^{-\beta q_i \phi} q_i = \rho^f \quad (2)$$

where the characteristic function $\lambda = 0$ in Ω_m or Ω_e and $\lambda = 1$ in Ω_s , $\beta = \frac{1}{k_B T}$ is the reciprocal of the Boltzmann energy, k_B is the Boltzmann constant, T is the absolute temperature. ϵ is a spatial-dependent dielectric coefficient, ϕ is the electric potential, a fixed charge distribution $\rho^f = \sum_j q_j \delta(x - x_j)$ is the cumulation of point charge in the solute region, and q_j is the singular charge located at x_j . Solving eq 2⁴⁴ yields the ΔG_{ele} , which accounts for the electrostatic contribution to the solvation free energy as shown in eq 3.

$$\Delta G_{\text{sol}} = \Delta G_{\text{ele}} + \Delta G_{\text{np}} \quad (3)$$

The remaining term, the nonpolar part ΔG_{np} in eq 2 was calculated by the solvent accessibility (SA) model^{45–47} with a surface tension parameter γ expanded over the atomic coordinates:⁴⁸

$$\Delta G_{\text{np}} = \gamma \sum_{i=1}^N S(z_i) \text{SA}_i \quad (4)$$

where SA_i is the solvent-accessible surface area of the i th atom, γ is an empirical surface tension parameter, and $S(z)$ introduces the variation of the surface tension along the z direction in the membrane

environment. The detailed solution procedure plus the ionic strength influence are given in the [Supporting Information](#) (S.I.).

We take $\Delta\Delta G_{\text{sol}}$ as the solvation energy difference when transferring a chemical from the solvent to the membrane environment.

$$\Delta\Delta G_{\text{sol}} = \Delta G_{\text{sol}}^{\text{membrane}} - \Delta G_{\text{sol}}^{\text{solvent}} \quad (5)$$

where $\Delta G_{\text{sol}}^{\text{membrane}}$ and $\Delta G_{\text{sol}}^{\text{solvent}}$ denote the solvation energy of a chemical in the solvent and in the membrane, respectively.

The lengths of the semiprincipal axes were calculated by approximating the shape of a chemical with an ellipsoid as shown in [Figure 2](#).

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad (6)$$

where the semiprincipal axes a , b and c obey the relation $a > b > c$.

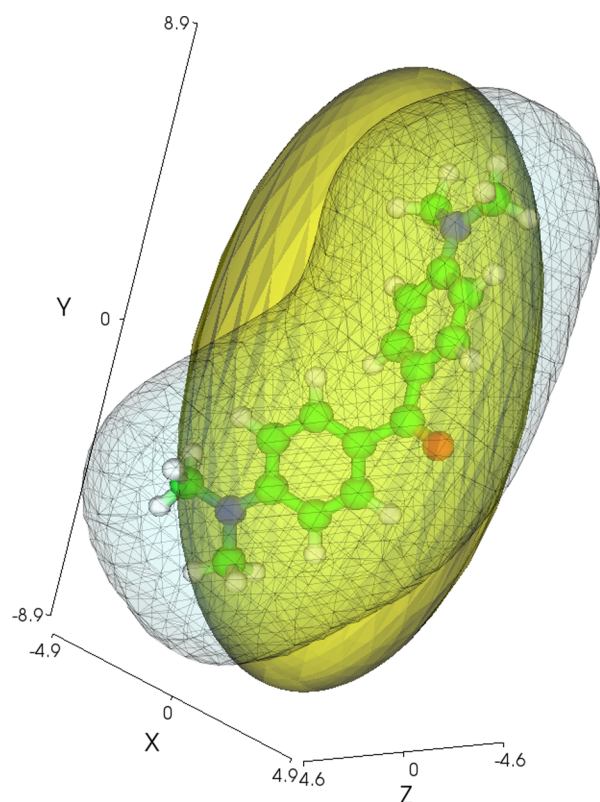


Figure 2. Illustration of the ellipsoid (yellow) enclosing a molecule (ball-and-stick model in the center). The meshed light blue profile represents the molecular surface. Three semiprincipal axes X , Y and Z have their origin $(0,0,0)$ coinciding with the center of the ellipsoid.

We used the eigenvalues of the covariance matrix of a set of uniformly sampled surface coordinates of a chemical to estimate the lengths of the semiprincipal axes. Derivations are available in the [S.I.](#)

Model Construction. The Naive Bayesian classifier⁴⁹ is an effective probabilistic classifier based on Bayes' theorem (eq 7) with independence assumptions between the covariates.

$$\pi(\theta|X) = \frac{\pi(X|\theta)\pi(\theta)}{\int \pi(X|\theta)\pi(\theta)d\theta} \quad (7)$$

where θ denotes the parameters, X denotes the random variables, $\pi(\theta|X)$ denotes the posterior probability, $\pi(\theta)$ denotes the prior probability ($\pi(\theta) \approx 0.5$ in this study), $\pi(X|\theta)$ denotes the likelihood function, and the denominator integral denotes the marginal likelihood. In this study, the parameter θ represents the class identifier and X represents the design variables obtained from the previous

section. Our interest was to calculate the posterior probability to represent chemical sustainability.

To select the variables with high predictive power and less dependence between each other, we examined the ROC AUC^{50,51} ([Figure 3](#)) and mutual information (MI) (eq 8) matrix ([Figure 4](#)).

$$I(X, Y) = \iint p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy \quad (8)$$

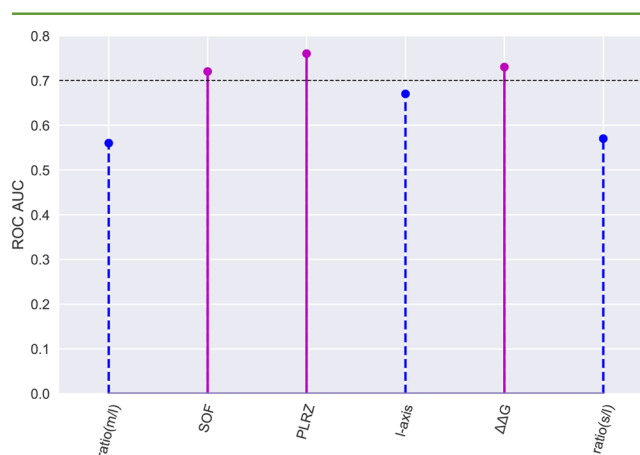


Figure 3. ROC AUC plots for design variables.

where $p(x, y)$ is the joint probability density function of X and Y , and $p(x)$ and $p(y)$ are the marginal probabilities.

	SOF	PLRZ	$\Delta\Delta G$	l-axis	ratio(m/l)	ratio(s/l)
SOF	1.00	0.40	0.26	0.26	0.17	0.20
PLRZ	0.40	1.00	0.66	0.51	0.24	0.24
$\Delta\Delta G$	0.26	0.66	1.00	0.48	0.24	0.28
l-axis	0.26	0.51	0.48	1.00	0.58	0.61
ratio(m/l)	0.17	0.24	0.24	0.58	1.00	0.81
ratio(s/l)	0.20	0.24	0.28	0.61	0.81	1.00

Figure 4. Mutual information matrix for design variables.

From the ROC AUC plot ([Figure 3](#)), we noticed that the predictive power of the variables follow the descendant sequence, $\text{PLRZ} > \Delta\Delta G > \text{SOF} > \text{l-axis} > \text{m/l} \approx \text{s/l}$. The ROC AUCs for three variables PLRZ, SOF and $\Delta\Delta G$ are above 0.7. The same three variables ([Figure 4](#)) also illustrate relatively low dependence between each other. (The categorized histograms for each variable is provided in the [Figure S1](#).) Therefore, we selected them as the independent variables for the predictive model.

The model construction, data analysis and the graphical visualization in this study were coded with the Python programming language,^{52,53} libraries and packages.^{54–61}

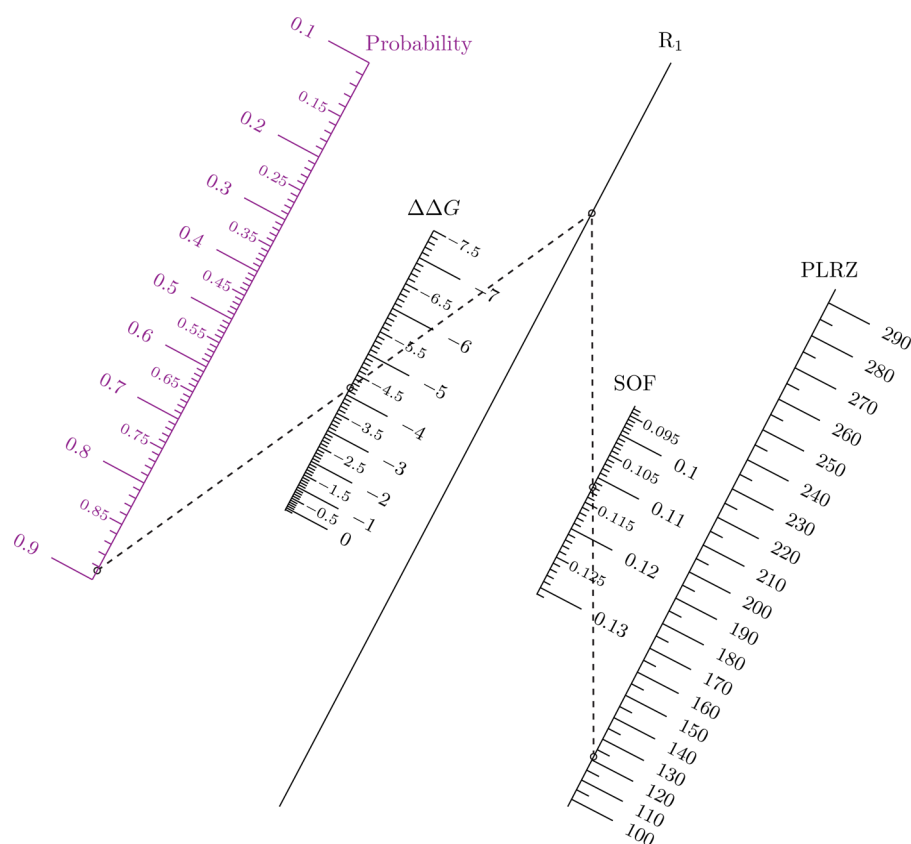


Figure 5. Assessment diagram to predict the likelihood of chemicals inducing cytotoxicity and to design chemicals to meet desired probabilities not incurring cytotoxicity. The probability axis is shown in purple. The design variables axes are printed black. R1 is an auxiliary axis. The physical units for the variables on the diagrams are kcal/mol for $\Delta\Delta G$, 1/eV for SOF and cm^3/mol for PLRZ.

RESULTS AND DISCUSSION

Model Evaluation and Mechanistic Rationale. The performance of the Naive Bayesian model was evaluated using the ROC AUC. It statistically measures the probability of a randomly selected positive case being ranked higher than a randomly selected control case.⁶² It has been widely used to quantify the overall discriminative ability of statistical classifiers. It has the excellent property of being applicable to scenarios with skewed class distributions and unequal classification error costs.⁶³ A value of 0.5 for ROC AUC indicates a random classification. Values above 0.5 and up to 1 rank the increasing discriminative power of a classifier.

In our study, we applied internal cross-validation and external validation to evaluate the model performance. To this end, we first split the total data into two groups (training and testing) with each group containing approximately 500 members. Then, we further partitioned the training group into 10 subgroups evenly. Next, we used the chemicals from 9 combined subgroups to train the model and the remaining one subgroup to test the model. We did so in a rotatory fashion among all the 10 subgroups so that every possible combination was considered. The ROC AUC for the cross-validation is 0.77 ± 0.06 . After completing the internal cross validation, we subjected the model to the other ~ 500 chemicals in the testing group (not seen by the model yet). The ROC AUC for this external validation is 0.82. The consistent performance between the training and testing sets indicates that the model has been properly fitted. This validates the use of this statistical model to construct the sustainability assessment diagram.

We attributed the satisfactory model performance primarily to the careful choice of the design variables that allowed for a meaningful mapping between the chemical and toxicological space. The mechanisms by which chemicals to incur cytotoxicity can be rather complex. Broadly speaking, three levels of interactions need to be taken into account.

First, in many instances, chemicals need to go across the cellular membrane to disrupt biological pathways or interfere with the functions of critical biological molecules. To transport chemicals inside the cellular membrane, there exist at least three known physical scenarios: passive diffusion, facilitated diffusion and active transportation. In our model, we focused on the case of passive diffusion and treated it at the thermodynamic level. The resultant $\Delta\Delta G$ term is the free energy difference between a chemical within and outside the membrane. The negative sign of this term indicates that transporting a chemical into the cellular membrane is energetically favored whereas the positive sign indicates that chemicals gain energy rewards if staying outside the membrane. In a sense, $\Delta\Delta G$ can be viewed as the theoretical ground of lipophilicity. Lipophilicity has long been recognized in medicinal chemistry as a proved parameter to infer bioavailability and drug toxicity.⁶⁴ To estimate lipophilicity, $\log P$ (water/octanol partition coefficient) has been used as a surrogate in computer-aided drug design and predictive toxicology fields. Various computational methods have been developed to calculate the $\log P$ values. The consistency and discrepancy among the predicted $\log P$ values using different methods are being actively researched in literature.⁶⁵ Our approach differs by starting from the fundamental thermodynamic principle instead of identifying

an analogue. $\Delta\Delta G$ physically reflects the energy change during the membrane transport process. Parameters in eq 2 can be modulated to adjust the model behavior to better mimic a specific membrane environment provided with better biological knowledge. The current success of the $\Delta\Delta G$ corroborates that thermodynamic parameters are effective at depicting the propensity of a chemical being transported inside the membrane and related to the likelihood of incurring cytotoxicity. In the future, we plan to further expand the model to explore the toxicokinetic aspect of the membrane transport process and events occurring at the interface between the cellular membrane and the solvent. These physical processes are anticipated to be directly influencing the membrane transport and cellular toxicity. We have not yet foreseen any surrogates for them. We therefore argue that the proposed new parameter $\Delta\Delta G$ is a superior metric to measure the energy drive for chemicals transpassing the cell membrane.

Second, in order to incur an adverse outcome, chemicals need to interact with certain biological molecules. A molecular recognition step is necessary before an effective coupling between a disrupting chemical and a biomolecule. The physical forces responsible for this interaction are often induced electric dipole moments and dispersion forces.⁶⁶ Polarizability (eq 9) is a physical quantity that describes the relative tendency of electron cloud distortion under the influence of an external electric field. It characterizes the energy change during the course of molecular recognition.⁶⁷ We thus included polarizability as a parameter in the cytotoxicity model.

$$\alpha = p/E \quad (9)$$

where p is the dipole moment and E is the electric field.

Third, chemicals can cause cell malfunction or death through undesired chemical reactions with critical biological molecules.⁶⁸ For instance, the covalent modification of proteins, especially with the thiol groups, has been shown to trigger cellular toxicity.^{69,70} To address this phenomenon, we applied the hard–soft acid–base (HSAB) theory to estimate the tendency for chemicals to form covalent bonds, which is linked to the tendency for some chemicals to be toxic.⁷¹ Molecular softness based upon DFT provides a means to quantify this propensity.⁷²

Assessment Diagram. From the Naive Bayesian model (eq 7), we arrived at the sustainability assessment diagram (Figure 5) by following a procedure described in our earlier publication¹¹ (log odds ratio between the nontoxic vs toxic chemicals are shown in S.I. Figure S2). The application of this diagram to predict cytotoxicity or to guide safer chemicals design is fairly straightforward. The only needed action is to connect dots between the lines as shown in Figure 5. For example, if we have the prior knowledge (either from experiments or computations) about the values of all the three design variables, we can pinpoint them (−4.18 for $\Delta\Delta G$, 0.112 for SOF and 122 for PLRZ for instance) on the black axes and then link them sequentially to land at a specific estimated sustainability probability (0.89 in the example) on the purple axis. Alternatively, if we want to design a chemical with a predefined sustainability probability of not incurring cytotoxicity, we can reverse the process by connecting the dots from the purple axis to the black axes. Suppose we use the same dotted line in the figure as an illustration and start from the probability 0.89 on the purple axis. We have the free choice to go for any point on the immediate next black axis $\Delta\Delta G$ before proceeding further. We may have some knowledge about the

acceptable value range for $\Delta\Delta G$ based on our experience or certain functional needs. If that is the case, we can simply bridge the value on the probability axis (0.89) with all the acceptable values on the $\Delta\Delta G$ axis. Each of these lines can further be extended to the auxiliary axis R1 and the other two remaining black axes (SOF and PLRZ). The collection of all the possible lines represents the complete solution set that can exist for the chemicals with the desired sustainability probability score 0.89. If we do not have any knowledge about $\Delta\Delta G$, we can leave it for a moment and move on the SOF axis. We may have already obtained SOF from quantum calculations. As illustrated in the example, we fix 0.112 on the SOF axis. Then we draw a line backward between 0.112 on the SOF axis and a point on the auxiliary axis R1. This dot on R1 can help us determine the point on the previously suspended $\Delta\Delta G$ axis. We join this point on R1 and the probability point 0.89 on the purple axis by a line, and this line intersects with the $\Delta\Delta G$ axis at −4.18, which is the only possible value for $\Delta\Delta G$ in this example. Similarly, we can solve the value on the PLRZ axis by extending the line between the point on R1 and 0.112 on the SOF axis. As shown, we arrive exclusively at the point 122 for PLRZ. The dotted line shown in Figure 5 illustrates only one of many possible solutions in the solution space. By varying the standpoints on the black axes, we can obtain the complete set of possible solutions in the chemical space that can meet the predefined probability 0.89. The same procedure works for designing chemicals for any target probabilities on the purple axis.

CONCLUSIONS

In this work, we introduced $\Delta\Delta G$ as an effective metric to model the potency of chemicals to induce cytotoxicity measured by the “Tox21” assays which were tested up to 100 μM . The prediction and design diagram constructed based on $\Delta\Delta G$, SOF and PLRZ enriches the toolbox of green chemistry and sustainable assessment.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssuschemeng.7b03493.

Detailed descriptions for the algorithm to solve the nonlinear Poisson–Boltzmann equation, Detailed derivations for the estimation of the semiprincipal axes, Histograms plots for the design variables in categories (PDF)

AUTHOR INFORMATION

Corresponding Author

*B. Lu. E-mail: bzlu@lsec.cc.ac.cn. Phone: +86-10-82541904. Fax: +86-10-62542285.

ORCID

Longzhu Q. Shen: 0000-0001-5629-3007

Benzhuo Lu: 0000-0003-4159-1532

Author Contributions

[§]These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Nan and Benzhuo thank the support by National Key Research and Development Program China (2016YFB0201304) and China NSF (NSFC 21573274). Longzhu thanks the helpful discussions with Dr. Quan Wang. Longzhu thanks Dr. Sylvia Vitello, Ms. Helen Llewelyn, Dr. Agnes Bolinska, Dr. Jasmine Childs-Fegredo for the proof-reading and language revision. Longzhu acknowledges the computational support provided by Yale high performance computing platform.

■ REFERENCES

- (1) National Research Council (US). Steering Committee on Identification of Toxic and Potentially Toxic Chemicals for Consideration by the National Toxicology Program. *Toxicity Testing: Strategies to Determine Needs and Priorities*; National Academies Press: Washington DC, USA, 1984.
- (2) Hartung, T. Toxicology for the twenty-first century. *Nature* **2009**, *460*, 208–212.
- (3) Anastas, P. T.; Warner, J. C. *Green Chemistry: Theory and Practice*; Oxford University Press: New York, USA, 1998.
- (4) Beach, E. S.; Cui, Z.; Anastas, P. T. Green Chemistry: A design framework for sustainability. *Energy Environ. Sci.* **2009**, *2*, 1038–1049.
- (5) Tickner, J. A.; Becker, M. Mainstreaming green chemistry: The need for metrics. *Curr. Opin. Green. Sustainable Chem.* **2016**, *1*, 1–4.
- (6) Helma, C. In silico predictive toxicology: The state-of-the-art and strategies to predict human health effects. *Curr. Opin. Drug. Discover. Devel.* **2005**, *8*, 27–31.
- (7) Rusyn, I.; Daston, G. P. Computational Toxicology: Realizing the Promise of the Toxicity Testing in the 21st Century. *Environ. Health Perspect.* **2010**, *118*, 1047–1050.
- (8) Raies, A. B.; Bajic, V. B. In silico toxicology: computational methods for the prediction of chemical toxicity. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2016**, *6*, 147–172.
- (9) Voutchkova, A. M.; Kostal, J.; Steinfeld, J. B.; Emerson, J. W.; Brooks, B. W.; Anastas, P.; Zimmerman, J. B. Towards rational molecular design: derivation of property guidelines for reduced acute aquatic toxicity. *Green Chem.* **2011**, *13*, 2373–2379.
- (10) Kostal, J.; Voutchkova-Kostal, A.; Weeks, B.; Zimmerman, J. B.; Anastas, P. T. A free energy approach to the prediction of olefin and epoxide mutagenicity and carcinogenicity. *Chem. Res. Toxicol.* **2012**, *25*, 2780–2787.
- (11) Shen, L. Q.; Judson, R.; Melnikov, F.; Roethle, J.; Gudibanda, A.; Zimmerman, J. B.; Anastas, P. T. A Probabilistic Diagram to Guide Chemical Design with Reduced Potency to Incur Cytotoxicity. *Green Chem.* **2016**, *18*, 4461–4467.
- (12) Shen, L. Q.; Melnikov, F.; Roethle, J.; Gudibanda, A.; Judson, R. S.; Zimmerman, J. B.; Anastas, P. T. Coupled molecular design diagrams to guide safer chemical design with reduced likelihood of perturbing the NRF2-ARE antioxidant pathway and inducing cytotoxicity. *Green Chem.* **2016**, *18*, 6387–6394.
- (13) Kramer, J. A.; Sagartz, J. E.; Morris, D. L. The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. *Nat. Rev. Drug Discovery* **2007**, *6*, 636–649.
- (14) Rosenkranz, H.; Matthews, E.; Klopman, G. Relationships between cellular toxicity, the maximum tolerated dose, lipophilicity and electrophilicity. *ALTA* **1992**, *20*, 549–562.
- (15) Schrage, A.; Hempel, K.; Schulz, M.; Kolle, S. N.; van Ravenzwaay, B.; Landsiedel, R. Refinement and reduction of acute oral toxicity testing: A critical review of the use of cytotoxicity data. *ATLA* **2011**, *39*, 273–295.
- (16) Kleinstreuer, N. C.; Yang, J.; Berg, E. L.; Knudsen, T. B.; Richard, A. M.; Martin, M. T.; Reif, D. M.; Judson, R. S.; Polokoff, M.; Dix, D. J.; Kavlock, R. J.; Houck, K. A. Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nat. Biotechnol.* **2014**, *32*, 583–591.
- (17) Cragg, G. M.; Newman, D. J. Plants as a source of anti-cancer agents. *J. Ethnopharmacol.* **2005**, *100*, 72–79.
- (18) Hotta, K.; Ueoka, H. New cytotoxic agents: A review of the literature. *Crit. Rev. Oncol. Hematol.* **2005**, *55*, 45–65.
- (19) Ismael, G. F.; Rosa, D. D.; Mano, M. S.; Awada, A. Novel cytotoxic drugs: Old challenges, new solutions. *Cancer Treat. Rev.* **2008**, *34*, 81–91.
- (20) Committee on Toxicity Testing and Assessment of Environmental Agents, Board on Environmental Studies and Toxicology, Institute for Laboratory Animal Research, Division on Earth and Life Studies, National Research Council. *Toxicity Testing in the 21st Century: A Vision and a Strategy*; National Academies Press: Washington DC, USA, 2007.
- (21) Dix, D. J.; Houck, K. A.; Martin, M. T.; Richard, A. M.; Setzer, R. W.; Kavlock, R. J. The ToxCast Program for Prioritizing Toxicity Testing of Environmental Chemicals. *Toxicol. Sci.* **2007**, *95*, 5–12.
- (22) Collins, F. S.; Gray, G. M.; Bucher, J. R. Transforming Environmental Health Protection. *Science* **2008**, *319*, 906–907.
- (23) Judson, R. S.; Houck, K. A.; Kavlock, R. J.; Knudsen, T. B.; Martin, M. T.; Mortensen, H. M.; Reif, D. M.; Rotroff, D. M.; Shah, I.; Richard, A. M.; Dix, D. J. In Vitro Screening of Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project. *Environ. Health Perspect.* **2009**, *118*, 485–492.
- (24) Shukla, S. J.; Huang, R.; Austin, C. P.; Xia, M. The future of toxicity testing: a focus on in vitro methods using a quantitative high-throughput screening platform. *Drug Discovery Today* **2010**, *15*, 997–1007.
- (25) Kavlock, R.; Chandler, K.; Houck, K.; Hunter, S.; Judson, R.; Kleinstreuer, N.; Knudsen, T.; Martin, M.; Padilla, S.; Reif, D.; Richard, A.; Rotroff, D.; Sipes, N.; Dix, D. Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* **2012**, *25*, 1287–1302.
- (26) Attene-Ramos, M. S.; Miller, N.; Huang, R.; Michael, S.; Itkin, M.; Kavlock, R. J.; Austin, C. P.; Shinn, P.; Simeonov, A.; Tice, R. R.; Xia, M. The Tox21 robotic platform for the assessment of environmental chemicals - From vision to reality. *Drug Discovery Today* **2013**, *18*, 716–723.
- (27) Panda, S.; Mohapatra, P.; Mohanty, R. Comparative toxicity of two organophosphorus insecticides on membrane integrity of *Chlorella vulgaris*. I. Effect on membrane permeability. *Microbiol. Res.* **1999**, *153*, 363–368.
- (28) Van Winkle, L.; Johnson, Z.; Nishio, S.; Brown, C.; Plopper, C. Early events in naphthalene-induced acute Clara cell toxicity - Comparison of membrane permeability and ultrastructure. *Am. J. Respir. Cell Mol. Biol.* **1999**, *21*, 44–53.
- (29) Mathieu-Denoncourt, J.; Wallace, S. J.; de Solla, S. R.; Langlois, V. S. Influence of Lipophilicity on the Toxicity of Bisphenol A and Phthalates to Aquatic Organisms. *Bull. Environ. Contam. Toxicol.* **2016**, *97*, 4–10.
- (30) Pliška, V.; Testa, B.; van de Waterbeemd, H. *Lipophilicity in Drug Action and Toxicology* **2008**, *4*, 1–6.
- (31) Horváth, I. T.; Anastas, P. T. Innovations and green chemistry. *Chem. Rev.* **2007**, *107*, 2169–2173.
- (32) US EPA. ToxCast & Tox21 Summary Files Released Dec. 2014; <http://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data>.
- (33) DeVito, S. C.; Garrett, R. L. *Designing Safer Chemicals. Green Chemistry for Pollution Prevention*; American Chemical Society: Washington DC, USA, 1996.
- (34) Judson, R.; et al. Analysis of the Effects of Cell Stress and Cytotoxicity on In Vitro Assay Activity Across a Diverse Chemical and Assay Space. *Toxicol. Sci.* **2016**, *152*, 323–339.
- (35) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
- (36) Marvin Calculator Plugins. Calculator Plugins were used for structure property prediction and calculation. Marvin 6.3.4; ChemAxon, 2013; <http://www.chemaxon.com>.

- (37) Johnson, B. G.; Gill, P. M.; Pople, J. A.; Fox, D. J. Computing molecular electrostatic potentials with the PRISM algorithm. *Chem. Phys. Lett.* **1993**, *206*, 239–246.
- (38) Boese, A. D.; Martin, J. M. L. Development of density functionals for thermochemical kinetics. *J. Chem. Phys.* **2004**, *121*, 3405–3416.
- (39) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ó.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian Inc.: Wallingford, CT, USA, 2009.
- (40) Parr, R. G.; Szentpály, L. V.; Liu, S. Electrophilicity index. *J. Am. Chem. Soc.* **1999**, *121*, 1922–1924.
- (41) Ji, N.; Liu, T.; Xu, J.; Shen, L. Q.; Lu, B. A Finite Element Solution of Laterally Periodic Poisson-Boltzmann Model for Membrane Channel Proteins. [Submitted to *Int. J. Mol. Sci.*]
- (42) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (43) Baker, N. A. Improving implicit solvent simulations: a Poisson-centric view. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137–143.
- (44) Lu, B. Z.; Zhou, Y. C.; Holst, M. J.; Mccammon, J. A. Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications. *Commun. Comput. Phys.* **2008**, *3*, 973–1009.
- (45) Eisenberg, D.; McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* **1986**, *319*, 199–203.
- (46) Sharp, K. A.; Nicholls, A.; Fine, R.; Honig, B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* **1991**, *252*, 106–109.
- (47) Cramer, C. J.; Truhlar, D. G. An SCF Solvation Model for the Hydrophobic Effect and Absolute Free Energies of Aqueous Solvation. *Science* **1992**, *256*, 213–217.
- (48) Tanizaki, S.; Feig, M. A generalized Born formalism for heterogeneous dielectric environments: Application to the implicit modeling of biological membranes. *J. Chem. Phys.* **2005**, *122*, 124706.
- (49) John, G. H.; Langley, P. Estimating Continuous Distributions in Bayesian Classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, San Mateo, 1995; pp 338–345.
- (50) Bradley, A. P. The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recogn.* **1997**, *30*, 1145–1159.
- (51) Lasko, T. A.; Bhagwat, J. G.; Zou, K. H.; Ohno-Machado, L. The use of receiver operating characteristic curves in biomedical informatics. *J. Biomed. Inf.* **2005**, *38*, 404–415.
- (52) Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20.
- (53) Millman, K. J.; Aivazis, M. Python for Scientists and Engineers. *Comput. Sci. Eng.* **2011**, *13*, 9–12.
- (54) Jones, E.; Oliphant, T.; Peterson, P. SciPy: Open source scientific tools for Python, 2001; <http://www.scipy.org/>.
- (55) van der Walt, S.; Colbert, S.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30.
- (56) pandas: Python Data Analysis Library, 2012; <http://pandas.pydata.org/>.
- (57) Pérez, F.; Granger, B. E. IPython: a System for Interactive Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 21–29.
- (58) Pedregosa, F.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (59) Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.
- (60) Waskom, M. et al. seaborn, v0.6.0 (June 2015), 2015; <http://dx.doi.org/10.5281/zenodo.19108>.
- (61) Roschier, L. PyNomo - a program to create nomographs with Python, 2009; <http://sourceforge.net/projects/pynomo/>.
- (62) Hanley, J. A.; McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36.
- (63) Fawcett, T. An introduction to ROC analysis. *Pattern. Recogn. Lett.* **2006**, *27*, 861–874.
- (64) Testa, B.; Crivori, P.; Reist, M.; Carrupt, P.-A. The influence of lipophilicity on the pharmacokinetic behavior of drugs: Concepts and examples. *Perspect. Drug Discovery Des.* **2000**, *19*, 179–211.
- (65) Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of Log P Methods on More Than 96,000 Compounds. *J. Pharm. Sci.* **2009**, *98*, 861–893.
- (66) Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, U. K., 2013.
- (67) Quinn, D. M.; Nair, H. K.; Seravalli, J.; Lee, K.; T, A.; Radić, Z.; C, V. D.; N, P.; Taylor, P. In *London Dispersion Interactions in Molecular Recognition by Acetylcholinesterase*; Quinn, D. M., Balasubramanian, A. S., Doctor, B. P., Taylor, P., Eds.; Springer: Weinheim, Germany, 1995.
- (68) Parke, D. V. Mechanisms of chemical toxicity – a unifying hypothesis. *Regul. Toxicol. Pharmacol.* **1982**, *2*, 267–286.
- (69) Hinson, J. A.; Roberts, D. W. Role of Covalent and Noncovalent Interactions in Cell Toxicity: Effects on Proteins. *Annu. Rev. Pharmacol. Toxicol.* **1992**, *32*, 471–510.
- (70) LoPachin, R. M.; DeCaprio, A. P. Protein adduct formation as a molecular mechanism in neurotoxicity. *Toxicol. Sci.* **2005**, *86*, 214–225.
- (71) LoPachin, R. M.; Gavin, T. Molecular mechanisms of aldehyde toxicity: A chemical perspective. *Chem. Res. Toxicol.* **2014**, *27*, 1081–1091.
- (72) Parr, R. G.; Pearson, R. G. Absolute hardness: companion parameter to absolute electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516.