

# Toxicology testing steps towards computers

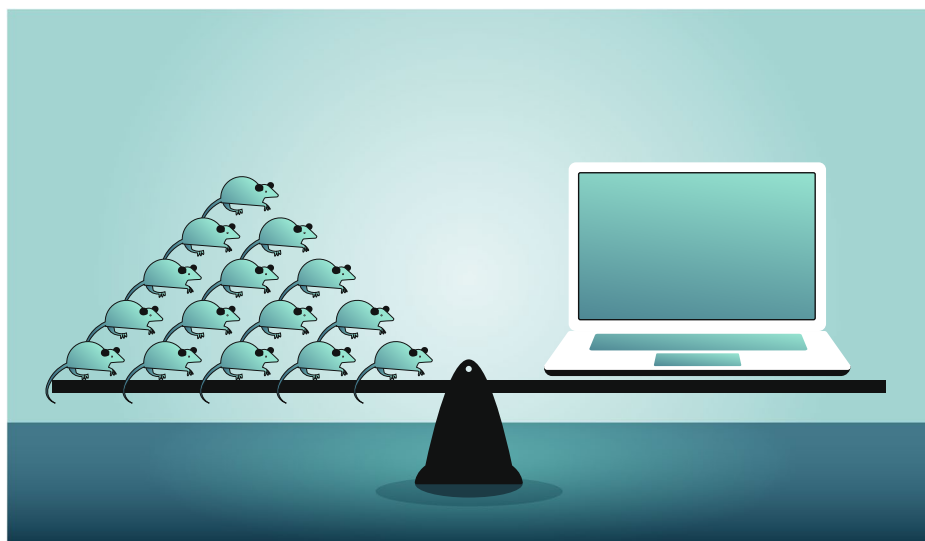
Can the computer eliminate the lab animal? As computational methods become more advanced and data more freely available, *in silico* modeling approaches have growing potential to help reduce the number of animals needed to test chemical toxicity.

Jim Kling

The 2016 overhaul of the United States Toxic Substances Control Act (TSCA), originally passed in 1976, was meant to help curb animal use in determining the potential toxicity of drugs and other chemicals. But in the short term, at least, the opposite seems to have happened. *Science* reported<sup>1</sup> a surge in animal testing, from 7,000 animals used in a few dozen tests in 2016, to more than 300 conducted a year later that involved about 75,000 rats, rabbits and other animals.

The specific cause of the jump in animal testing is unknown, but it is ironic given that the law also required the Environmental Protection Agency (EPA) to “reduce, refine, or replace” animals in toxicological testing. The trend is alarming to animal welfare and industry groups, and frustrating to researchers working on alternatives. One such alternative avenue that has made strides in recent years is to move *in vivo* toxicology studies *in silico*: a number of computational methods have been developed that could be used to guide more targeted animal studies, potentially reducing the number of animals needed, or even replace them in some regulatory cases. “For certain endpoints with a sufficient amount of data that is collected and properly curated and analyzed, *in silico* methods come close to experimental accuracy of measurement, and could replace animals. In some less studied (systems), *in silico* methods can point to gaps in existing data and prompt new data generation to fill those gaps,” said Alexander Tropsha, who is associate dean for pharmacoinformatics and data science at the University of North Carolina Eshelman School of Pharmacy, where he uses informatics technologies to assist in the development of new drugs.

*In silico* approaches capitalize on recent advances in computing power and publicly available toxicological databases to estimate the probability that a given chemical will pose a danger to humans or the environment. Such models rely on decades of *in vivo* toxicology data; researchers can use them to search for close structural relatives of a chemical of interest and find positive or negative toxicological tests in



***In vivo* vs. *in silico*:** Computer models are in the works that might help shift the balance away from animal use in toxicity testing. Credit: E. Dewalt/Springer Nature

those cases that could help predict the new chemical's toxicity.

But regulatory agencies have been slow to adopt these approaches, perhaps because the more accurate models use opaque ‘black box’ statistical analyses. Other methods rely on experts, either human or computer, to provide an alert based on specific structural elements that suggest a chemical has a high likelihood of being toxic.

Researchers developing novel *in silico* methods are looking to both improve the accuracy of their predictions, and to convince regulators to trust them.

## Opening the black box

Regulators naturally like transparency, but it doesn't equate to accuracy, according to Tropsha. That's because transparent methods, often referred to as read-across or alert-based, rely on linking the potential toxic effects of chemicals to critical contributions from individual functional groups on molecules—small parts of the whole that grant it certain chemical properties. On the surface, this seems reassuring, because a regulator

can look at two molecules and see obvious chemical similarities. But the contributions of a functional group can depend on other parts of the molecule in unpredictable ways. Functional groups are a bit like a piece of sticky tape that could wrap paper over a package, or fold over itself into any number of shapes. “It's quite a stretch to extrapolate from a single functional group to an entire chemical's behavior, given that functional groups interact with other functional groups within the chemical and influence each other's behavior and physical and, collectively, physical, electronic and biological properties,” said Tropsha.

Another method, known as Quantitative Structure-Activity Relationships (QSARs) can account for some of that behavior by analyzing and cross-referencing large data sets. QSARs analyze large data sets of molecules with known toxicological data, and then infer the likely properties of the novel chemical. The approach is more accurate because it inherently accounts for the mix of functional groups and other characteristics of the molecules, according to Tropsha. But it's also more opaque, since it

relies on computer code to reach conclusions rather than simple chemical principles.

In one example, Tropsha and colleagues modeled the aquatic toxicity of nitroaromatic compounds against a model protozoan. They used an alert-based method to classify molecules according to one of two known mechanisms of toxicity, or into a third category if a compound lacked a mechanistic-based alert. The researchers then developed two separate QSAR models, each based on one of the two known mechanisms, and combined them together to form a universal, mechanism-free model. This model was useful for a wide variety of molecules, but it tended to be less accurate.

The system worked best when researchers categorized any individual molecule according to its potential mechanism, and then ran it against the corresponding QSAR model. They ran compounds with no known mechanism against the universal model. The approach led to greater predictive accuracy and larger coverage than traditional models<sup>2</sup>.

“I think this hybrid approach will sit well with regulators because it connects with what they are used to, but has this additional appeal of statistical robustness,” said Tropsha.

### A map of the (chemical) world

Tropsha and others working to improve existing models are getting an assist from new data sources. A lack of toxicology data, historically kept behind corporate firewalls, has long been a problem for in silico methods. But in recent years, EPA, the National Institutes of Health, the National Center for Advancing Translational Sciences, and even the pharmaceutical industry have started creating extensive public databases. “There’s a trend for making more and more such data available,” said Thomas Hartung, chair for evidence-based toxicology at Johns Hopkins University.

One particularly rich source of information developed after Europe enacted the REACH (Registration, Evaluation, Authorisation, and Restriction of Chemicals) regulations in 2006, which required publication of toxicological tests for all candidate chemicals (<https://www.echemportal.org/echemportal/propertysearch/index.action>). Hartung and his group tapped this source to create a massive chemical topography that can be used to map a new chemical and estimate its toxicological properties.

Using 10 million representative molecules (out of about 140 million known), they created a map of the known chemical universe that placed structurally similar molecules close to one another.

To determine structural proximity, their model/algorithm had to compare each molecule to every other molecule in the database. That added up to 50 trillion comparisons of individual chemical pairs, which demanded about two days of calculations on an Amazon cloud service.

The researchers then pooled chemicals by 74 labels of available toxicological data, under 19 categories like acute toxicity, reproductive toxicity, or skin irritation. Any modern computer can run a comparison of a novel chemical against one of those data sets to see where it fits in the chemical space, and then examine the data of the chemical’s closest structural relatives. “If it’s negative all around, we can say it’s extremely unlikely that something will suddenly be positive,” said Hartung. An analysis where 190,000 chemicals with known classification as toxic or not were compared to the respective prediction showed that the model was 87% accurate. By contrast, when an animal study is repeated using the identical molecule, the same result occurs only 81% of the time<sup>3</sup>. “Our computational approaches outperform the reproducibility of the animal tests,” said Hartung.

Hartung’s work is looking for regulatory acceptance, which he hopes will be forthcoming in the United States as his group works to get the method validated through The Interagency Coordinating Committee on the Validation of Alternative Methods, a group established in 2000 that includes 16 US regulatory and research agencies.

### A stronger whole

Individual models like these can be effective, but other researchers are looking at ways to crowdsource in silico models to improve accuracy. Nicole Kleinstreuer, deputy director of the National Toxicology Program’s Interagency Center for the Evaluation of Alternative Toxicological Methods, coordinates such a program across multiple US government agencies and stakeholders. Their initial focus is on acute toxicity.

Kleinstreuer’s team first collects and curates training sets and then provides them to academic, industry, and government research groups. Each group then builds a prediction model using the training set, less a portion of the data that Kleinstreuer’s group reserves to evaluate the performance of the models. High-performing models, further selected based on availability of the code, whether the model can provide mechanistic insight, and other factors, are then combined to create a consensus model that outdoes the individual components. The program has had success,

she says, and has found some regulatory acceptance. “The EPA uses these consensus structure-based models to screen the entire chemical space and decides which compounds are likely to be endocrine disruptors and therefore should go into the first tier of testing, which is in vitro-based,” she said.

Crowdsourcing can increase the confidence of modeling results, since independent teams use a range of structural features, chemical properties, and descriptors. “Each of those unique strategies has strengths and weaknesses. Certain models based on certain features, using certain algorithms, are going to perform better on different parts of the chemical universe. So if you combine a bunch of different models, then you end up with a consensus modeling approach, where the whole is stronger than the parts,” said Kleinstreuer.

These models aren’t intended to fully replace animal studies, but they represent a first step. “Before you even run any in vitro assays, you’re screening them using in silico approaches so you can optimize your resources and test the chemicals that are more likely to be an issue,” said Kleinstreuer.

To improve their training sets, Kleinstreuer and her colleagues have branched beyond animal data and included in vitro assays, such as cell-based assays, receptor-binding assays, and DNA-binding assays. In collaboration with the EPA, the team employed that approach with androgen and estrogen-receptor models, using data from 18 assays applied to more than 2,000 chemicals. “We could provide a really strong weight of evidence that a chemical, based on its pattern of activity across all these different assays, was actually affecting the estrogen disruption pathways or not,” said Kleinstreuer.

Such an approach can’t be taken lightly, however. “It requires that you have really robust coverage of different parts of that biological pathway in your in vitro assays,” said Kleinstreuer.

### A step further

Some approaches are taking that philosophy one step further. EPA’s Virtual Embryo project aims to use computational models to simulate how chemicals might affect development, and what exposure thresholds pose a threat. The effort will eventually include a range of developmental processes, but the group recently published results from a cleft palate model<sup>4</sup>, which represents a developmentally complex but physiologically simple event—the fusion of the two skull plates to form the roof of the mouth.

The research is applying systems biology to understand how chemical perturbations of individual cells can lead to abnormal development. It draws on data from EPA's ToxCast database (<https://actor.epa.gov/dashboard/>), initiated in 2009, which includes data for more than 1,800 chemicals that span a broad range of uses, including industrial processes, consumer products, and environmental contaminants. ToxCast screens chemicals in over 700 high-throughput assay endpoints that cover a range of high-level cellular responses, including biochemical targets, signaling pathways, and cellular and developmental phenotypes. The researchers used that data to simulate the pathogenesis of cleft palate in a cell-by-cell, bottom-up, and interaction-by-interaction computational model of the biological processes that govern skull plate development and closure.

The work takes modeling a step further than typical in silico models. "(Those) are not tuned to look for things such as chemical-chemical interactions or gene-environment interactions, which are readily done with virtual embryo models," said team leader Thomas Knudsen, Developmental Systems Biologist at the EPA's National Center for Computational Toxicology.

It also provides information about exposure. The models can estimate a concentration of a chemical that would be expected to induce a developmental defect.

"We can take an in vitro concentration response for, say, a particular gene function that might be important for palatal closure, and we can divide that data into 32 concentrations and run the simulation on them, and predict what concentration might set the system off into an abnormal developmental trajectory," said Knudsen. Then they can model the exposure to the mother that would be required to reach that concentration in the embryo.

The researchers have previously published models for angiogenesis and urethral fusion. Together with other models in development, Knudsen hopes to create a complete virtual embryo that can predict a wide range of human toxicity. For the moment, the team must rely on predicting animal toxicity, where the latest ToxCast models range from 70–85% accuracy. "So it's pretty good, but it's not perfect," said Knudsen.

Perfection will no doubt have to wait, but Knudsen and others hope that their models will help inform toxicological testing, improving accuracy and reducing animal use when appropriate.

That's especially important given the update to TSCA, which has increased the regulatory burden on new chemicals. It is leading to longer wait times at EPA, and a 25% withdrawal rate of new chemical

applications, compared to about 5% before the overhaul, according to a recent Bloomberg report<sup>5</sup>.

That environment will continue to put pressure on industry and regulators alike to streamline toxicology testing. In silico methods are a start, though they won't yet eliminate in vivo testing. "It's only reproducing the (existing) animal data," said Hartung.

Still, things are looking up. In some cases, where there is sufficient data that is properly analyzed, in silico methods can likely reduce and replace animal testing. And even when the data is sparse, it can at least help guide the way. □

**Jim Kling**

Freelance writer, Bellingham, Washington, USA.  
e-mail: [jkling@gmail.com](mailto:jkling@gmail.com)

Published online: 14 January 2019  
<https://doi.org/10.1038/s41684-018-0227-0>

#### References

1. Zainzinger, V. *Animal tests surge under new U.S. chemical safety law*. *Science* May 8, 2018. <https://www.sciencemag.org/news/2018/05/animal-tests-surge-under-new-us-chemical-safety-law>
2. Alves, V. et al. *Green Chem.* **18**, 4348–4360 (2016).
3. Luechtefeld, T., Marsh, D., Rowlands, C. & Hartung, T. *Toxicol Sci* **165**, 198–212 (2018).
4. Hutson, M. S., Leung, M. C. K., Baker, N. C., Spencer, R. M. & Knudsen, T. B. *Chem. Res. Toxicol* **30**, 965–979 (2017).
5. Rizzuto, P. *Rate of EPA chemical regulation ramps up since toxics law update*. Bloomberg BNA May 17, 2017. <https://www.bna.com/rate-epa-chemical-n73014451140/>