Published in final edited form as:

Nat Mater. 2019 May ; 18(5): 418–422. doi:10.1038/s41563-019-0332-5.

Opportunities and Challenges Using Artificial Intelligence (AI) in ADME/Tox

Barun Bhatarai¹, W. Patrick Walters², Cornelis E.C.A. Hop³, Guido Lanza⁴, and Sean Ekins^{5,*}

¹Novartis Institutes for Biomedical Research, Cambridge, Massachusetts, USA.

²Relay Therapeutics, Cambridge, Massachusetts, USA.

³Genentech, 1 DNA Way, South San Francisco, CA 94080.

⁴Numerate, San Francisco, CA.

⁵Collaborations Pharmaceuticals Inc., Raleigh, North Carolina, USA.

Abstract

A recent conference organized a panel of scientists representing small and big pharma companies, who work at the interface of machine learning (ML) and absorption, distribution, metabolism, excretion, and toxicology (ADME/Tox). With the recent rebirth of AI related to pharma, it is timely to present this collaborative commentary to capture the diverging opinions on the past, present and future role of AI for ADME/Tox and how it can be applied in newer areas such as nanomaterials.

A brief introduction to the history of AI in ADME/Tox

The term Artificial Intelligence (AI) can have several meanings. In this context, we will be referring specifically to applications of machine learning (ML) using various methods (Table 1) in drug discovery, and more specifically to the prediction of absorption, distribution, metabolism, excretion, and toxicology (ADME/Tox) properties (Table 2) .

We should differentiate this approach from other computational approaches that are widely used such as physiologically based Pharmacokinetic (PBPK) and pharmacokinetic pharmacodynamic / quantitative systems pharmacology (PKPD/QSP) modeling; however, if the decision-making process is automated these could be applied in AI frameworks. *In silico* ADME/Tox models have made considerable progress over ~40 years (Table 3) and we are now squarely in the era of ML with many of the models serving their intended purpose in the industry and complementing experimental methods. The big question is whether the latest generation of computational tools such as deep neural networks (DNNs, also referred to as deep learning) will further improve the quality of the models and decision making in lead optimization.

*Corresponding Author: Sean Ekins, Collaborations Pharmaceuticals, Inc., 840 Main Campus Drive, Lab 3510, Raleigh, NC 27606, USA. sean@collaborationspharma.com, Tel: 215-687-1320.

A typical large pharma drug discovery team is generating data for between 20 and 50 different assays including biochemical, cellular, pharmacokinetic, ADME/Tox, selectivity, and more (Table 2). One of us has participated in an antibacterial project with 92 assays! Computational models, when appropriately implemented, can provide a means of interpreting large volumes of data and influencing the decision making and, hence, the direction of the team.

Ultimately, we need to be able to model the outcomes of compounds in a human disease state. However, we do not have enough *in vivo* data to build predictive models, so we end up building models based on *in vitro* surrogates. In some cases, the quality of the ML models for these ADME/Tox or physicochemical properties can approach the *in vitro* assays¹. In addition to being able to predict single properties, we also need multi-parametric optimization tools to combine the many different properties and activities². Some of us are advocating early human dose prediction as a way to prioritize compounds³. Over the last few decades there have been many efforts to investigate the *in vitro* – *in vivo* correlations (IVIVC). These correlations build confidence in extrapolation from *in vitro* to *in vivo*⁴ of relevant species and, by extension, to computational models derived from *in vitro* data.

What are we trying to do?

Advances in modeling approaches and 2D molecular descriptors have definitely shaped the prediction paradigm of ADME/Tox (Table 3), but not all endpoints are predicted equally well. For example, difficult to model endpoints such as time dependent CYP inhibition (TDI), where 3D as well as structural fragments play a big role⁵, lag behind other datasets. Mitochondrial toxicity⁶ is another example where the inherent toxicity might be due to several mechanisms of cytotoxicity. Public data could be of some help if it can be made homogeneous and compatible with in house results for such properties.

The role of AI in pre-clinical discovery and ADME/Tox is to help teams in the pharmaceutical industry make the right decisions. If we can do this, we should be able to get to better compounds faster. Models can be used to guide design prior to synthesis or narrow down the number of compounds. As a result, drug discovery teams save time and resources by selecting only a few compounds to test in *in vitro* or *in vivo* assays. These approaches are also translatable to aid in the development of nanomaterials⁷ and other materials to be used in delivering treatments. While machine learning has been applied to predict properties of nanoparticles, to our knowledge to date ADME/Models have not been applied for nanomaterials. This represents a potential opportunity that can learn from the progress in this area for small molecules.

We have to see the quality of the model in light of the purpose it serves. In the early stages of drug discovery, a ‘perfect’ *in silico* – *in vitro* correlation is not required. These models should provide a level of enrichment in order to enhance the optimization process. In the later stages of drug discovery, the bar goes up considerably and models must differentiate molecules with very small structural changes. In the later stages of optimization, a local model may outperform a global model. In other words, the ADME/Tox ML model used at each stage should be ‘fit for purpose’.

There has been some success in being able to predict surrogate toxicology endpoints (e.g. hERG⁸), but we are still far from being able to replace animal studies. *In vivo* computational models for ADME/Tox are not at the same level of proficiency as *in vitro* models. Models of *in vivo* data are confounded by the limited amount of data, and the fact that the data are not obtained in a consistent fashion. *In vivo* data comes with greater variability even for a single species or one animal strain because of dosage regimen, formulation, route of administration, cassette or one-on-one study, etc. In addition, the time points used for data collection in *in vivo* studies may not be optimal for model building. *In vitro* data are low hanging fruit and simpler, as it is 'one assay - one hypothesis' modeling with less variability. There are a few confounding factors in predicting *in vivo* outcomes. The first is the sheer complexity of what we are trying to predict computationally. Sometimes, the observed result is a function of a metabolite/s rather than the input compound that we are using. The second factor is the paucity of available data to model. While large pharma often have large PK datasets, the data is typically collected using multiple species, doses and formulations, all of which make it difficult to derive a consistent dataset. The statistical challenges of combining heterogeneous data, which have variable noise and bias, are also real, as are those of dealing with relatively small amounts of data (as in toxicology) to model highly complex phenomena. We also spend considerable effort trying to generate models from data in databases that are not in a format for model building.

There is perhaps also a historical overestimation of the value of ADME/Tox data in big pharma to consider as few companies have shared their data while there is call for more openness⁹ and pre-competitive consortia, such as ATOM (<https://atomscience.org/>). It would be a step forward if more companies participate.

Do we have the right tools?

The requirements for a computational model are clearly different for each person and company and it isn't a 'one size fits all'. For a model to be most useful in a medicinal chemistry setting (the 'Computer Aided Drug Design' view), it should be interpretable and provide insights into specific chemical features contributing to the measured endpoint. In addition to driving design, an interpretable model inspires buy-in and trust from the chemists using the model. Others would say (the 'AI approach') that the model does not need to be explanatory to be useful at all. The compounds that are selected with the model may not be obvious to a human and just need to work or be active. For those involved in ML model building, this is holding them to a very high standard. This also highlights a very important philosophical aspect. While there are scientific hurdles to overcome, the cultural barriers that impede uptake of ML are just as high if not higher because of inherent lack of trust.

Is data ready for tools to deploy?

One of the challenges is the need to create meta-data to connect the many siloed databases within companies. An end-to-end data connection and data flow will then enable the application of ML techniques. Data from early discovery is usually captured and annotated well (but not always). AstraZeneca has proposed an approach using control compounds to calibrate the output from different company sites¹⁰. This is also illustrated by the hepatocyte

model conditions at Genentech and Roche (which are effectively the same company) as they are different and researchers just cannot simply pool their data for modeling. Many organizations use PowerPoint and PDF documents to capture results so data needs to be extracted and text mined. Another key data curation issue is extracting molecule structures that are encoded in publications, patents or reports. In several cases, the molecules are represented using various combinatorial descriptions using methods such as R-groups, which must subsequently be reassembled into molecules. Ultimately, we need all important data to be captured in a machine readable form¹¹. Big Pharma therefore is primarily a technology-enabled industry and needs appropriate infrastructure to integrate ML and AI on top of the existing system or data. Nevertheless, implementation has been slow in part due to lack of employees skilled in these computational approaches.

Are the models any good?

A wide variety of ML techniques have been applied in virtual screening¹². Efforts to explore more diverse training data, appropriate molecular features, and consensus modeling^{13, 14} are ongoing in many groups. The availability of several online ML based ADME model repositories^{15, 16} and prediction servers provides some utility to the community, but underlying models are limited by the data used. The availability of Open Source code or open source models are likely to enhance uptake. NIH/NCATS is putting ML models in the public domain (<https://predictor.ncats.io/>). These and other public repositories and efforts to share models are steps towards more broadly accessible ADME/Tox models¹⁷.

As scientists, we often ask how we will know when our ML models will be useful in a drug discovery program? This often comes down to asking “does the molecule I’m predicting look like anything in my training set”^{18, 19}. History has shown that ML models are better at interpolating than they are at extrapolating. This also brings us back to a model being ‘fit for purpose’, and overcoming unrealistic expectations regarding the few false negatives observed. We should focus on enrichment and identifying more promising leads quickly. Part of the challenge is how we measure success for ML. We need to frame the question correctly. Far too often, people ask, “Where are the drugs that came from ML?” This is like asking, “Where are the drugs that came from high throughput screening, or combinatorial chemistry?” One way to theoretically test this would be to put two drug discovery teams in the same company on the same project, one using ML for all aspects and the other not. It is unlikely that any company would undertake such an expensive experiment. In the end, a model is therefore only useful if a team uses it to drive decision making. Most pharma companies are employing ADME/Tox models as part of their drug discovery process and regularly track model performance and applicability. Still, one needs to ensure the models are updated as a project advances in order to avoid models becoming outdated as the chemistry changes.

Which Algorithm is best?

As DNNs have become more accessible another common question we are asked is “will Deep Learning be the solution to our ADME/Tox problems?” The answer to this may depend on the size and quality of the dataset in question. Given the same molecular

representation (e.g. chemical fingerprints) the differences between ML algorithms tend to be very small when compared to DNN. A group from Merck and the University of Toronto compared the performance of DNN with another ML algorithm using 15 large datasets²⁰. Overall the DNN method produced a modest improvement of 0.04 in R^2 ²⁰. Another promising area of research for DNNs is the use of multitask methods which simultaneously train models to predict multiple properties (e.g. aqueous solubility, and hERG inhibition). The idea is that the underlying knowledge of molecular interactions can be transferred between tasks²¹. Recent research has shown that multitask models may be more efficient than making separate models when the endpoints are correlated. Performance is also increased when similar compounds are used to predict multiple tasks^{22,23, 24}. So far, traditional ML models and DNNs are more or less on par^{17, 25, 26}, with DNN occasionally outperforming the other methods marginally²¹, but DNNs may allow predictions based on SMILES strings rather than many chemical descriptors as required by traditional ML methods. In some cases, with very large datasets for physicochemical properties like logD the DNN models perform slightly better than other ML algorithms. Literature reports have shown that, to date, multitask NNs have not quite lived up to expectations^{27, 28} though the results are not worse either.

One potential advantage to DNNs is the ability to create new representations (e.g. graph convolution) that can be more expressive than the chemical fingerprinting representations used for the last 30 years. One disadvantage of these representations is that they are not easily interpretable²⁹. DNNs have the promise of learning from very complex data, making them especially attractive in modeling *in vivo* ADME/Tox data. To date, this promise appears largely unfulfilled. Deep learning can also generate spurious correlations to noise. Many, if not most, of the *in vivo* end points we deal with are noisy and can be ambiguous. Whether we can model something like *in vivo* half-life as the endpoint with only a chemical structure as input remains to be seen. However, we are aware of early efforts to use all rat *in vitro* ADME data as input with a DNN to predict *in vivo* PK. DNN might be best utilized with pathology studies relating to toxicity e.g. to assess hepatotoxicity, nephrotoxicity etc. when given a set of cellular images and the algorithms can train themselves to identify damaged cells or subtle differences that are drug induced. In summary, for small to medium ADME/tox datasets, deep learning may not offer a big enough advantage to left the effort unless they are complex datasets or use image data.

The future

It is abundantly clear from the discussion above that we need to have closer collaboration between those generating ADME/Tox data for small molecule drug discovery and those building models so that the data being generated is appropriate for model building. Also, we should consider the end users to facilitate uptake of the models and their continued/enhanced utilization. We need to find scientists who understand the domain, have the appropriate computational skills and a sufficient understanding of statistics so that they know when models work and when they don't. One of the clearest observations we made during our discussion was the understaffing of large pharma in this area relative to the smaller companies. It seems like most of the ML effort in pharma has been dedicated toward either low-lying fruit like cellular images or high-value return like clinical trial modeling. Efforts

toward ADME/tox models still seem to lie with the those of us who have been building QSAR models for many years.

Hype versus reality

There are several early stage companies that are using AI for drug discovery or aspects of drug development (Table 4). It will be interesting to see whether they ultimately are more successful, faster, cheaper, etc. compared with traditional drug discovery. There is already a growing list of applications and alliances between pharmaceutical companies and some of these AI companies, some of which may be addressing ADME/Tox.

We have mixed opinions on whether what we are seeing currently with deep learning and AI is just the next hot fad or a glimpse of greater things to come. So far, the gains of DNN alone may not be great, but there may be more impact if they are used in consensus with other ML methods. We also need to consider how to obtain sufficient data of the appropriate quality to enable DNN models. There is not a single solution in drug discovery and one has to remain receptive to new technology.

In summary, the pharmaceutical industry is already embracing AI as evidenced by their investments and collaborations, as well as conferences, reviews and commentaries on the topic³⁰. Our consensus is the need to apply the right ML technologies in the right way with the right dataset at the right time and ultimately ask the right ADME/Tox questions!

Acknowledgments

Debra Chipman and Elizabeth Cutler are kindly acknowledged for organizing the AI Applications in Biopharma Summit. SE acknowledges Dr. Alex Clark, Dr. Joel Freundlich and Dr. Antony Williams for their many discussions on machine learning and ADME/Tox models. SE acknowledges funding to Collaborations Pharmaceuticals Inc. from NIGMS R44 GM122196-02A1.

References

1. Gupta RR et al. Drug Metab Dispos 38, 2083–2090 (2010). [PubMed: 20693417]
2. Ekins S, Honeycutt JD& Metz JT Drug Discov Today 15, 451–460 (2010). [PubMed: 20438859]
3. Page KM Mol Pharm 13, 609–620 (2016). [PubMed: 26696327]
4. Webborn PJ Future Med Chem 6, 1233–1235 (2014). [PubMed: 25162995]
5. Zientek M et al. Chem Res Toxicol 23, 664–676 (2010). [PubMed: 20151638]
6. Zhang H et al. Toxicol In Vitro 23, 134–140 (2009). [PubMed: 18940245]
7. Butler KT et al. Nature 559, 547–555 (2018). [PubMed: 30046072]
8. Wang S et al. Mol Pharm 13, 2855–2866 (2016). [PubMed: 27379394]
9. Ekins S& Williams AJ Lab on a Chip 10, 13–22 (2010). [PubMed: 20024044]
10. Winiwarter S et al. J Comput Aided Mol Des 29, 795–807 (2015). [PubMed: 25697964]
11. Clark AM, Williams AJ& Ekins S J Cheminform 7, 9 (2015). [PubMed: 25798198]
12. Martin EJ et al. J Chem Inf Model 57, 2077–2088 (2017). [PubMed: 28651433]
13. Ericksen SS et al. J Chem Inf Model 57, 1579–1590 (2017). [PubMed: 28654262]
14. Verras A et al. J Chem Inf Model 57, 445–453 (2017). [PubMed: 28257198]
15. Capuzzi SJ et al. J Chem Inf Model 57, 105–108 (2017). [PubMed: 28045544]
16. Sushko I et al. J Comput Aided Mol Des 25, 533–554 (2011). [PubMed: 21660515]
17. Russo DP et al. Mol Pharmaceutics, (2018).

18. Sheridan RP J Chem Inf Model 53, 2837–2850 (2013). [PubMed: 24152204]
19. Roy K, Kar S& Ambure P Chemometrics and Intelligent Laboratory Systems 145, 22–29 (2015).
20. Ma J et al. J Chem Inf Model 55, 263–274 (2015). [PubMed: 25635324]
21. Liu, K; 2018. [cited]Available from: arXiv:1803.06236v2
22. Ramsundar B et al. J Chem Inf Model 57, 2068–2076 (2017). [PubMed: 28692267]
23. Hop P, Allgood B& Yu J Mol Pharm 15, 4371–4377 (2018). [PubMed: 29863875]
24. Rodriguez-Perez R& Bajorath J ACS Omega 3, 12033–12040 (2018). [PubMed: 30320286]
25. Korotcov A et al. Mol Pharm 14, 4462–4475 (2018).
26. Lane T et al. Mol Pharm, (2018).
27. Xu Y et al. J Chem Inf Model 57, 2490–2504 (2017). [PubMed: 28872869]
28. Ramsundar B et al. 2015 [cited]Available from: <https://arxiv.org/abs/1502.02072>
29. Kearnes S et al. J Comput Aided Mol Des 30, 595–608 (2016). [PubMed: 27558503]
30. Ekins S Pharm Res 33, 2594–2603 (2016). [PubMed: 27599991]

Table 1.

Examples of common machine learning algorithms applied in ADME/Tox

Algorithm	Summary
Random Forest (RF)	An ensemble learning method that constructs many decision trees and outputs class or mean prediction.
Support Vector Machine (SVM)	A supervised learning method. The examples are mapped in space and classes separated by a hyperplane.
Neural Network	A simple neural network has input, hidden and output layers.
<i>K</i> -nearest Neighbors (KNN)	A non-parametric method that uses the <i>K</i> closest training examples in the feature space and classifies objects by a majority vote or in regression uses the average of the values of the nearest members.
Naïve Bayes (NB)	A probabilistic classifier, considers features to contribute independently to the probability.
Deep Learning (DNN)	Uses multiple layers of a neural network, where each layer uses the output from the previous one. It can learn multiple levels of representations at different levels of abstraction.

Table 2.

Examples of ADME/Tox properties that are modeled using ML algorithms from *in vitro* or *in vivo* data.

Properties	Measurement
Solubility	Kinetic solubility
Clearance	Rodent <i>in vivo</i> - Snapshot PK
Permeability	Caco-2, MDCK, PAMPA
Transporters (e.g., P-gp)	Transporter overexpressing cell lines
Metabolic stability	Liver microsomes and hepatocytes
Drug Drug Interactions	CYP450s, transporters
Blood Brain Barrier (BBB)	Mouse brain endothelial cell line
Cardiotoxicity (hERG)	Binding or flux in different cell types

Table 3.Progress of *in silico* ADME/Tox

Decade	Summary of <i>in silico</i> ADME/Tox developments
1980s	Small datasets (10s of compounds) Quantitative Structure Activity Relationship (QSAR), seminal work from Corwin Hansch etc. Comparative Molecular Field Analysis, Partial Least Squares.
1990s	Slightly larger datasets (100s), many more algorithms. Lipinski's Rule-of-5, pharmacophores, decision trees, recursive partitioning.
2000s	More data due to high-throughput ADME (1000s), rise of machine learning, RF, SVM etc., more pharma rules (BBB, <i>in vivo</i> etc.), multi-parametric optimization. Big pharma routinely gathered up to 1000 compounds' data per week.
2010s	10,000 to >100,000 data for <i>in vitro</i> ADME properties in big pharma, public data in the low 1000s, Growth of open projects (e.g. etox, OpenTox, Tox21, ToxCast etc) wide variety of ML algorithms (list here) and ADME/Tox models routinely used, increasing application of deep learning.

Table 4.

Companies using AI for drug discovery and beyond

Company	Summary	URL
Healx	Repurposing for rare diseases	https://healx.io/
Atomwise	Deep learning and structure-based drug design	https://www.atomwise.com/
Recursion Pharma	Deep learning and high content screening combined	https://www.recursionpharma.com/
Whole biome	Combining AI with the microbiome	https://www.wholebiome.com/
Benevolent AI	Knowledge graphs and drug repurposing	https://benevolent.ai/
Berg Health	AI and patient data	https://www.berghealth.com/
InSilico Medicine	Deep learning for drug discovery and biomarkers	http://insilico.com/
Numerate	Multiple ML techniques, including Deep Learning for drug discovery.	http://www.numerate.com/
Exscientia	AI for drug discovery	https://www.exscientia.co.uk/
PathAI	AI for pathology	https://www.pathai.com/
Flatiron Health	AI and clinical data	https://flatiron.com/
Innoplexus	ML and text analytics	https://www.innoplexus.com/
twoXAR	AI based drug discovery platform	http://www.twoxar.com/
Data2Discovery	AI and knowledge graphs for drug discovery	http://d2discovery.com
Insitro	ML models for drug discovery	http://insitro.com
Collaborations Pharmaceuticals	Using ML to develop treatments for neglected and rare diseases	http://www.collaborationspharma.com/