Self-Driving Laboratories: A Paradigm Shift in Nanomedicine Development

Authors:

Riley J. Hickmana,b,c, Pauric Bannigand, Zeqing Baoa, Alán Aspuru-Guzikab,c,e,f,g,h, Christine Allena

† R.J. Hickman and P. Bannigan contributed equally to this work.
* Lead contact and correspondence: cj.allen@utoronto.ca

Affiliations:

aDepartment of Chemistry, University of Toronto, Toronto, ON M5S 3H6 Canada
bDepartment of Computer Science, University of Toronto, Toronto, ON M5S 2E4 Canada
cVector Institute for Artificial Intelligence, Toronto, ON M5S 1M1, Canada
dLeslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON
eLebovic Fellow, Canadian Institute for Advanced Research (CIFAR), Toronto, ON, M5S 1M1, Canada
fDepartment of Chemical Engineering & Applied Chemistry, University of Toronto, Toronto, ON M5S 3E5, Canada
gDepartment of Materials Science & Engineering, University of Toronto, Toronto, ON M5S 3E4, Canada
hCIFAR Artificial Intelligence Research Chair, Vector Institute, Toronto, ON, M5S 1M1, Canada

Keywords: self-driving laboratories, materials acceleration platforms, nanomedicines, lipid-based nanoparticles, liposomes, polymer nanoparticles, drug delivery, machine learning, active learning, automation, high-throughput experimentation

Summary: Nanomedicines have transformed promising therapeutic agents into clinically approved medicines with optimal safety and efficacy profiles. This is exemplified by the mRNA vaccines against COVID-19, which were made possible by lipid nanoparticle technology. Despite the success of nanomedicines to date, their design remains far from trivial in part due to the complexity associated with their preclinical development. Herein we propose a nanomedicine materials acceleration platform (NanoMAP) to streamline the preclinical development of these formulations. NanoMAP combines high-throughput experimentation with state-of-the-art advances in artificial intelligence (including active learning and few-shot learning) as well as a web-based application for data sharing. The deployment of NanoMAP requires interdisciplinary collaboration between leading figures in drug delivery and artificial intelligence to enable this data-driven design approach. The proposed approach will not only expedite the development of next generation nanomedicines, but also encourage participation of the pharmaceutical science community in a large data curation initiative.
1. Introduction

Nanomedicines have enabled significant advancement in pharmaceutical formulation due to their ability to confer improvements in drug safety and efficacy, thereby transforming promising active agents (such as small molecule drugs, biologicals, DNA and RNA) into viable drug products.\textsuperscript{1,2} The clinical success of nanomedicines has afforded global improvement in the treatments and lives of patients suffering from a range of diseases. For example, two of the clinically approved mRNA COVID-19 vaccines were formulated in lipid nanoparticles that provide stability to the mRNA cargo, allowing delivery to their sites of action in the body.\textsuperscript{3,4} Similarly, nanomedicines have been demonstrated to increase the efficacy and reduce the side-effects (\textit{i.e.}, toxicity) of several chemotherapeutics.\textsuperscript{5} Prominent examples include DOXIL\textsuperscript{®} (liposomal doxorubicin) and VYXEOS\textsuperscript{®} (liposomal daunorubicin hydrochloride and cytarabine, encapsulated at a synergistic ratio). Indeed, numerous active agents require formulation in advanced drug delivery systems such as nanomedicines to be commercially viable. However, despite the great potential and clinical success of nanomedicines, their design and development remain challenging in comparison to the development of conventional pharmaceutical products such as tablets and capsules.\textsuperscript{6} As a result, we propose the use of a materials acceleration platform (MAP),\textsuperscript{7–12} named NanoMAP. NanoMAP will harness artificial intelligence (AI), machine learning (ML), and high-throughput experimentation to automate the design process and expedite the development of nanomedicines.

2. Rising to the Challenge

\textbf{NanoMAP addresses two of the key challenges that confront the field of nanomedicine development.} First, the application of AI and ML in the pharmaceutical sciences is currently challenged by the lack of high-quality data necessary for effective deployment of this approach.\textsuperscript{13} This may seem paradoxical given the tremendous investment into applications of nanotechnology in drug delivery in the last two decades. Indeed, there is a vast body of scientific literature on both nanotechnology-based formulation development and preclinical characterization of nanomedicines.\textsuperscript{14} A Web of Science search for either “nanoparticles”, “polymer nanoparticles”, “liposomes” or “lipid nanoparticles” and the secondary keyword “drug delivery” returns more than 70,000, 20,000, 14,000, and 10,000 research articles in the last 20 years, respectively (\textbf{Figure 1}). However, the utility of this research for data-driven nanomedicine design is limited by poor publication practices and a lack of standardization in reporting key components or properties of drug formulations.\textsuperscript{15,16} For example, when one combs through the articles related to polymer nanoparticles and drug delivery to compile the properties of materials and active agents in an effort to build a database, the extent to which important meta-data is omitted from manuscripts becomes clear (\textbf{Table 1}). This lack of complete, high-quality data in scientific articles is recognized as a key challenge for researchers who rely on data mining to train ML models, with extracted datasets rarely exceeding 1000 data instances and more often limited to <200 samples.\textsuperscript{13} In order to address this issue, NanoMAP will implement standardized high-throughput experimentation, open access database, and ready-to-use experimental protocols.
Second, there are a plethora of active agent properties, material properties and manufacturing parameters that must be considered in the development of nanomedicines. The currently employed design–build–test loop is largely reliant on an iterative, trial-and-error-based experimental approach. Owing to the combinatorial explosion of possible experiments, full resolution of the relationship between parameters and formulation performance is typically intractable. Moreover, it is well established that material-drug (or active agent) compatibility significantly influences the properties and performance of a formulation (i.e., loading capacity, stability, release kinetics, etc). Yet, most current research efforts in this space focus on use of the same small pool of materials that are clinically approved. This is largely due to the additional time and cost associated with potential regulatory hurdles for new materials and excipients. In order to justify the cost of bringing such a material to market, its inclusion would need to result in a formulation with performance that is unachievable with any combination of currently approved materials or excipients. Additionally, as each active agent has its own unique physicochemical properties, it is understood that no one material can serve as the ideal candidate for the formulation of all active agents. Thus, restriction of the formulation design space clearly deters innovation. The proposed MAP harnesses advances in AI to enable data-driven development of innovative nanomedicines within the boundaries of the established, currently available and clinically approved materials. Moreover, NanoMAP might allow for the design of new materials that could afford formulations with leap-step advances in performance.

![Graph showing number of publications on nanoparticles, polymer nanoparticles, liposomes, and lipid nanoparticles from 2002 to 2022.](image)

**Figure 1.** Summary of the number of publications on nanoparticles (yellow bars), polymer nanoparticles (red bars), liposomes (orange bars) and lipid nanoparticles (purple bars) and drug delivery in the past 20 years. Data sourced from the Web of Science using keyword combinations of “nanoparticles” and “drug delivery”, “polymer nanoparticles” and “drug delivery”, “liposomes” and “drug delivery”, or “lipid nanoparticles” and “drug delivery”, with the search results filtered by research articles only.

**Table 1.** A snapshot of the information reported on drug-loaded polymer nanoparticle formulations. Data is extracted from a Web of Science literature search for publications on “polymer nanoparticles” and “drug
delivery”. Specifically, the search results were sorted in order of the number of citations, and the top ten articles deemed relevant were inspected further and summarized below.

<table>
<thead>
<tr>
<th>Active agent</th>
<th>Polymer</th>
<th>Polymer molecular weight (kDa)</th>
<th>Particle size (nm)</th>
<th>Particle PDI</th>
<th>Zeta potential (mV)</th>
<th>Drug loading capacity (%)</th>
<th>Encapsulation efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coumarin-6</td>
<td>PLGA</td>
<td>40 to 75</td>
<td>201 ± 2</td>
<td>0.04</td>
<td>-37</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Docetaxel</td>
<td>PLGA (PEG)</td>
<td>- 15 (3)</td>
<td>~50</td>
<td>-</td>
<td>~29</td>
<td>~4</td>
<td>~28</td>
</tr>
<tr>
<td>Paclitaxel</td>
<td>PLA</td>
<td>106</td>
<td>589 ± 245</td>
<td>0.33</td>
<td>-</td>
<td>-</td>
<td>43</td>
</tr>
<tr>
<td>Paclitaxel</td>
<td>PLGA</td>
<td>22</td>
<td>112 ± 4</td>
<td>0.18 ± 0.005</td>
<td>-0.6 ± 6</td>
<td>0.7 ± 0.04</td>
<td>70 ± 4</td>
</tr>
<tr>
<td>Curcumin</td>
<td>Alginate</td>
<td>-</td>
<td>100 ± 20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tetanus toxoid</td>
<td>Chitosan</td>
<td>&gt;50</td>
<td>354 ± 27</td>
<td>-</td>
<td>+37 ± 6</td>
<td>-</td>
<td>55 ± 3</td>
</tr>
<tr>
<td>Silk peptide</td>
<td>Chitosan</td>
<td>-</td>
<td>80</td>
<td>206 ± 22</td>
<td>+25 ± 3</td>
<td>8</td>
<td>82</td>
</tr>
<tr>
<td>Gatifloxacin</td>
<td>Chitosan</td>
<td>-</td>
<td>65 to 90</td>
<td>205 ± 16</td>
<td>+27 ± 4</td>
<td>-</td>
<td>78 ± 5</td>
</tr>
<tr>
<td>Estradiol</td>
<td>PLGA</td>
<td>15</td>
<td>98 ± 3</td>
<td>0.16 ± 0.008</td>
<td>+79 ± 2</td>
<td>-</td>
<td>51 ± 6</td>
</tr>
<tr>
<td>Bovine serum albumin</td>
<td>PLGA</td>
<td>- 120</td>
<td>110</td>
<td>-</td>
<td>-</td>
<td>23</td>
<td>90</td>
</tr>
</tbody>
</table>

3. Nanomedicine Materials Acceleration Platform (NanoMAP)

a. Physical platform for high-throughput nanoparticle development

**Nanoprecipitation for high-throughput nanomedicine screening.** Amongst the various methods to prepare nanomedicines, nanoprecipitation is perhaps the most amenable to automation and high-throughput experimentation via liquid handling robots. Briefly, this method involves the addition of molecularly dissolved excipients and active agents (within the organic phase) to an aqueous phase (often containing surfactant to stabilize the resulting nanoparticles). This process has been used to prepare lipid nanoparticles, polymer nanoparticles, and liposomes. Miniaturization of the nanoprecipitation process onto a 96-well plate format using a liquid handling robot has been reported in an effort to screen the lipid composition for mRNA-loaded lipid nanoparticles. Moreover, it has also been demonstrated that nanomedicines screened in this manner can be scaled up on microfluidics platforms, thus eradicating potential downstream manufacturing concerns. It is straightforward to envision the development of similar workflows for small molecules, proteins, or peptides given the necessary equipment to characterize the resulting nanomedicines. For instance, drug loading capacity (DLC) and encapsulation efficiency (EE) can be automated with appropriate extraction methods and analysis via high-pressure liquid chromatography (HPLC). High throughput in vitro stability and release assays in biorelevant media are also amenable to automation using 96-well dialysis plates and similar sampling protocols to those mentioned above for DLC and EE.
Dynamic light scattering plate readers enable the high throughput measurement of particle size.\textsuperscript{21,22} Therefore, all of the necessary equipment and components to conduct such high-throughput experiments already exist. In the NanoMAP framework, we unite these components in order to systematically prepare and characterize nanomedicines in an automated fashion (Figure 2).

\textbf{Figure 2.} A scheme summarizing the physical platform for high-throughput nanomedicine screening. Candidate excipients and a formulation design space are derived from the existing literature (A). The design space can then be explored to identify nanomedicines with desirable performance (e.g., size, loading, and release kinetics) using automated experimentation (B). This high-throughput approach will allow for lead candidate formulations that meet key design criteria (i.e., active agent loading levels, release kinetics) to be identified in a manner that saves time and resources. Potential lead candidate formulations can then be scaled up for further preclinical studies as required (C). Created with BioRender.com.

\textbf{b. Machine learning strategies to enhance efficient formulation development}

\textbf{Active learning to efficiently generate informative datasets.} Active learning (AL)\textsuperscript{23} is a model-based sequential learning strategy in which a ML model can achieve greater predictive accuracy with fewer training data if it is allowed to choose the data on which it is trained. AL is a well-established framework for tasks in which unlabeled data is abundant (e.g., potential drug-lipid-surfactant formulations) but obtaining such labels incurs significant expense. Here, we intend to use an active learner to efficiently construct an informative dataset of nanomedicines (Figure 3A). Our active learner will sequentially recommend formulations to be prepared and measured by the automated laboratory via a utility function which prioritizes certain formulations based on their expected informativeness. Resulting formulation performance measurements will be passed to the model for re-training. This process will repeat until a predetermined
and tackle the reproducibility crisis in the pharmaceutical sciences. Indeed, AL strategies such as Bayesian optimization \textsuperscript{24–26} and variance-based sampling \textsuperscript{23} have recently been applied in the context of closed-loop design of nanomaterials, including inorganic nanoparticles,\textsuperscript{27,28} polymeric nanoparticles encapsulating nucleic acids,\textsuperscript{29} and polymer-protein hybrids.\textsuperscript{30}

**Few-shot learning for accurate prediction of nanomedicines for novel active agents.** The availability of high-quality data in drug formulation and development is to some extent limited by a lack of standardization in experimental design and poor reporting practices. Although the present study will aid in the generation of a relatively large, standardized dataset of nanomedicines and will enable the rapid testing of conventional supervised ML models, the paramount practical advantage to such a dataset would be to enable accelerated nanomedicine development for novel active agents. However, extrapolation to out-of-distribution examples is a challenging task for conventional supervised ML. As such, we argue that nanomedicine design can be cast as a **few-shot learning problem** (Figure 3B).\textsuperscript{31} In this paradigm, the entire available dataset is organized by active agents into subsets called **source tasks**. Few-shot learners are then trained on all source tasks individually. Test time refers to commencement of a novel campaign in which an optimal formulation must be discovered for a novel active agent. Sequential experimentation can now commence, but with the few-shot learner guiding experimentation with access to the novel campaign and the source task information. Such approaches have been shown to effectively transfer knowledge from source tasks onto novel campaigns, thus increasing predictive accuracy in the low-data regime and ultimately accelerating the discovery process.\textsuperscript{32–34}

c. **Web-based application for data sharing**

**Graphical user interface for dissemination of research progress.** Historically, in addition to the lack of freely available high-quality data to train models, one of the challenges with the integration of ML into the pharmaceutical sciences has been a lack of the necessary programming skills needed to implement these tools.\textsuperscript{12,13} To ensure the maximum impact and adoption of the data-driven tools described herein, the NanoMAP platform will include a user-friendly web-based application (Figure 3C). This application will allow access to the curated database, detailed instructions on how to reproduce our low-cost physical platform, as well as access to the trained few-shot learning models via an interface which does not require programming expertise. Similar interfaces have been created to democratize data-driven molecular property prediction \textsuperscript{35–38} and chemical reaction optimization.\textsuperscript{39,40} Crucially, we will also provide an option to upload datasets and meta-data, which will promote community engagement and tackle the reproducibility crisis in the pharmaceutical sciences.
4. Roadmap to NanoMAP

NanoMAP will combine high-throughput experimentation with state-of-the-art advances in AI (including active learning and few-shot learning) and a web-based application for data sharing to not only expedite the design of innovative next-generation nanomedicines, but also to promote community participation in a larger data curation project. As a roadmap to the development of NanoMAP we have identified three key milestones.

**Milestone 1:** An initial formulation design space of excipients and active agents will be established to develop the necessary automation workflows for NanoMAP. This initial design space will consist of commonly used excipients or materials and a panel of up to five active agents selected to have diverse physicochemical properties. The hardware required to conduct the proposed high-throughput experiments is commercially available. Most research labs skilled in nanomedicine development have the necessary analytical equipment and consumables to conduct this research. The key piece of equipment preventing automation is likely access to a liquid handling robot. However, these have become significantly more affordable in recent years, with entry level models available for as little as $7K.
Milestone 2: Expansion of the initial design space with new active agents for the integration of active learners and few-shot learners into NanoMAP. At this point, key design criteria (i.e., figures of merit) for the nanomedicines will be established as target objectives for the active learners and few-shot learners. Various active learning algorithms for optimization tasks have been developed and are freely available via GitHub, including several that have been developed by members of our research team.\textsuperscript{41,42} Conversely, few-shot learners are much less explored in the pharmaceutical sciences, and the full potential of such models has yet to be realized.

Milestone 3: Construction of a web-based application to host a database of formulation design (meta-)data and automated protocols, as well as detailed instructions on how to reproduce the physical NanoMAP platform. Additionally, a graphical user interface will be accessible through the website that will host programming-free access to pre-trained ML models. Various platforms exist to host such web-based applications. For example, the Aspuru-Guzik group has developed MOLAR, an open-source database management system for PostgreSQL tailored to the needs of materials scientists.\textsuperscript{43}

Feasibility of approach: The necessary hardware to conduct the proposed high-throughput experimentation is commercially available. Moreover, advances in AI and ML research over the past two decades have afforded much of the software framework for the model-based optimizers necessary to close the design–build–test loop for NanoMAP. What remains is for the components to be assembled into a single platform. This is not a trivial task. The successful implementation of NanoMAP (and other such self-driving labs in the pharmaceutical sciences) requires interdisciplinary collaboration between leading figures in drug delivery and AI. Herein, we have assembled a team with the skills, experience, and knowhow necessary to build NanoMAP. However, measurable impact on the field requires widespread adoption of such an approach with open sharing of data, methods and statistical models throughout the community. For this reason, we propose construction of a web-based application to enable and promote access and dissemination. Importantly, the platform will enable those who are not experts in formulation or computer science to deploy these data-driven tools.

5. Societal Impact of NanoMAP

While the COVID-19 pandemic has challenged the operations of scientific laboratories around the globe, it is science and medicine that are leading us out of this pandemic. In fact, advanced pharmaceutical formulation is amongst the key areas that have played a critical role in the fight against COVID-19. NanoMAP will lead to a better understanding of the composition-performance relationships of nanomedicines, to the discovery of innovative nanomedicines, and to a decrease in the financial cost and time associated with bringing new nanomedicines to market. This research will also result in the development of cutting-edge few-shot learning models that will help solve the “big data” issue that has thus far limited the deployment of ML models in the pharmaceutical sciences. These models, as well as the curated datasets, will be shared openly on-line through open access repositories such as GitHub, for use by the global research community. To further increase the usability of the data-driven tools that will result from this research, we will also develop and deploy a free to use and user-friendly
web-based application to allow programming-free access to pre-trained models. While NanoMAP will focus on the development of nanomedicines, we see no reason to believe that the foundational workflows and models cannot be improved or refined and made applicable to an extended range of sectors where formulation is critical to product development (e.g., agriculture, cosmetics, paints and coatings). Ultimately, we believe that the deployment of NanoMAP, and the development of similar self-driving labs, has the potential to make a transformative impact on the formulation development process and improve the bench-to-bedside translation of innovative medicines for patients who suffer from life-threatening diseases.

Acknowledgements

The authors are thankful to Dr. Matteo Aldeghi and Jonathan Zaslavsky for helpful discussion. R.J.H. gratefully acknowledges the Natural Sciences and Engineering Research Council of Canada (NSERC) for provision of the Postgraduate Scholarships-Doctoral Program (PGSD3-534584-2019) as well as support from the Vector Institute. C.A. acknowledges an NSERC Discovery grant (RGPIN-2022-04910). A.A.-G. acknowledges support from the Canada 150 Research Chairs program and CIFAR, as well as the generous support of Anders G. Fröseth.

Conflicts of interest

A.A.-G. is the Chief Visionary Officer and founding member of Kebotix, Inc.

References


