Artificial Intelligence for Electrolyte Design: Going Beyond the Molecular Paradigm

by Austin D. Sendek and Venkatasubramanian Viswanathan

Electrolytes are fundamental to the function of batteries. They transport ions between the electrodes, while remaining inert in the harsh electrochemical environment the electrodes create. Designing electrolytes is a challenging, multidimensional problem that is unique from typical materials discovery or molecular design efforts: for one, a candidate electrolyte must satisfy tens of properties simultaneously. Second, the design space is immense. Electrolytes are chemical mixtures consisting of organic molecules and inorganic salt; thus, designing an electrolyte formulation for Li-ion batteries involves the combinatorial challenge of selecting a small number of organic molecules and a lithium-containing inorganic salt from a seemingly endless number of options, while accurately predicting all the key properties for each candidate formulation.

Rick Nason of Dalhousie University famously distinguished between *complicated* and *complex* challenges: the former are processes where the solution is difficult to reach but a rational map to the solution exists.[1] For the latter, the solution is difficult to reach, and no map exists to get you there. Historically, electrolyte design has been treated as a complex problem and much of the innovation has occurred through happenstance and trial-and-error. However, the rapid scaling of data and high-performance computing are poised to completely disrupt the process of electrolyte design, turning it from a complex and intractable problem into a complicated but deterministic and rational problem. The key to reducing the complexity of electrolyte design is to recognize that, though electrolytes are made up of molecules, this is not a molecular design problem. The most important information exists in the cross-terms: how molecules interact with each other, and how they interact with the electrodes. In this article, we discuss how designing an electrolyte formulation in the computing age requires a re-thinking of AI tooling beyond the standard molecular property prediction paradigm.

AI- and machine learning-guided materials discovery efforts have demonstrated extraordinary progress in the last decade. During that time, we have gone through three generations of machine learning models for material- and molecular property prediction. The first generation of models leveraged handcrafted descriptors to build models that can predict the properties of single-component systems (i.e., single molecules or homogeneous crystals), an extension of the quantitative structure–activity relationship (QSPR) model framework developed nearly 60 years ago. The second generation involved utilizing graphs, a natural representation for bond connectivity within molecules, pairing them with two core machine learning advances from image processing—convolution and deep neural networks. This family of models, referred to as graph neural networks (GNNs), provided unprecedented accuracy and revolutionized the ability to screen large libraries of molecules. The third-generation models are molecular foundation models which are inspired by the structure of natural language and can leverage large swathes of unlabeled data (i.e., training on large databases of synthetically accessible molecules), using the model to build a more robust representation which can be fine-tuned for downstream property prediction.

We, the authors, have been working on this problem for the past decade—initially as academic scientists, and then as cofounders of Aionics, Inc., a company that uses AI and high-performance computing to develop high-performance electrolytes for diverse electrochemical systems. Beginning in 2014, we have been publishing studies developing data-driven and machine learning models for predicting battery material properties: first for liquid electrolyte electrochemical stability (VV),[2] then for solid-state Li-ion conductivity (ADS),[3,4,5,6] dendrite suppression properties (VV),[5,7] then for cathode properties including voltage, capacity, strain upon lithiation, and anion redox (VV, ADS)[8,9,10] and electrolyte-electrode interface kinetics (ADS, 2024).[11] Riding this wave of innovation, we founded Aionics in 2020 and began developing and commercializing machine learning models for battery performance. Working with manufacturers of materials, batteries, and battery-powered devices, descriptor-based methods were initially found to excel in cases where training data was extremely limited due to the expense of its generation—notably, for predicting the cycle life of a battery based on information about the electrolyte.[12] A few years later, Viswanathan and colleagues developed a suite of GNN-based electrolyte property prediction models with superior performance to descriptor-based methods.[13] These models were then commercialized by Aionics and deployed to the public under the Advanced Material Property Prediction Model framework.[14] Looking forward to the future, we are now developing molecular foundation models. Early data show unequivocally that these big-data models can vastly outperform all other models on some key tasks.[15]

Despite the algorithmic advances taking place in both the pure AI/ML community and the materials informatics community, electrolyte design remains a major challenge. Consider the simple case study below: we perform an electrolyte screening study of a known chemical space, and apply filtering criteria that would seem natural: (i) the molecule has to be stable in the battery (i.e., having appropriate values for the single component molecular levels—highest occupied molecular orbital and lowest unoccupied molecular orbital), and (ii) the molecule is a liquid at room temperature, having appropriate melting and boiling points. This simple set of screening criteria would eliminate ethylene carbonate, one of the most widely used electrolyte components. The issue is that the properties of individual molecules by themselves do not translate to electrolyte performance: information on the rest of the system is required to determine whether the formulation will remain liquid, or whether the cell will remain electrochemically stable. Thus, methods for incorporating information on the intricacies of the formulation and its interfaces with the electrodes are required to make significant advances in electrolyte formulation design—not just simply predicting the properties of molecules alone, no matter how sophisticated the algorithms or how accurate the predictions.

In general, electrolyte performance is a function of two sets of properties: innate properties of the formulation itself (ex situ) and properties of the formulation interacting with the chemical environment of the battery (in situ). The former are mixture properties, whereas the latter are interface properties. In both cases, molecular information alone is insufficient. Predicting the properties of mixtures requires information on the interactions of molecules with each other, while predicting the properties of interfaces requires information on the interactions of molecules with the electrode. No amount of information on individual molecules in vacuum can infer these cross-terms. Thus, a new class of ML models is required to supplement the simplified "molecule-only" framework.

(continued on next page)

Sendek and Viswanathan

(continued from previous page)

The canonical electrolyte design challenge involves adding a small amount of an additive to a baseline electrolyte formulation that otherwise works well under most of the required operating conditions. The goal of the additive is then to provide improved functionality in one or more axes (e.g., fast charging or lower temperature operation) while not affecting any of the other properties (e.g., cycle life). Thus, to minimize the overall cost of the formulation as well as to ensure that the majority of the properties of the baseline electrolyte formulation are carried forward, additive amounts are limited to 5% or less. The question then arises: how will the addition of a small amount of a new molecule change the overall properties of the electrolyte?

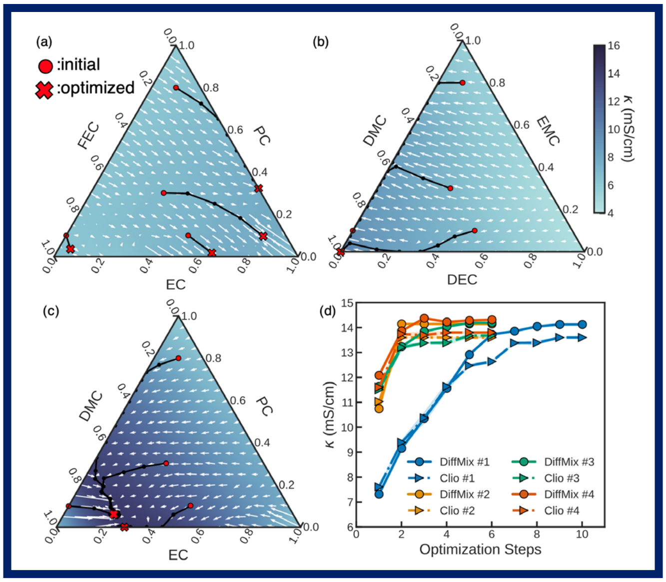

Fig. 2. *Differentiable battery electrolyte optimization with DiffMix and robotic experimentation. Reproduced with permission from ref. [13]. In (a)-(c), optimization for ionic conductivity is performed over various liquid electroyte compositions. In each optimization case, a batch of four trajectories has been simulated starting from the dot sign and ending at the cross sign. The white arrows are the gradient information obtained by auto-differentiating the ionic conductivity against compositions through DiffMix. In (d) the optimization curves of ionic conductivities in (c) is shown along the four trajectories, including both DiffMix results and the robotic experimentation results generated by Clio. Reproduced with permission from ref. [13].*

Mixture Properties

The simplest theory to predict the properties of a chemical mixture is linear mixing (i.e., the mixture properties are a molar-weighted average of the component properties). However, assuming linear mixing creates a big challenge here: adding a mere 5% additive will not substantially alter any linear property. As an example, assume that the melting point of the baseline electrolyte is -20^{o}C and an additive with a melting point of -100^{o}C is added at the 5% molar level. If the additive and baseline electrolyte linearly mix, then the new electrolyte formulation will have a melting point of -24^{o}C, a very small decrease of 4^{o}C in the overall electrolyte performance despite identifying an additive component that has an 80^{o}C difference in melting point. The goal then is to find additives which break the linear mixing rule for their given formulation. Fortunately, many properties are not naturally linear, though the direction and magnitude of the nonlinearity is often not obvious a priori. For example, in Fig. 1 we plot the predicted melting point of a set of electrolyte mixtures based on a linear mixing model versus the experimentally observed melting points. The correlation is poor; the components clearly do not linearly mix.

The deviation from linear mixing is called the *excess property*, and thus, designing functional electrolyte formulations requires identification of molecular mixtures that have large excess properties. We must understand how the additive and formulation interact, and then design new additives that mix super-linearly for the beneficial properties while mixing sub-linearly for harmful properties. Information on the individual molecules alone is an incomplete piece of the puzzle.

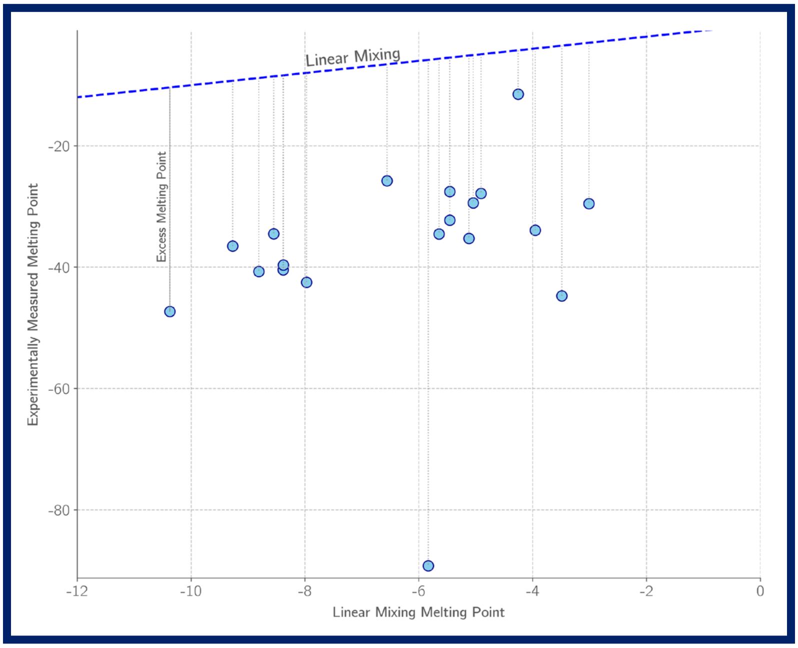

Fig. 1. *Linear mixing versus reality: We plot the experimentally measured melting point of electrolyte mixtures versus the predicted melting point assuming linear mixing of the individual components. The data would fall along the diagonal dotted line if linear mixing were correct. Instead, the correlation is poor and a substantial deviation from the linear mixing line can be seen. We acknowledge Stephanie Tarczynski, Jiayi Wu, Thuy Kim, and Varun Kumar for generating this image and data.*

Excess properties have been at the heart of chemical mixtures but have not been widely used in battery electrolyte design until recently. Excess functions have been formulated, with Redlish-Kister polynomials being one of the popular polynomial expansions that leverage the permutation invariance property of mixtures. Even for simple mixtures, there can be significant excess properties as molecular interaction amongst the components plays a critical role.

This situation highlights the need for incorporating additional physics into otherwise "straightforward" molecular ML models. Recently, we found a way to merge the two worlds: (i) machine learning advances involving graph convolution and (ii) mixture physics. The core insight involves learning the coefficients of mixture physics laws using machine learning, and then using these coefficients to add corrections from linear mixing. These coefficients are directly linked to the molecular identity of each of the constituent molecules, leading to mixture-specific corrections.

To realize this idea, the mixture physics laws need to be written in a differentiable programming framework, and then we can chain this to all the machine learning advances. This novel idea, which we called DiffMix,[16] has been put to test to design a high conductivity electrolyte formulation. DiffMix was used to navigate a ternary chemical space and the predictions were then validated using experiments in our robotic test stand, Clio.[17] The optimization trajectory measured from the robotic electrolyte setup, Clio, matched the predictions of the DiffMix model (Fig. 2).

The design of an electrolyte mixture is a balancing act among the various components of the mixture. Satisfying ex-situ property requirements cannot be accomplished by single molecular prediction models; instead a

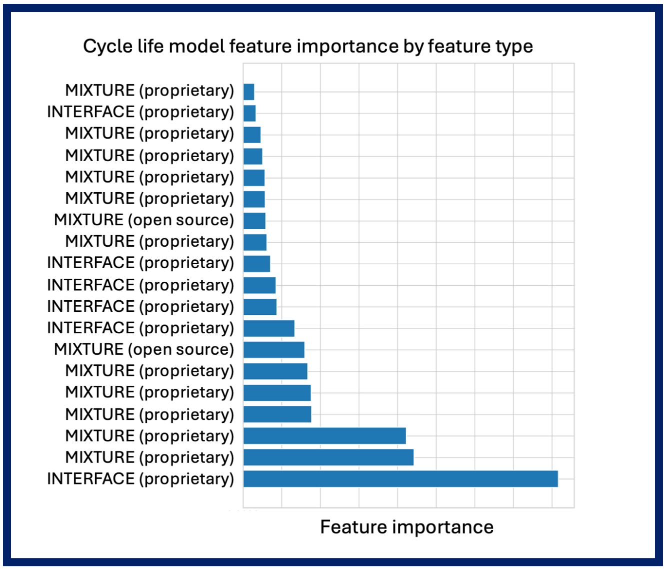

Fig. 3. *Importance of mixture and interface information for cycle life prediction. A model is trained on battery cycling data to learn to predict cycle life as a function of electrolyte composition. Proprietary interface features, proprietary mixture features, and open source mixture features are used. The features in the trained model are ranked by their importance. The single most important feature, by far, is an interface descriptor. We acknowledge Drs. Noushin Omidvar, Handong Ling, and Mohamed Elshazly for generating this image and data.*

sophisticated machine learning stack must be employed that can handle mixture physics and the underlying molecular interactions amongst the mixture components.

Interface Properties

Of equal or greater importance to the ex-situ properties of electrolytes are the in-situ properties of the electrolyte when added to a battery. In this case, the properties are determined by the interactions of the electrolyte components with themselves but also primarily with the electrode surface. Again, the properties of the individual molecules alone are insufficient.

Relevant in-situ properties include cycle life, charging rate capability, discharging rate capability, internal resistance, coulombic efficiency, round-trip efficiency, and self-extinguishing time.

All these properties require some knowledge of the interactions at the electrolyte-electrode interface, where electrolyte molecules may reversibly bond with the electrode surface. By extending concepts from computational surface science and catalysis, we published the first series of studies showing that the properties of batteries can be predicted directly from density functional theory (DFT) simulations of electrolyte molecules on electrode surfaces.[18,19] In these simulations, electron transfer is simulated explicitly and breakdown of the molecules under the electrochemical and chemical forces of the interface can be observed.

Consider cycle life as an example property. The cycle life of a battery is a simple metric of battery life that arises as a multidimensional function of many phenomena happening in the bulk of the electrolyte solution and at the electrode interface. However convoluted the process may be, cycle life is a deterministic function of the electrolyte and electrodes and thus should be learnable with sufficient examples. In a recent internal study, we built a model to predict the cycle life of a Li-ion battery as a function of its electrolyte based on a small dataset of under 200 examples. The resulting model had a cross-validation error of 41 cycles, meaning the model is expected to predict the cycle life of a cell with an arbitrary (in-distribution) electrolyte to within 41 cycles on average.

It is instructive to look at the information required to make these predictions. We trained the model on a set of proprietary interface-based descriptors, proprietary mixture descriptors, and open-source descriptors; the relative importance of these various descriptors is shown in Fig. 3. Instead of focusing on the exact definition of each descriptor, the emphasis is on whether the descriptors depend on mixture alone or of the mixture and electrode interface. The single most important descriptor in the model contains interface information.

Once the appropriate descriptors are identified for in-situ property prediction, they must be extracted for the entire chemical space of interest to enable large-scale screening. However, this screening quickly becomes computationally intractable with traditional density functional theory methods. Surface calculations require many atoms, some of which are best simulated with Gaussian-type basis sets (i.e., the molecules) and some of which require plane wave basis sets (i.e., the surface). Many molecular orientations and possible surface adsorption sites exist. The computational cost to exhaustively study the interfacial effects of all molecules on all relevant surfaces can easily reach into the trillions of dollars. Thus, some machine learning acceleration is required.

Using the output of a large, expensive model to train a simpler model with minimal accuracy loss is a common task in machine learning, and it is a helpful strategy here. This approach of building computationally cheap surrogate models is sometimes referred to as distillation, pseudo-labeling, or transfer learning, depending on the context. In the case of interfacial features, we have found that simple models can learn these features efficiently from hundreds to thousands of examples and can perform inference on new systems with high fidelity and a 10^6-fold increase in speed over DFT, enabling screening across the entire space of molecules for any given interface.

Discussion

Currently, there are approximately 10^{10} molecules that can be synthesized commercially, and there may be a total of over 10^{50} synthetically accessible molecules. There are many efforts underway to study the properties of these known and hypothetical molecules for varied kinds of applications, particularly within medicine. For electrolyte design, however, this is just the tip of the iceberg. Electrolytes cannot be designed successfully without considering the effects of combining these molecules into mixtures and explicitly considering their interfaces with electrodes. This combinatorial effect makes the design space unfathomably large: there are 10^{55} ways to combine all 10^{10} commercially accessible molecules into two-component mixtures; when you mix three of them, you have 10^{165} options to consider. For comparison, the current age of the universe is about 10^{17} seconds. Like the astronomical universe, the sheer size of the mixture universe is staggering, but also a source of great opportunity and excitement if it can be successfully navigated.

There is one more important set of considerations in electrolyte design: commercial requirements. Fortunately, this is the only set of properties where information on each molecule individually is sufficient. Cost, synthetic accessibility, feasible vendors, supply chain resiliency and more can all be inferred directly from the structure of each component alone. This has been a historically overlooked area of electrolyte design, but some capabilities are beginning to emerge: for example, models have been trained to predict the cost to procure a small sample of a molecule by learning directly on its molecular structure and by considering its retrosynthetic pathway.[20,21]

In sum, the recipe for reducing electrolyte design from a complex problem to a merely complicated one then requires not just knowing the properties of molecules; it requires knowing the excess properties of their mixtures and understanding their interactions with electrode interfaces. Molecular property prediction can be done using existing methods, but mixture and interfacial properties require new methods that incorporate these complex interactions. Then, a second layer of machine learning models must be deployed to decrease the computational cost of these predictions such that a significant portion of the space of mixtures can be traversed.

This philosophy underpins the Aionics electrolyte design platform, which has been built and deployed to accelerate electrolyte design for many battery designs and applications. Our platform contains a library of over 10^{10} molecules and their predicted physical and commercial properties, plus the ability to predict values for new

(continued on next page)

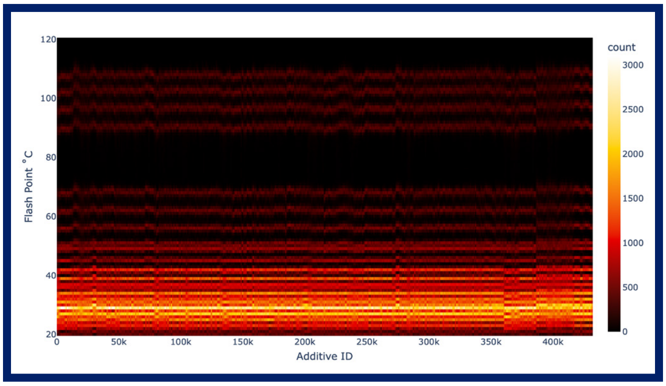

Fig. 4. *High-throughput flash point optimization. Using a flash point model trained on experimental data and the DiffMix framework, the Aionics platform computes the predicted flash point of approximately 18 million modifications of a baseline formulation. Each modification contains one novel molecular additive and a unique ratio of the baseline components. Many of the sampled variations have mixture flash points around 30C, but there are notable outliers with substantially higher and lower flash points. Mixture models that contain excess property physics like DiffMix are necessary to identify these unintuitive and highly important outliers. We acknowledge Drs. Mohamed Elshazly and Shreyas Honrao for generating this image and data.*

molecules that are not yet commercially known. Relationships with chemical suppliers have enabled the development of models for predicting how cost will scale with volume. Most importantly, the platform supports computationally combining these molecules into any possible mixture so that their emergent mixture and interfacial properties can be predicted. Millions of variants on existing mixtures can be simulated and assessed in the span of minutes. To illustrate this, in Fig. 4 we show predictions of the flash point of approximately 18 million unique electrolyte formulations, each of which is composed of one of 36 variants of a baseline formulation plus any one of approximately 500,000 commercially available additives. These computations execute at a rate of approximately 3,000 mixtures per second on modest CPU resources; this map of 18 million mixtures took less than two hours of wall time to generate. This capability represents one of the largest databases of electrolyte formulations and their physical properties, and to our knowledge is the only major database of electrolytes that also contains commercial properties. The methods described in this article—molecular property prediction, mixture excess property prediction, interface interaction surrogates, and price prediction—are among the key advances underlying this state-of-the-art electrolyte design platform. All relevant intellectual property that we developed in the university setting has been licensed exclusively to Aionics.[22]

Looking forward to the future, we encourage the development of new methods that capture the cross-terms that are inherent in electrolyte design solutions: molecule-molecule ex-situ interactions, and molecule-surface in-situ interactions. Despite the massive space of molecules available to us, we believe that approaching the problem as merely a molecular design challenge is an oversimplification. The space of mixtures made up of these molecules is orders of magnitude larger, raising the prospect that we can discover new, high performance electrolyte blends for any electrochemical cell. Existing efforts in molecular property ML models, particularly by the pharmaceutical chemistry community, have laid a solid foundation; it is now up to the computational electrochemistry community to develop the domain specific models that are required to push the field forward.

Acknowledgements

This article was enabled and supported by the following contributors: Mohamed K. Elshazly, Shreyas Honrao, Varun Kumar Karrothu, Thuy B. Kim, Handong Ling, Noushin Omidvar, Stephanie Tarczynski, and Jiayi Wu. We thank and acknowledge their direct contributions to the data, models, and figures. We also wish to thank the remainder of the team at Aionics, Inc. whose hard work helped bring this analysis to life. Additionally, we thank all our colleagues over the years who contributed substantially to the previous papers discussed here, including Drs. Yumin Zhang, Shang Zhu, Zeeshan Ahmad, Gregory Houchins, Vikram Pande, Emil Annevelink, Rachel Kurchin, Adarsh Dave, Hongyi Lin, Ekin Dogus Cubuk, Brandi Ransom, Nicholas Grundish, Qian Yang, and Evan J. Reed. We thank Meghan Hayes and the Center for Technology Transfer and Enterprise Creation at Carnegie Mellon University for supporting our commercialization efforts of the intellectual property developed in the Viswanathan lab.

© The Electrochemical Society. DOI:10.1149/2.F08252IF

About the Authors

Austin D. Sendek, Co-founder and CEO, Aionics, Inc.
Education: BS Applied Physics (UC Davis); PhD Applied Physics (Stanford University)
Research Interests: Energy storage, Machine learning, Materials informatics
Work Experience: Adjunct Professor of Materials Science and Engineering, Stanford University, Stanford, CA, USA
Pubs + Patents: 19 publications, h-index 12
Awards: CB Insights AI 100 (Aionics), Fast Company's World Changing Ideas 2024 (Aionics), Forbes 30 Under 30 in Energy 2019
Website: aionics.io
https://orcid.org/0000-0003-3338-1615

Venkatasubramanian Viswanathan, Co-founder and Chief Scientist, Aionics, Inc.
Education: BTech in Mechanical Engineering (Indian Institute of Technology, Madras); PhD in Mechanical Engineering (Stanford University)
Research Interests: Energy storage, Machine learning, Materials informatics
Work Experience: Associate Professor of Aerospace Engineering, University of Michigan
Work with ECS: Active member of Electrochemical Society and winner of San Francisco Section Daniel Cubicciotti Student Award, Herbert H. Uhlig Summer Fellowship
Awards: MIT Technology Review Innovators Under 35, Office of Naval Research (ONR) Young Investigator Award, Alfred P. Sloan Research Fellowship in Chemistry, National Science Foundation CAREER award and American Chemical Society Energy Storage Lectureship
Pubs + Patents: 150+ publications, 50+ patents
Website: https://aionics.io; https://eeg.engin.umich.edu/
https://orcid.org/0000-0003-1060-5495