ACS OMEGA

Open Access



http://pubs.acs.org/journal/acsodf

Article

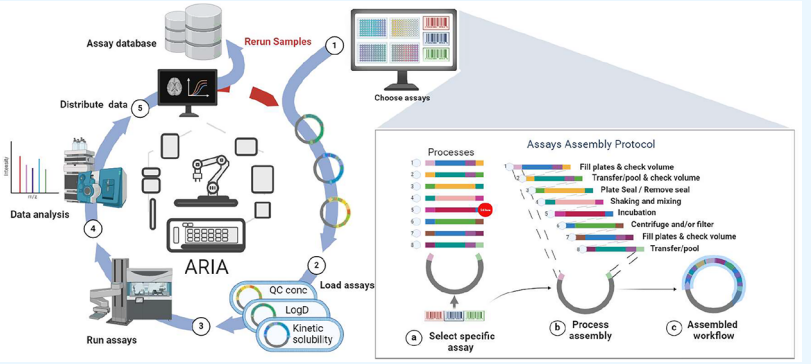
# An Automated Robotic Interface for Assays: Facilitating Machine Learning in Drug Discovery by the Automation of Physicochemical Property Assays

Newton P. Wu, Wenyi Wang, Dhireshan Gadiagellan, Mike Counsell, Nikkia K. Hamidi, Yuko Koike, and Huy Q. Nguyen*

Cite This: *ACS Omega* 2024, 9, 24948−24958

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information




**ABSTRACT:** Measuring the physicochemical properties of molecules is an iterative but integral process in the drug development process. A strategy to overcome the challenges in maximizing assay throughput relies on the usage of *in silico* machine learning (ML) prediction models trained on experimental data. Consequently, the performance of these *in silico* models are dependent on the quality of the utilized experimental data. To improve the data quality, we have designed and implemented an automated robotic system to prepare and run physicochemical property assays (**A**utomated **R**obotic **I**nterface for **A**ssays, ARIA) with an increase in sample throughput of 6 to10-fold. Through this process, we overcame major challenges and achieved consistent reproducible assay data compared to semiautomated assay preparation.

## INTRODUCTION

In most drug development pipelines, measuring physicochemical properties such as lipophilicity (Log*P*/Log$D_{7.4}$) and kinetic solubility of drug molecules is crucial to improve their drug like properties. Measuring Log$D_{7.4}$ and kinetic solubility has not changed much in the last 20 years.[1−7] The gold standard for measuring Log$D_{7.4}$ remains the shake-flask method.[8,9] Similarly, for kinetic solubility, high-throughput assays in microplates with concentration measurements at early time points (≤24 h) have been commonly used.[7] Additionally, it is important to validate synthesized drug molecules by performing quality control assays, such as those for determining purity and stock concentrations. Acquiring this data quickly and as accurately as possible is paramount in progressing project portfolios. To that end, many programs rely heavily on utilizing *in silico* models to predict these physicochemical properties to design and advance new drug candidates.[10−12]

Training machine learning (ML) models to predict Log$D_{7.4}$ and kinetic solubility is especially important early in the process when medicinal chemists rely on predicted physicochemical properties to design the next set of molecules for synthesis.[13,14] Additionally, these predicted properties are used to prioritize molecules in order to triage for downstream resource demanding biochemical assays (Figure 1). To make sure the machine learning models are performing well, the models are routinely updated with new experimental data. Therefore, it is important to make sure that the experimental assay conditions, and assay execution are consistent. Inconsistency in the data generated when there is a change in assay conditions (different technicians, locations, etc.) will lead to variability and directly affects ML models trained on this data.[15] To reduce variability and improve the data quality for training and updating ML models, we have automated all of

Received: February 29, 2024
Revised: May 14, 2024
Accepted: May 17, 2024
Published: May 24, 2024

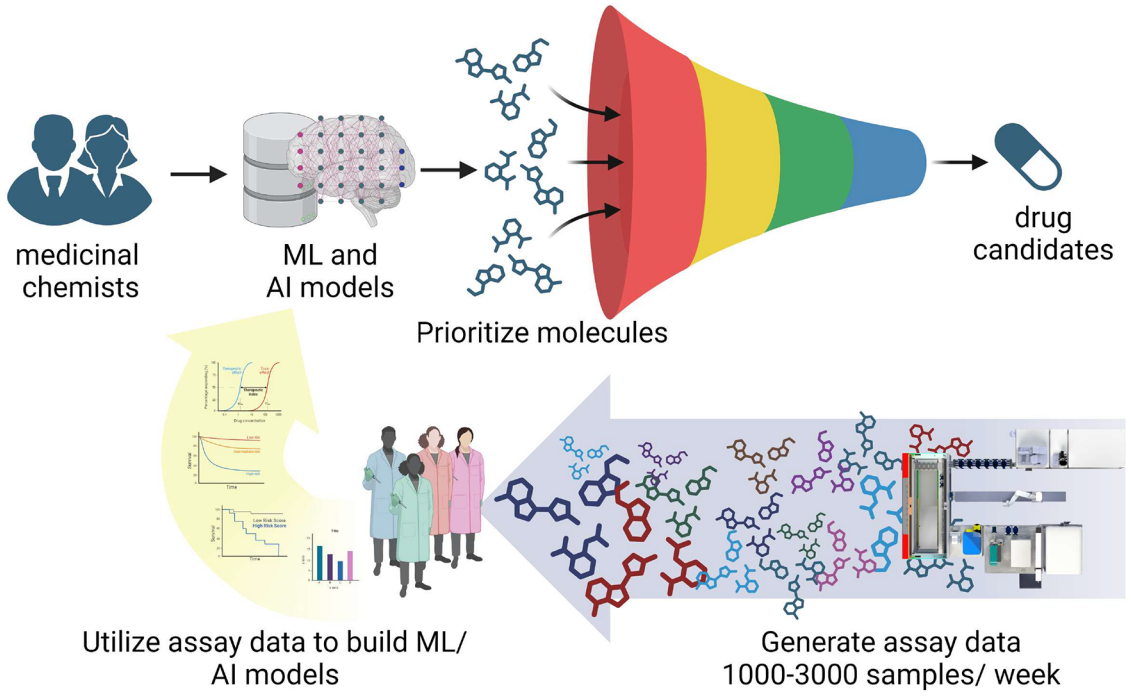



**Figure 1.** Drug development pipeline assisted by ML/AI models, built from high-throughput physicochemical properties experimental data, generated by analytical chemists and the ARIA system.

our physicochemical assays utilizing an automated robotic interface for assays (ARIA, Figure 2).

Designing and building robotic systems to automate different types of lab work is not a novel concept, as scientists and engineers have been improving these type of systems for the last 40 years.[16−21] Robotic systems are now able to automate different types of chemistry from analytical to total synthesis.[22−26] Additionally, automation in the pharmaceutical industry has been a necessity, especially for high-throughput screens.[27,28] Major advantages associated with robotics are the increase in throughput, consistency, and the reproducibility of the data generated. With the ever-growing and -evolving hardware and software options, having a modular and reconfigurable system is essential not only for long-term adoption but also for sustainability.[21] Alongside technological progress in hardware, recent advancements in software have changed the way robotic systems could be designed to be more accessible and open to collaboration.[23,29,30]

Even with the advantages that automation brings, there remains a need for validation steps to ensure that the data generated is accurate. Being able to identify and correct systematic errors is of utmost importance when migrating to an automated system. In this paper, we will discuss these validation steps which were engineered into the ARIA system to generate reliable and highly reproducible data for ML in-house models. Fundamentally, ARIA is a platform that brings together a multitude of assay modules, which facilitates the running of many relevant assay types (Figure 2). Careful workflow considerations were taken to address the space utilization and temporal optimization of the robot in order to minimize labware movements, maximizing system throughput. Additionally, we will also discuss software design and how we developed the digital workflow assembly line.

## RESULTS AND DISCUSSION

**ARIA: Platform for Modular Assembly.** To automate all of the physicochemical assays (Log$D_{7.4}$, kinetic solubility, QC/DMSO stock concentration determination), ARIA was designed and engineered as a platform to execute assays by assembling combinations of modules unique to each assay (Figure S1). With automated robotic platforms, a major challenge has been establishing software simple enough to develop new assay workflows as well as engaging collaborations with other robotic systems or external groups. Groups have already developed automation programming software to simplify and make their automation workflow development more robust.[23,29,30] In particular, inspired by the Cronin group's work on the "chemputation framework", which focuses on creating a universal machine controlled by a common software package for the execution of chemical processes, we have adopted a similar concept to develop and refine digital workflows for the ARIA system.[31]

A challenge in the development of systems like ARIA is the translation of devices used in manual assay execution to devices that can be incorporated into an automated system. Devices such as the LabRam II mixer (Resodyne) are very well-suited for manual human operation but lack automation ability off the shelf. Early decisions in the project to develop custom automation allowing for the incorporation of fit-for-purpose devices such as the Labram were instrumental in the success of ARIA.

Leveraging existing technologies, we integrated scheduling software (Revolution, UK Robotics ltd) with liquid handling controls (FluentControl, Tecan) to streamline our processes. This integration mirrors the principles of the Gibson assembly process, a method we adapted as a conceptual model for developing assays on the ARIA system.[32] Just as the Gibson assembly combines DNA fragments to construct genes, ARIA combines assay modules to construct digital workflows (Figure


https://doi.org/10.1021/acsomega.4c02003
ACS Omega 2024, 9, 24948−24958

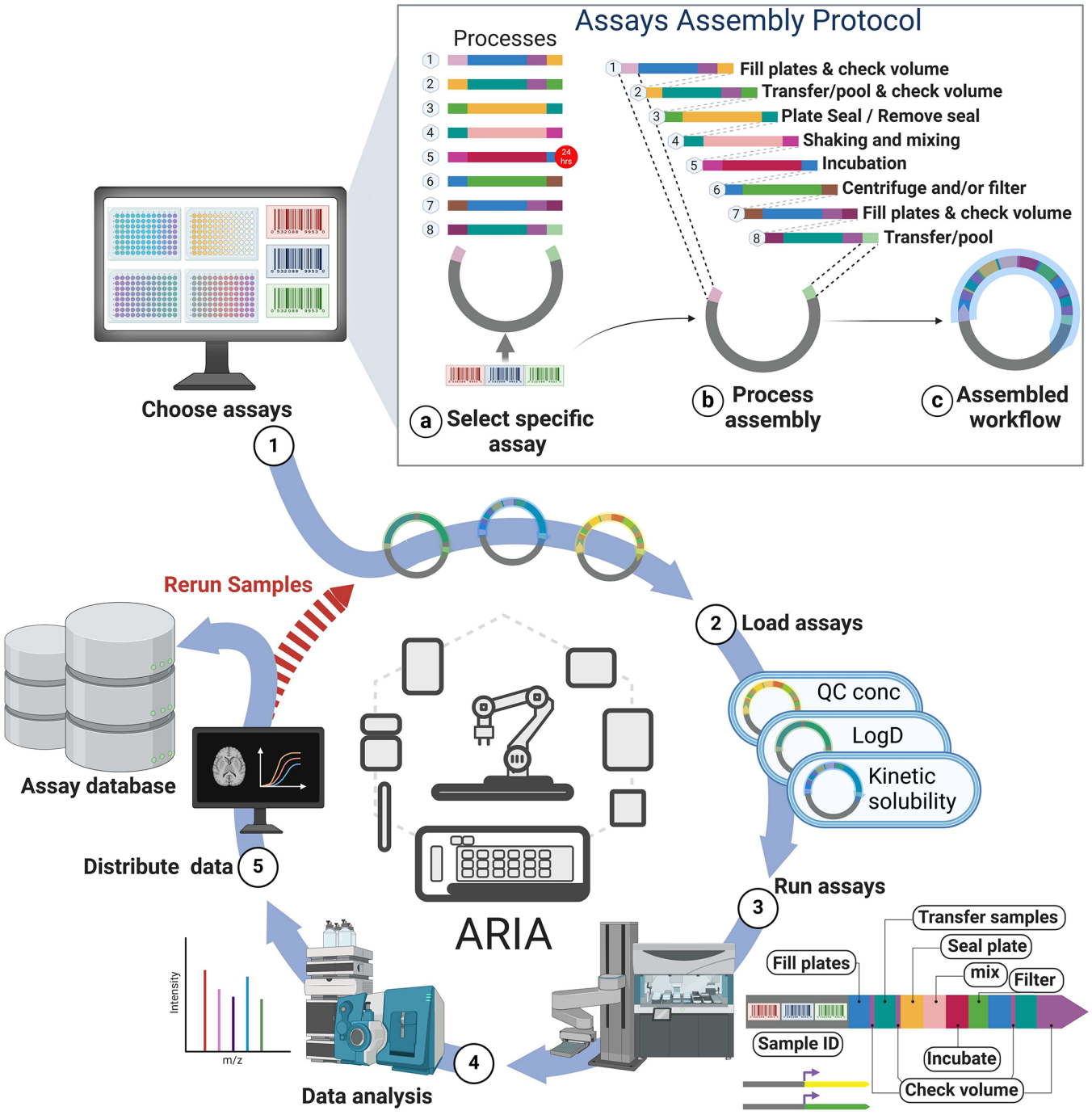



**Figure 2.** Overview of the ARIA workflow for physicochemical property assays. (Step 1) Gray: users prompt ARIA to scan samples from compound management and retrieve metadata for all the samples (ID, molecular weight, etc.). Specific assays are selected (option to modify assay parameters) and the necessary ARIA modules are assembled together. (Step 2) Selected assays are loaded with user defined parameters and initiated. (Step 3) ARIA prepares the sample plates and is optimized to run in parallel when possible to increase throughput. (Step 4) After completion the assays plates are analyzed by LCMS. (Step 5) Analyzed data is uploaded onto a central database or the samples are rerun dependent on the results from step 4.

2). By engineering the assay programming in this assembly process, developing new assays becomes trivial and allows more time spent on assay validation.

**System Validation.** To validate the ARIA system as a viable platform for running various physicochemical properties assays, we conducted a direct comparison between assay data from samples prepared semiautomated and those prepared by the ARIA system. Samples measured were received from our compound management team and prepared to be around 10 mM in DMSO (QC assay). Measuring DMSO stock concentrations for these samples confirms the actual stock concentrations for any downstream assays. Additionally, measuring the concentration of sample in DMSO is quantitative with only one dilution step before analysis, making this assay a good litmus test for comparing results between the two methods.

Taking a closer look at the QC data, the average DMSO stock concentration prepared by ARIA (9.7 ± 2.5 mM) was



https://doi.org/10.1021/acsomega.4c02003
ACS Omega 2024, 9, 24948−24958

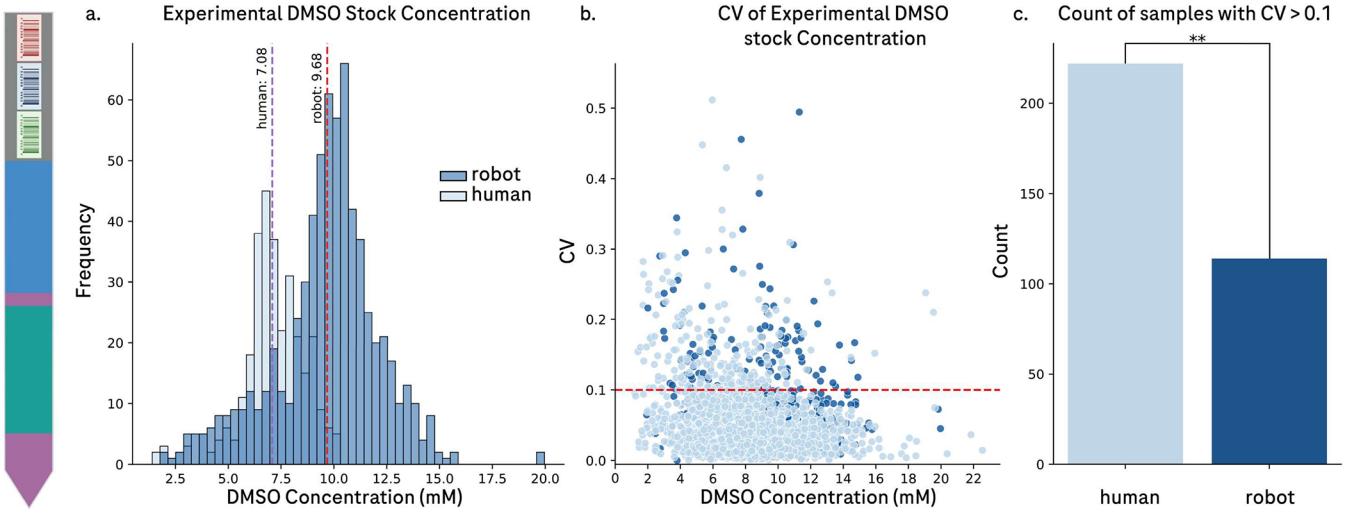



**Figure 3.** QC DMSO stock concentration data compared between experimental samples from semiautomated workflow and the ARIA system. (a) Histogram of measured DMSO stock concentrations determined by semiautomated methods (light blue) and by the ARIA system (dark blue) with the dotted lines representing the average concentration for their respective methods. (b) The CV calculated at every measured DMSO stock concentration for both semiautomated (light blue) and the ARIA system (dark blue) with the red dotted line at 10% CV. (c) Number of experimental DMSO concentrations above 10% CV for both methods and the statistical significance with $p$-value < 0.01.

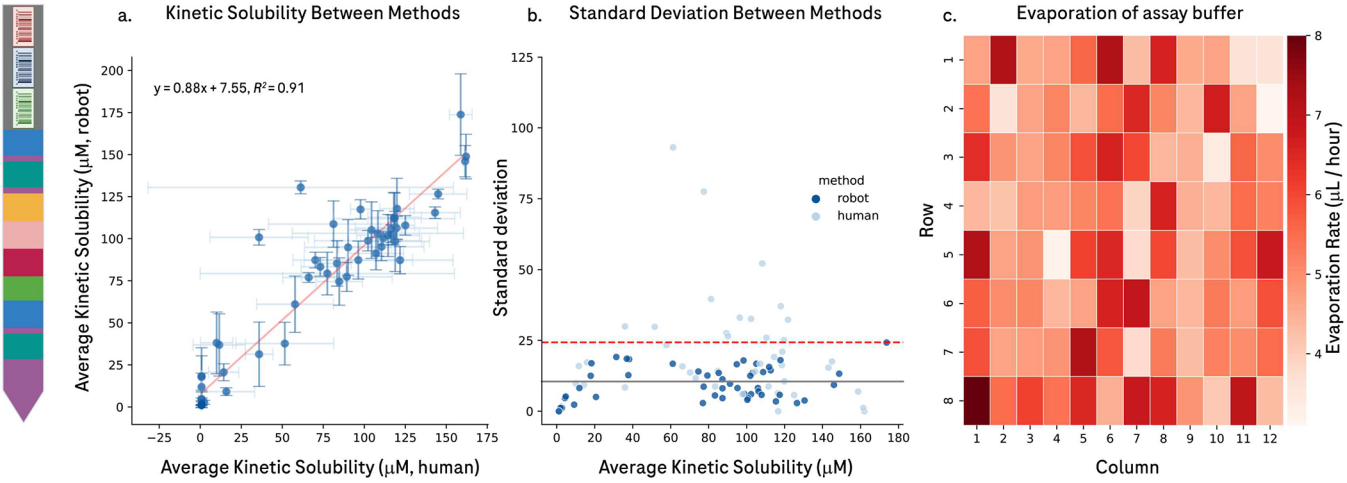



**Figure 4.** Experimental kinetic solubility between samples prepared using a semiautomated method (light blue) vs ARIA (dark blue). Buffer evaporation was taken into account for the kinetic solubility data prepared by the semiautomated method. (a) Error bars correspond to the standard deviation for the specific methods ($x$-axis: semiautomated, $y$-axis: ARIA). (b) Standard deviation of measured kinetic solubility for both methods with the average standard deviation (gray line) and mean of standard deviation (dotted red line). (c) Heat map representing a 96-well plate with the rate of evaporation ($\mu$L/hour) for each well position.

closer to the expected 10 mM compared to the observed average DMSO stock concentration from semiautomated preparation (7.1 ± 1.6 mM) (Figure 3a). After careful examination, the difference in the results was due to inaccurate liquid aspiration during the semiautomated preparation which uses a semiautomated pipettor (Sorenson Bioscience Inc.). When properly calibrated, the semiautomated pipetter performs very well; however, with extensive usage the equipment required more frequent calibrations to maintain optimal performance. Although we employ calibration standards for every sample batch to ensure accuracy, there are still occasions when calibration errors elude detection. In the ARIA system, each liquid handling step is recorded as part of an extensive audit trail and monitored through volume verification (ultrasonic, and gravimetric), facilitating the identification of errors vital for downstream troubleshooting. This tracking capability, along with ensuring continual calibration of every instrument, effectively addresses the issues noted with the semiautomated pipettor. As evident from the coefficient of variation (CV) where we observed twice the amount of samples with a %CV > 10% from the semiautomated workflow (Figure 3b,c). These results were promising and helped us realize the potential of the ARIA system, especially as we moved forward with more complex assays.

Building upon our previous success with the QC assay, data from the kinetic solubility assay further supported the ARIA system's efficacy. A direct comparison of the kinetic solubility data between the two methods demonstrated a strong correlation ($R^2$ = 0.91, as shown in Figure 4a), underscoring the consistency of the ARIA system. However, there were significantly larger standard deviation observed in the kinetic solubility data from samples prepared using a semiautomated


https://doi.org/10.1021/acsomega.4c02003
ACS Omega 2024, 9, 24948−24958

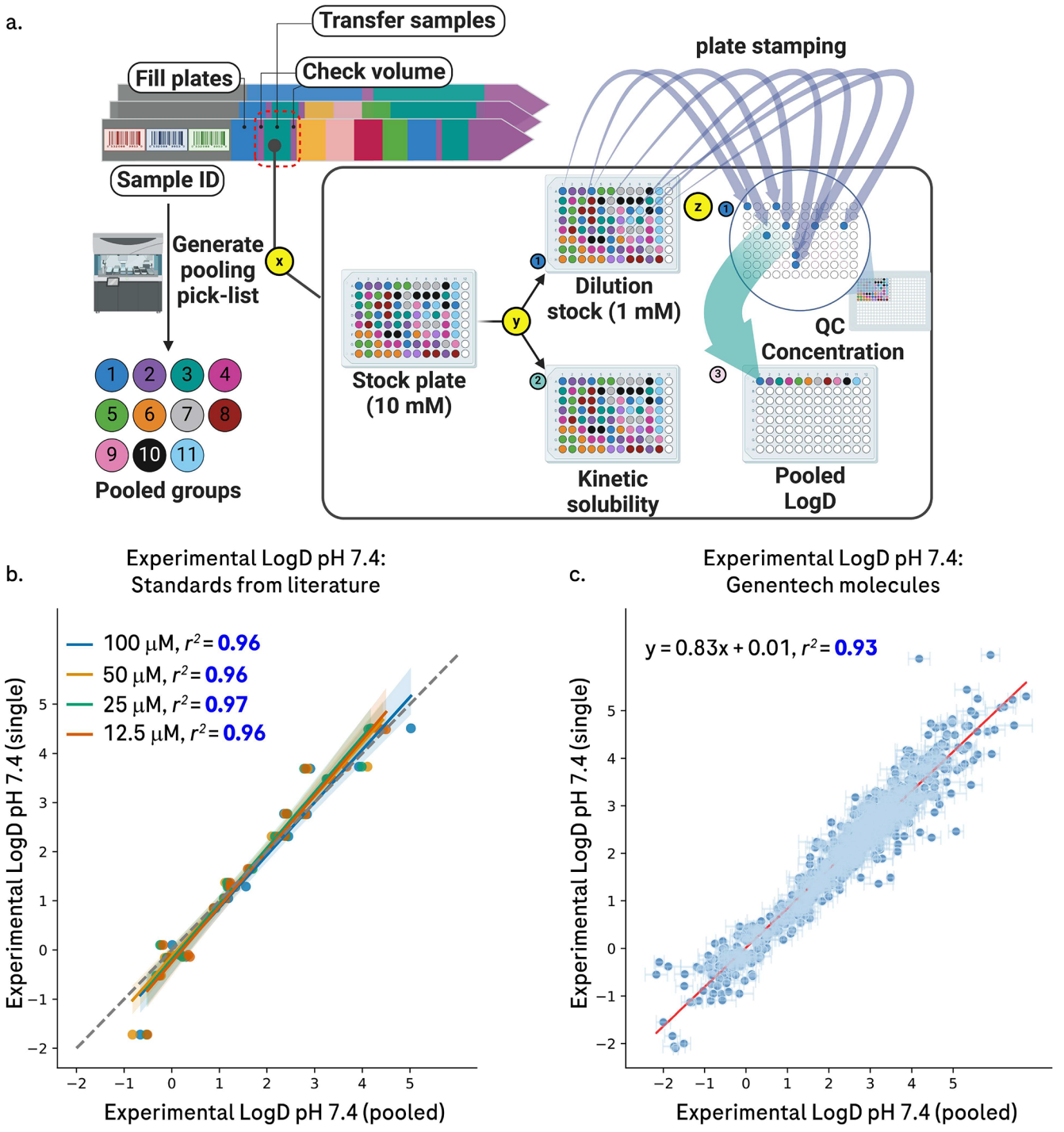



**Figure 5.** Parallel programming to optimize the liquid transfers for all three assays. (a) Schematic for 2 multidispensing steps, (y) to prepare the kinetic solubility plate (4 $\mu$L) and 1 mM stock dilution plate. (z) Another multidispense for the pooling Log$D_{7.4}$ samples and plate stamping into a 384-well plate for the QC assay. (b) Correlation between pooled Log$D_{7.4}$ and Log$D_{7.4}$ measured as single samples for a set of standards at varying concentrations (12.5 to 100 $\mu$M). (c) Correlation between pooled Log$D_{7.4}$ and Log$D_{7.4}$ measured as single samples for a set of Genentech compounds ($n$ = 1309 molecules).

workflow with standard deviations above 25 $\mu$M (Figure 4b). This variance was attributed to the more complex kinetic solubility assay, which involves numerous liquid transfers and filtering steps, increasing the likelihood for errors. Furthermore, after an extensive investigation into each assay step, we discovered buffer evaporation was a major contributor to the variability in the assay results (Figure 4c). Depending on the specific well location, we observed a consistent evaporation rate no lower than 3.1 $\mu$L/hour and up to 8.0 $\mu$L/hour. To account for buffer evaporation, careful steps were taken to ensure assays plates were sealed in between liquid handling steps to reduce volume loss. Importantly, buffer evaporation was not as significant with samples prepared by ARIA resulting in a lower standard deviation. In contrast to the Log$D_{7.4}$ assay, the ratio of sample intensity in PBS and octanol, the kinetic solubility assay is quantitatively similar to the QC assay. This


https://doi.org/10.1021/acsomega.4c02003
ACS Omega 2024, 9, 24948−24958

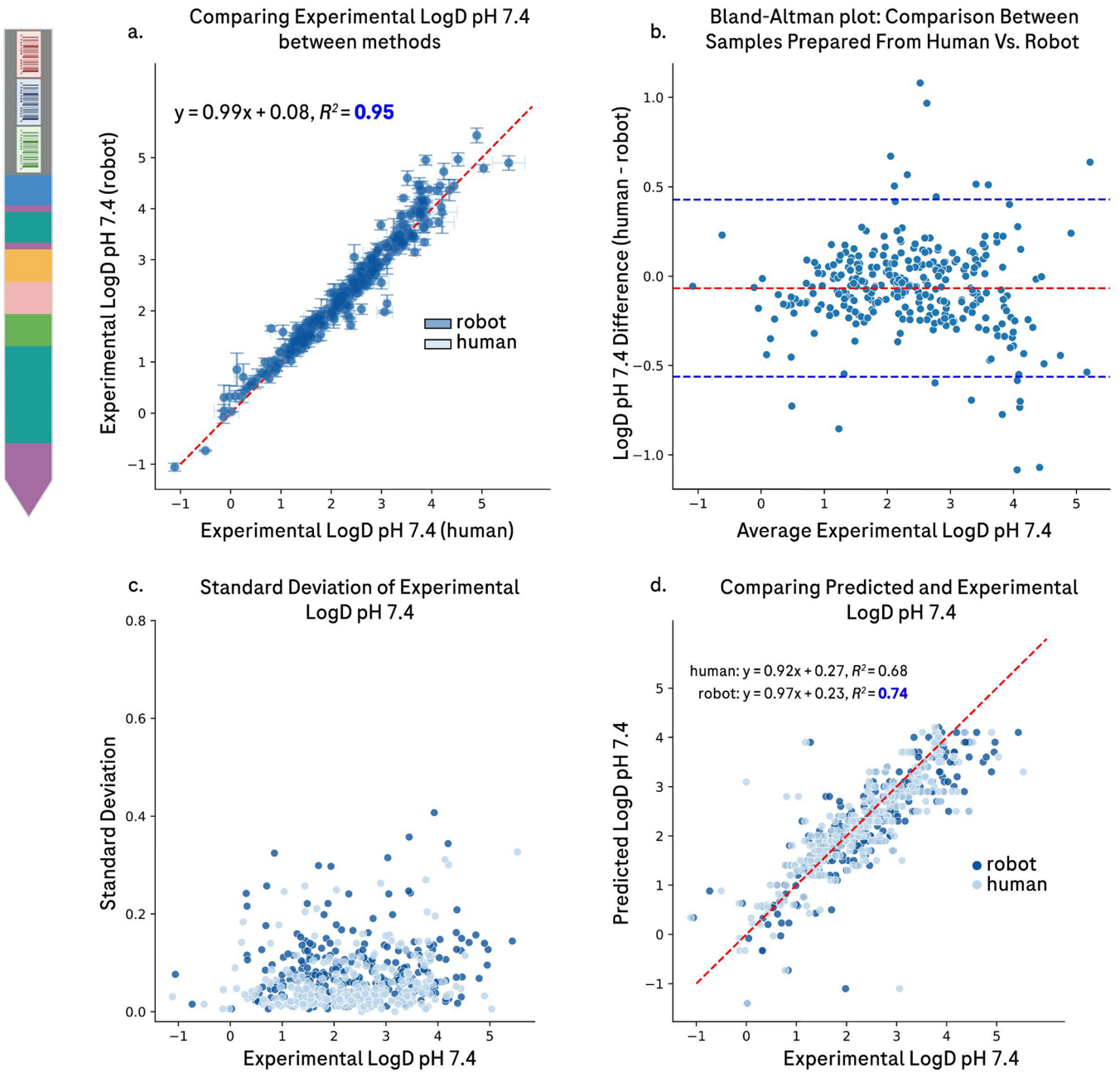



**Figure 6.** Experimental Log$D_{7.4}$ data at pH 7.4 measured from samples prepared by both methods, human (light blue) and by ARIA system (dark blue). (a) Direct comparison of experimental Log$D_{7.4}$ pH 7.4 between both methods ($n$ = 590 molecules). (b) Bland–Altman plot demonstrating the difference between experimental Log$D$ pH 7.4 between the two methods.[34] (c) Standard deviation for every sample for both methods. (d) Comparison of experimental Log$D_{7.4}$ for both methods and how they relate to the predicted Log$D_{7.4}$ using an in-house prediction model built on measured Log$D_{7.4}$ at Genentech.[35]

methodological difference leads to greater variance and measurement inaccuracies in the semiautomated preparations. The ARIA system addresses these validation challenges effectively and also provides additional opportunities to optimize current workflows.

**Increasing Throughput and Optimizing Workflow.** Leveraging on the ARIA system's ability for parallel programming, and a custom developed business logic layer (BLL), the Tecan liquid handling transfer steps for all three assays were combined to increase throughput and reduce the amount of plastic tips consumed (Figure 5). The first sample transfer from 10 mM DMSO stock was a multidispensing step that created the kinetic solubility assay plates and the dilution plate for the QC and Log$D_{7.4}$ assays. To increase the throughput and reduce the number of liquid handling steps, samples were pooled into groups of 8 with each sample in the group having a molecular weight that is different from the other samples in the same group. By pooling samples, we were able to quantify the mixture of samples with a triple quadrupole mass spectrometer's ability for multiple reactions monitoring (MRM).[33] To accomplish this with the ARIA system, a master pick list was automatically generated by the BLL for the Tecan to cherry pick and pool samples into 1 mM DMSO stock solutions (Figure 5). The pooled samples were


https://doi.org/10.1021/acsomega.4c02003
ACS Omega 2024, 9, 24948−24958

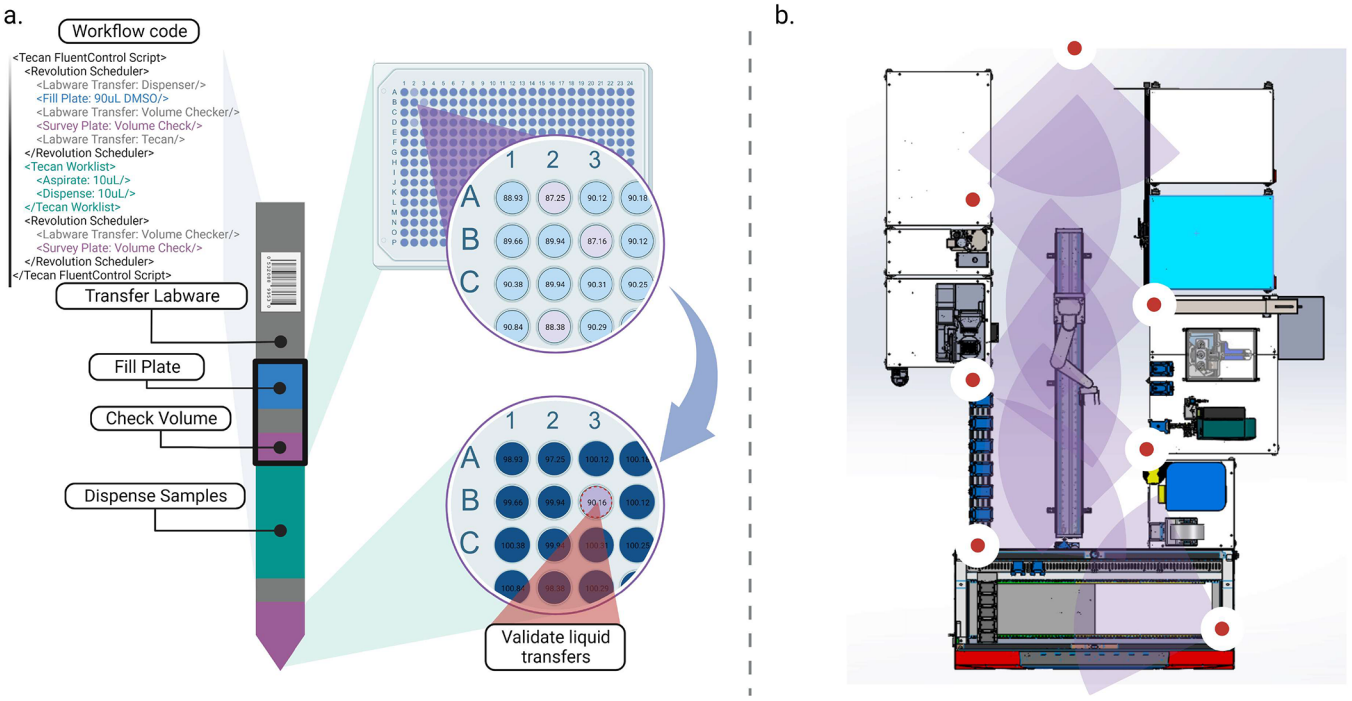



**Figure 7.** Strategies and hardware in place for assay validation. (a) Validating liquid transfer with an ultrasonic volume checker to assist in identifying liquid transfer errors. (b) Utilize CCTV cameras to troubleshoot in real time or observe and identify when errors occurred during the assay.

then transferred to the assay microplates containing PBS/octanol and then separated for analysis. By pooling the samples, an entire 96-well plate was condensed into one row (12 wells), and technical replicates would only be an additional 2 rows for triplicates. Additionally, using fewer plastic microplates simplified the workflow and supported our aim to reduce environmental impact in research. Careful considerations were taken to account for possible electrostatic interactions between samples (acid and base at pH = 7.4). There were no significant effects on measured Log$D_{7.4}$ for a group of compounds containing both acids and bases at varying concentrations (Figure 5b). Furthermore, measured pooled Log$D_{7.4}$ for a set of Genentech's molecules had good correlation ($R^2$ = 0.93, Figure 5c) with single Log$D_{7.4}$ values for the same set of compounds.

After optimizing the Log$D_{7.4}$ assay with pooling samples for MRM analysis, we observed promising results between observed Log$D_{7.4}$ values from samples prepared by a semiautomated method and by the ARIA system ($R^2$ = 0.95, Figure 6a). Furthermore, the Log$D_{7.4}$ difference for both preparation methods in relation to the mean Log$D_{7.4}$ values revealed that most of the observed Log$D_{7.4}$ values fall within the 95% confidence interval Figure 6b.[34] These results suggest that the samples prepared by the semiautomated method and by the ARIA system were similar. Moreover, the standard deviation for both methods were all below 0.5 log units for observed Log$D_{7.4}$ values Figure 6c. The Log$D_{7.4}$ values measured by the ARIA system demonstrated a notably higher correlation with predicted Log$D_{7.4}$ values ($R^2$ = 0.74, in-house model) than those obtained from the semiautomated method ($R^2$ = 0.68, in-house model, Figure 6d). This significant improvement in data quality from the ARIA system underscores the advantages of automation in enhancing the reliability of machine learning predictions.

**Enhancing Troubleshooting with Data Transparency and Hardware Monitoring.** The ARIA system's meticulous recording of each step by the BLL, opens new avenues for understanding and refining the assay methodology. This level of detail provided invaluable insights into automating the assay preparations with ARIA. The ARIA system's configuration, encompassing a web application, a custom business logic layer, scheduling software, and the Tecan liquid handling robot, establishes a robust framework for both data transparency and hardware monitoring.

Central to the ARIA system is an in-house web application (ARIA's interface) and business logic layer that plays a pivotal role in data collection (Figure S3, a detailed description of ARIA's architecture). Every plate processed is tagged with a unique barcode, which is directly linked to the specific assay it belongs to. This meticulous tracking of metadata for each plate ensures thorough documentation throughout the assay process. The web application not only facilitates control over the assay selection but also efficiently manages the sample data. The system scans barcodes on sample tubes, retrieves essential information like molecular weight and chemical structure, and tracks each sample throughout the assay to ensure completion of all designated tests. The storage of sample data in a database is an added advantage, allowing for future reruns and experiments.

A critical feature of the ARIA system is its ability to record each liquid dispensing step with an extensive audit trail. The volume of each dispensed liquid is verified, logged and associated with the unique barcode of each plate at a well level. This data is represented in a visually accessible format, such as heat maps or table formats (CSV files), providing a detailed account of the dispensing process (Figure 7a). Furthermore, every process in the workflow is diligently tracked to ensure comprehensive process documentation.


https://doi.org/10.1021/acsomega.4c02003
ACS Omega 2024, 9, 24948−24958

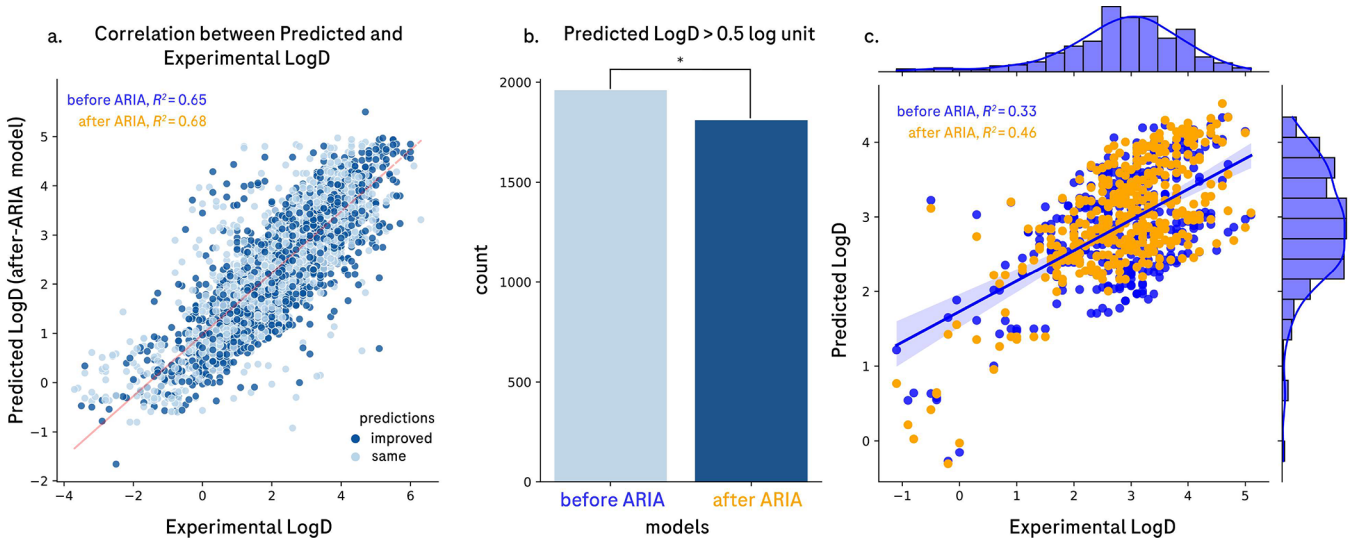



**Figure 8.** Evaluation of ML models generated from data produced with (after-ARIA model) and without ARIA (before-ARIA model). The before-ARIA model was trained on 148073 measured Log$D_{7.4}$, and the after-ARIA model had 7255 more data points. (a) Correlation between predicted Log$D_{7.4}$ from the after-ARIA model and experimental Log$D_{7.4}$ ($n$ = 4139 molecules across multiple projects). Data points representing an improved Log$D_{7.4}$ prediction (dark blue) and no improvement relative to the experimental Log$D_{7.4}$ (light blue). (b) Number of predicted Log$D_{7.4}$ values greater than 0.5 log unit in regards to the corresponding experimental Log$D_{7.4}$. With the before-ARIA model containing 1960 predictions above 0.5 log units and 1810 for the after-ARIA model, the $p$-value was 0.174. (c) For some projects, the after-ARIA model displayed a significant increase in performance, from $R^2$ = 0.33 for the before-ARIA model (blue) to $R^2$ = 0.46 for the after-ARIA model (orange).

The system's scheduling software, Revolution (by UK Robotics), orchestrates operations outside the Tecan environment. It plays a crucial role in identifying and responding to errors such as idle or stuck plates, missing plates, or insufficient plate availability and critically, it robustly handles device errors. Upon detection of such errors, Revolution pauses the workflow and alerts the user, allowing for timely intervention and workflow restart. This feature is essential for preserving the integrity of the run, avoiding the need to start over.

Additionally, the ARIA system is equipped with CCTV cameras throughout its setup (Figure 7b). These cameras offer an extra layer of monitoring, enabling users to visually inspect the process and identify potential errors. An illustrative example of this utility is the resolution of incidents like plate collisions due to human errors or the resolution of device errors remotely by ensuring it is in a safe, recoverable state. The camera recordings can be reviewed to understand why the system halted and to verify whether a plate was incorrectly positioned, providing valuable feedback to both the software and the user. This helps facilitate the development of robust standard operating procedures (SOP) that further promotes system consistency of operation.

The ARIA system's comprehensive setup, combining a sophisticated web application, BLL, detailed volume checking, efficient scheduling software, and thorough security monitoring, offers a robust platform for troubleshooting. This system not only enhances data transparency but also ensures effective monitoring and resolution of hardware-related issues, making it a highly reliable tool in assay methodology.

**ML Models Built from Data Generated with ARIA.** To evaluate the effect ARIA has in facilitating ML, a direct comparison was made between ML models built with data generated from ARIA. Building two comparable models based on a similar sized training data was not possible because ARIA has only been implemented for less than a year (4.7%, 7255 molecules measured with ARIA out of 155328 total molecules). Furthermore, the new data generated from ARIA will always be new molecules and covers a chemical space quite different to the training set for the model before ARIA. With these constraints in mind, we generated 2 ML models with before-ARIA and after-ARIA data to predict Log$D_{7.4}$ (Figure 8a). Using the same test set (4139 molecules) for both models, we observed a $R^2$ = 0.65 for the before-ARIA model and an increase in performance for the after-ARIA model, $R^2$ = 0.68. It is not surprising we do not observe a significant increase given the small amount of data provided by the ARIA system. Additionally, the increase in performance we observed could be a result of other factors like the addition of more chemically similar compounds. Nonetheless, there is an improvement in the after-ARIA model and we observed fewer Log$D_{7.4}$ predicted greater than 0.5 log units compared to the experimental Log$D_{7.4}$ (Figure 8b). Finally, we analyzed the performance of the after-ARIA model at the project level to narrow the scope of chemical space. For some projects, we observed a significant increase in predicting Log$D_{7.4}$ with the after-ARIA model (Figure 8c). These promising results was a clear indicator for ARIA as a platform to generate higher quality data for ML purposes.

**Future Prospects of Automation in Drug Discovery and ML.** Reflecting on our journey with ARIA, we ponder on what could be altered or implemented if we were to start anew. If we were to reimagine ARIA from the ground up, then identifying the core components that are indispensable and those that could be streamlined or omitted is crucial. This process is not just theoretical but essential for evolving ARIA into a more efficient, user-friendly system, better suited to the rapid technological advancements in the field.

A key immediate implementation would be the integration of a LCMS (liquid chromatography mass spectrometry) directly into the ARIA system. This addition would effectively close the loop from assay execution to data analysis, creating a seamless flow of information and significantly reducing the


https://doi.org/10.1021/acsomega.4c02003
ACS Omega 2024, 9, 24948−24958

time between sample processing and result interpretation. Another significant enhancement would be the implementation of computer vision technology within the ARIA system. This advancement would render the system "smarter" and more autonomous. There are numerous scenarios where routine errors, such as misaligned plates, insufficient tips on the Tecan deck, or inaccuracies in specified locations, necessitate human intervention. By incorporating computer vision, many of these issues could be automatically detected and resolved by the system itself, further reducing the need for manual oversight and improving overall efficiency.

## CONCLUSION

The implementation and validation of the ARIA system marks a significant stride in the field of drug discovery. Our comprehensive analysis throughout this paper underscores the efficacy of ARIA in enhancing the accuracy, reproducibility, and throughput of physicochemical assays (6 to 10-fold increase in number of samples assayed per month). By successfully addressing the inherent limitations of the semi-automated sample preparation, ARIA has set a new standard in the measurement of key drug properties like Log$D_{7.4}$ and kinetic solubility. Looking forward, the integration of automated systems like ARIA in drug discovery will be important to advancing the field. The high-quality data generated by ARIA can be used to update *in silico* ML/AI models and improve overall predictions. Even though the comparison between the performances of ML models trained on data prepared with and without ARIA was not as significant as we hoped, the results were still promising. These models are crucial in the early stages of drug development, and any improvement in the data quality and the assay integrity would assist in designing new drug candidates and optimizing lead series. The consistent and reliable data from ARIA provides a robust foundation for training and refining these predictive models. Furthermore, the automation and precision of ARIA opens new avenues in high-throughput screening in our early discovery laboratory, to not only measure physicochemical properties but also develop assays to interrogate downstream questions related to chemical stability and activity.

## EXPERIMENTAL SECTION

**Instrumentation.** An overview of all the instruments installed on the ARIA system in depicted in Figure S2 in the Supporting Information: PreciseFlex 400 robotic arm (Brooks Automation, Fremont, CA), Cytomat 2 incubator (Thermo Scientific, Langenselbold, Germany), Fluent 1080 liquid handler (Tecan, Männedorf, Switzerland), D2 Dispenser (UK Robotics Ltd., Bolton, UK), XPeel (Azenta, Burlington, MA), LabRAM II acoustic shaker (Resodyne, Butte, MT - Automated by UK Robotics), HiG centrifuge (Bionex, San Jose, CA), I.D. Decapper (Hamilton, Reno, NV), PlateLoc (Agilent, Santa Clara, CA), and BioMicroLab VC384 volume checker (SPT Labtech, Concord, CA).

**Reagents and Materials.** Standard compounds for method validation were acquired from Sigma-Aldrich (St. Louis, MO). DMSO was acquired from EMD Chemicals (Philadelphia, PA). MS-grade LC solvents were purchased from OmniSolv (Charlotte, NC). 1-Octanol was purchased from Sigma-Aldrich (St. Louis, MO). Phosphate-buffered solution (PBS) was prepared with concentrations of NaCl (136 mM), KCl (2.6 mM), $Na_2HPO_4$ (7.96 mM), and $KH_2PO_4$ (1.46 mM). The buffer pH was adjusted to 7.4. Buffer solution was prepared with a saturated organic phase (200 mL of 1-octanol/20 mL of PBS pH = 7.4) and saturated aqueous phase (1 mL of 1-octanol/200 mL of PBS pH = 7.4). The 96-well plates comprised the 900 $\mu$L vial Nunc, 96-deep well polystyrene plates, and 300 $\mu$L vial Greiner 96-well polypropylene plates. The 384-well plates were 100 $\mu$L vial Greiner polypropylene plates.

*Kinetic Solubility and DMSO Stock Concentration Determination Instrument Setup.* The system consist of a UPLC (Agilent 1290 series, Agilent, Santa Clara, CA) with two binary pumps (precolumn A/B, postcolumn C/D), an autosampler, a column oven, a diode array detector (DAD, G7115A, Agilent), a single quadrupole mass spectrometer (G6160A, LC/MSD iQ, Agilent), and charged aerosol detector (CAD, Corona Veo RS, Thermo Scientific). The stationary phase was a reverse phase Poroshell 120 EC-C18 column (InfinityLab, 2.1 × 50 mm$^2$, 1.9 $\mu$m, Agilent).

*Log$D_{7.4}$ Instrument Setup.* Liquid chromatography measurements were taken with LC-MS/MS systems. The first system involved a Shimadzu Nexera X2 HPLC with two LC-30AD pumps, SIL-30ACMP autosampler, SPD-20AV diode array detector (Columbia, Maryland), 6 port Valco valve (Houston, TX), and Sciex QTRAP4000 mass spectrometer (Foster City, CA). The latter system consisted of a Shimadzu Nexera X2 HPLC with two LC-30AD pumps, SIL-30ACMP autosampler, SPD-20AV diode array detector, and Sciex API4000 mass spectrometer. Both systems utilized a Waters Xbridge C18 (2.1 × 30 mm$^2$, 2.5 $\mu$m, 130 Å) column under gradient method maintained at 40 °C.

*Semiautomated Sample Preparation.* Solvent and liquid sample handling was automated with a Tecan Freedom Evo robot (Männedorf, Switzerland) configured with two liquid handling arms using 8-channel disposable tips and 8-channel fixed tips. Samples were mixed with a Resodyn LabRAM acoustic mixer (Butte, MT).

*Log$D_{7.4}$ Assay.* The Log$D_{7.4}$ assay has been reported elsewhere, and we will briefly go over the protocol along with some modifications to increase throughput.[8] From 10 mM DMSO stock solutions, 7 $\mu$L was dispensed into 96-deep-well plates containing saturated 1-octanol (350 $\mu$L) using a Tecan Freedom Evo robot. Plates were sealed with aluminum film and shaken for 5 min at 25G of acceleration (Labram acoustic shaker). After shaking, the plates were centrifuged at 3700 rpm for 10 min. Plates were then placed into the Tecan liquid handler to add the remaining 350 $\mu$L of saturated PBS buffer solution (pH = 7.4). The plates were sealed, shaken, and centrifuged one final time (25G, 5 min). Liquid separation was achieved with the Tecan robot by first transferring 250 $\mu$L of 1-octanol phase into a 96-well plate. Prior to aspirating the saturated PBS phase of the mixture, the tips were first filled with 10 $\mu$L of PBS to act as a plug to prevent contamination with 1-octanol.[36] The Tecan transferred 250 $\mu$L of the PBS phase to a corresponding 96-well plate. Both separated compound phases were then sorted and prepared into mixtures of 8 compounds by mixing 20 $\mu$L of each in the well of a 96-well plate. Mixtures were sorted based upon monoisotopic mass such that each compound had a minimum 2 Da mass difference within the same group to prevent overlapping signal response during mass spectrometry quantification. The sorting was handled by a bespoke algorithm within the BLL, while also generating work lists operated by the Tecan liquid handler.


https://doi.org/10.1021/acsomega.4c02003
ACS Omega 2024, 9, 24948−24958

Samples were then analyzed by LCMS and Log$D_{7.4}$ was calculated using eq 1.

$$\log D = \log_{10}\left(\frac{\text{octanol}}{\text{PBS pH} = 7.4}\right) \quad (1)$$

*Kinetic Solubility Assay.* This assay has been reported elsewhere; briefly, assay plates were prepared from compounds as DMSO stock solutions (10 mM).[9] Samples (4 μL, 10 mM) were transferred into a 96-well microplate containing PBS buffer (pH = 7.4, 196 μL, affording a final concentration of 200 μM for each sample). To prevent evaporation, samples were sealed with a heat activated aluminum sheet. The sample plates were shaken on a vibrating platform shaker for 24 h at 1000 rpm and kept at room temperature. After 24 h, sample plates were removed from the shaker, centrifuged, and transferred (200 μL) to 96-well filter plates (Millipore Multiscreen HTS). A positive pressure manifold (Waters, Positive Pressure-96 Processor) was then utilized to filter precipitate and collect the filtrate into separate 96-well microplates. After filtration, 100 μL of filtrate was diluted into 100 μL of DMSO for a final theoretical maximum compound concentration of 100 μM. Plates were then sealed and shaken before transferring to a LCMS-UV-CAD for analysis.

To quantify the sample concentrations, a LCMS system equipped with a CAD was utilized (LCMS-UV-CAD). The injection volume was 5.00 μL for all sample and flow rate was 0.400 mL/min. Mobile phases A/C were water with 0.01% formic acid, and mobile phases B/D were methanol with 0.01% formic acid. The first pump (precolumn, A/B) started at 98% A to 98% B over 2.10 min and back to 98% A at 3 min. The second binary pump (postcolumn C/D) had a corresponding counter gradient complementary to the first binary pump, starting with 98% D to 98% C over 2.10 min and back to 98% D at 3 min. By introducing a counter gradient with the second binary pump postcolumn, a 50% methanol solvent mixture is maintained for the CAD analysis and helps reduce baseline noise.[37] A concentration calibration curve for the CAD was utilized to calculate the concentrations of unknown samples.

*Purity and DMSO Stock Concentration.* Assay plates were prepared by transferring sample from 10 mM in DMSO stock solutions (1 μL) into 384-well microplates containing DMSO (99 μL). The assay plates were then mixed on a shaker for 25 min at 1000 rpm and then transferred to a LCMS-UV-CAD for analysis. Sample purity was measured using the UV absorption data. To measure sample concentration in DMSO, a concentration calibration curve for the CAD generated from a set of standards was utilized to calculate the concentrations of unknown samples.

*LogD ML Models.* Both models used MoKa descriptors (from Molecular Discovery, www.moldiscovery.com/software/moka/), RDkit (www.rdkit.org/) molecular descriptors, and Morgan fingerprint descriptors (also available in RDkit) as features and were trained using the Extremely Randomized Tree (XRT) algorithm (ExtraTreesRegressor in sklearn).[38] The models were established in Python 3.9.14.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c02003.

List of instruments installed on the ARIA system, scheme of ARIA's software architecture, and photographs of the ARIA system(PDF)

Video of the ARIA system in action(MP4)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Huy Q. Nguyen** − *Analytical Research, Discovery Chemistry Department, Genentech, Inc., South San Francisco, California 94080, United States;* orcid.org/0000-0002-5252-0062; Phone: +1 (650) 467-6559; Email: nguyen.huy_quoc@gene.com

### Authors

**Newton P. Wu** − *Analytical Research, Discovery Chemistry Department, Genentech, Inc., South San Francisco, California 94080, United States;* orcid.org/0009-0009-6261-7106

**Wenyi Wang** − *Drug Metabolism and Pharmacokinetics, Genentech, Inc., South San Francisco, California 94080, United States;* orcid.org/0000-0001-5641-2205

**Dhireshan Gadiagellan** − *UK Robotics, Inc., Manchester BL5 3EH, United Kingdom;* orcid.org/0009-0003-2939-1914

**Mike Counsell** − *UK Robotics, Inc., Manchester BL5 3EH, United Kingdom*

**Nikkia K. Hamidi** − *Analytical Research, Discovery Chemistry Department, Genentech, Inc., South San Francisco, California 94080, United States*

**Yuko Koike** − *Analytical Research, Discovery Chemistry Department, Genentech, Inc., South San Francisco, California 94080, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c02003


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

All the figures in this manuscript were created using BioRender.com. We thank the compound management team at Genentech for providing us with samples and assistance.

## ■ REFERENCES


(1) Zhou, L.; Yang, L.; Tilton, S.; Wang, J. Development of a high throughput equilibrium solubility assay using miniaturized shake-flask method in early drug discovery. *J. Pharm. Sci.* **2007**, *96*, 3052.

(2) Bharate, S. S.; Vishwakarma, R. A. Thermodynamic equilibrium solubility measurements in simulated fluids by 96-well plate method in early drug discovery. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 1561.

(3) Könczöl, A.; Dargó, G. Brief overview of solubility methods: Recent trends in equilibrium solubility measurement and predictive models. *Drug Discovery Today: Technol.* **2018**, *27*, 3.

(4) Sou, T.; Bergström, C. A. Automated assays for thermodynamic (equilibrium) solubility determination. *Drug Discovery Today: Technol.* **2018**, *27*, 11.

(5) Wenlock, M. C.; Potter, T.; Barton, P.; Austin, R. P. A Method for Measuring the Lipophilicity of Compounds in Mixtures of 10. *SLAS Discovery* **2011**, *16*, 348−355.

(6) Kestranek, A.; Chervenak, A.; Longenberger, J.; Placko, S. Chemiluminescent Nitrogen Detection (CLND) to Measure Kinetic Aqueous Solubility. *Curr. Protoc. Chem. Biol.* **2013**, *5*, 269.

(7) Lin, B.; Pease, J. H. A high throughput solubility assay for drug discovery using microscale shake-flask and rapid UHPLC-UV-CLND quantification. *J. Pharm. Biomed. Anal.* **2016**, *122*, 126.

(8) Ma, Y.; Chen, X.; Javeria, H.; Du, Z. High-throughput screening of Log*D* by using a sample pooling approach based on the traditional shake flask method. *J. Chromatogr B Analyt Technol. Biomed Life Sci.* **2023**, *1227*, 123804.
(9) Lin, B.; Pease, J. A novel method for high throughput lipophilicity determination by microscale shake flask and liquid chromatography tandem mass spectrometry. *Comb Chem. High Throughput Screen* **2013**, *16*, 817−25.
(10) Komura, H.; Watanabe, R.; Mizuguchi, K. The Trends and Future Prospective of In Silico Models from the Viewpoint of ADME Evaluation in Drug Discovery. *Pharmaceutics* **2023**, *15*, 2619.
(11) Sadybekov, A. V.; Katritch, V. Computational approaches streamlining drug discovery. *Nature* **2023**, *616*, 673.
(12) Lombardo, F.; Desai, P. V.; Arimoto, R.; Desino, K. E.; Fischer, H.; Keefer, C. E.; Petersson, C.; Winiwarter, S.; Broccatelli, F. Silico Absorption, Distribution, Metabolism, Excretion, and Pharmacokinetics (ADME-PK): Utility and Best Practices. An Industry Perspective from the International Consortium for Innovation through Quality in Pharmaceutical Development: Miniperspective. *J. Med. Chem.* **2017**, *60*, 9097.
(13) Waring, M. J. Lipophilicity in drug discovery. *Expert Opin. Drug Discovery* **2010**, *5*, 235.
(14) Arnott, J. A.; Planey, S. L. The influence of lipophilicity in drug discovery and design. *Expert Opin. Drug Discovery* **2012**, *7*, 863.
(15) Wenlock, M. C.; Carlsson, L. A. How Experimental Errors Influence Drug Metabolism and Pharmacokinetic QSAR/QSPR Models. *J. Chem. Inf. Model.* **2015**, *55*, 125.
(16) Isenhour, T. L.; Eckert, S. E.; Marshall, J. C. Intelligent Robots-The Next Step in Laboratory Automation. *Anal. Chem.* **1989**, *61*, 805A.
(17) Kerr, R. A. Sample Preparation by a One-Armed Robot. *Science* **1982**, *216*, 164.
(18) Dessy, R. Robots in the Laboratory: Part II. *Anal. Chem.* **1983**, *55*, 1232A.
(19) Olsen, K. The First 110 Years of Laboratory Automation: Technologies, Applications, and the Creative Scientist. *SLAS Technol.* **2012**, *17*, 469.
(20) Prabhu, G. R. D.; Urban, P. L. The dawn of unmanned analytical laboratories. *TrAC, Trends Anal. Chem.* **2017**, *88*, 41.
(21) Vescovi, R.; Ginsburg, T.; Hippe, K.; Ozgulbas, D.; Stone, C.; Stroka, A.; Butler, R.; Blaiszik, B.; Brettin, T.; Chard, K. Towards a modular architecture for science factories. *Digital Discovery* **2023**, *2*, 1980.
(22) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A mobile robotic chemist. *Nature* **2020**, *583*, 237−241.
(23) Salley, D.; Manzano, J. S.; Kitson, P. J.; Cronin, L. Robotic Modules for the Programmable Chemputation of Molecules and Materials. *ACS Cent. Sci.* **2023**, *9*, 1525.
(24) Ha, T.; Lee, D.; Kwon, Y.; Park, M. S.; Lee, S.; Jang, J.; Choi, B.; Jeon, H.; Kim, J.; Choi, H. AI-driven robotic chemist for autonomous synthesis of organic molecules. *Sci. Adv.* **2023**, *9*, eadj0461.
(25) Manzano, J. S.; Hou, W.; Zalesskiy, S. S.; Frei, P.; Wang, H.; Kitson, P. J.; Cronin, L. An autonomous portable platform for universal chemical synthesis. *Nat. Chem.* **2022**, *14*, 1311.
(26) Shiri, P.; Lai, V.; Zepel, T.; Griffin, D.; Reifman, J.; Clark, S.; Grunert, S.; Yunker, L. P.; Steiner, S.; Situ, H. Automated solubility screening platform using computer vision. *iScience* **2021**, *24*, 102176.
(27) Wu, J.; et al. Integrated System Built for Small-Molecule Semiconductors via High-Throughput Approaches. *J. Am. Chem. Soc.* **2023**, *145*, 16517−16525.
(28) Schneider, G. Automating drug discovery. *Nat. Rev. Drug Discovery* **2018**, *17*, 97.
(29) Mahjour, B.; Zhang, R.; Shen, Y.; McGrath, A.; Zhao, R.; Mohamed, O. G.; Lin, Y.; Zhang, Z.; Douthwaite, J. L.; Tripathi, A. Rapid planning and analysis of high-throughput experiment arrays for reaction discovery. *Nat. Commun.* **2023**, *14*, 3924.
(30) Vargas Medina, D. A.; Maciel, E. V. S.; Lanças, F. M. Modern automated sample preparation for the determination of organic compounds: A review on robotic and on-flow systems. *TrAC, Trends Anal. Chem.* **2023**, *166*, 117171.
(31) Wilbraham, L.; Mehr, S. H. M.; Cronin, L. Digitizing Chemistry Using the Chemical Processing Unit: From Synthesis to Discovery. *Acc. Chem. Res.* **2021**, *54*, 253.
(32) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343.
(33) Wenlock, M.; Austin, R.; Potter, T.; Barton, P. A highly automated assay for determining the aqueous equilibrium solubility of drug discovery compounds. *J. Lab Autom* **2011**, *16*, 276−84.
(34) Bland, J. M.; Altman, D. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *Lancet* **1986**, *327*, 307−310.
(35) Aliagas, I.; Gobbi, A.; Lee, M.-L.; Sellers, B. D. Comparison of logP and logD correction models trained with public and proprietary data sets. *Journal of Computer-Aided Molecular Design* **2022**, *36*, 253−262.
(36) Yamashita, T.; Nishimura, I.; Nakamura, T.; Fukami, T. A System for Log*D* Screening of New Drug Candidates Using a Water-Plug Injection Method and Automated Liquid Handler. *Journal of the Association for Laboratory Automation* **2009**, *14*, 76−81.
(37) Almeling, S.; Ilko, D.; Holzgrabe, U. Charged aerosol detection in pharmaceutical analysis. *J. Pharm. Biomed. Anal.* **2012**, *69*, 50−63. (Review Papers on Pharmaceutical and Biomedical Analysis 2012).
(38) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Machine Learning* **2006**, *63*, 3−42.


https://doi.org/10.1021/acsomega.4c02003
*ACS Omega* 2024, 9, 24948−24958