An Artificial Intelligence Approach to Proactively Inspire Drug Discovery with Recommendations

Steven L. Rohall, Lydia Auch, Jonathan Gable, Jacob Gora, Johanna Jansen, Yipin Lu, Eric Martin, Margaret Pancost-Heidebrecht, Bill Shirley, Nikolaus Stiefel, and Mika Lindvall*

ABSTRACT: Artificial intelligence (AI) is becoming established in drug discovery. For example, many in the industry are applying machine learning approaches to target discovery or to optimize compound synthesis. While our organization is certainly applying these sorts of approaches, we propose an additional approach: using AI to augment human intelligence. We have been working on a series of recommendation systems that take advantage of our existing laboratory processes, both wet and computational, in order to provide inspiration to our chemists, suggest next steps in their work, and automate existing workflows. We will describe five such systems in various stages of deployment within the Novartis Institutes for BioMedical Research. While each of these systems addresses different stages of the discovery pipeline, all of them share three common features: a trigger that initiates the recommendation, an analysis that leverages our existing systems with AI, and the delivery of a recommendation. The goal of all of these systems is to inspire and accelerate the drug discovery process.

INTRODUCTION

Many pharmaceutical companies, including Novartis, are applying artificial intelligence, machine learning, and other predictive technologies to challenges in drug discovery. Examples include the application of machine learning to the analysis of disease in tissue samples, the identification of promising drug targets, and the determination of which compounds to synthesize next.1,2 The goal of these projects is to provide insights that people may not have found on their own.

Our organization has taken an additional approach that leverages the value of artificial intelligence in order to enable scientists to better prioritize and perform their research. By providing scientists with the right information at the right time via recommendations, we endeavor to use artificial intelligence as a means to amplify human intelligence. For example, awareness of how others have addressed a troublesome synthesis and the availability of commercial compounds with similar properties are daily issues in the medicinal chemistry community. The systems exist for our chemists to search our electronic lab notebook for similar work, or vendor databases for related compounds. It has been our experience, however, that the chemists will not avail themselves of such systems unless they have hit a roadblock in their work—and sometimes not even then—relying instead on informal social networks to solicit advice. By “reversing” the typical workflow, our systems conduct searches and perform analyses on behalf of the chemists without the need for them to perform any additional work.

The challenge is to deliver the right information to the right person at the right time. This is where artificial intelligence is employed. Standard activities (e.g., lab notebook entries, 3D protein co-crystal depositions) trigger our systems. The systems then perform analyses, and the information is delivered to the scientist. Our goal is not to replace targeted search activities, but rather to augment the scientists’ work and provide additional information and inspiration. Our recommendations are succinct; we do not attempt to be exhaustive in what we suggest. Rather, we highlight a few relevant possibilities that will cause the scientists to stop and think about ways to improve their work. Our goal is to provide the right timing and the right targeting for our recommendations. However, if a recommendation is off the mark (e.g., not relevant, or sent to the wrong person), the scientist can quickly spot that in our succinct messages and move on.

Special Issue: Artificial Intelligence in Drug Discovery

Received: December 20, 2019
Published: February 26, 2020

https://dx.doi.org/10.1021/acs.jmedchem.9b02130
PRIOR WORK

We are not the first to consider using AI-based recommendations for medicinal chemistry, although there has been comparatively little work done on enterprise recommender systems in general, and particularly those in the chemistry space. Boström et al. have applied the technology of recommender systems to the problem of chemical reagent selection. They built a system which employs item-to-item collaborative filtering (i.e., “people who bought this also bought that”) to suggest reagents similar to one being searched for. Their goal is to increase the “odds for serendipity” happening in drug discovery. Their system showed that by suggesting reagents that other chemists had used they received a more diverse set of results than if they had used more traditional similarity-based suggestions.

Corkery proposed, but did not implement, applying standard collaborative filtering techniques to the task of ordering chemical compounds from a supplier, the idea being that additional molecules would be suggested based upon the contents of a chemist’s shopping cart. This is no different than existing shopping recommenders but may actually be more difficult to implement since, without the knowledge of what the chemist is planning to do with a given compound, it would be difficult to recommend other molecules out of the hundreds or thousands that could be recommended that would actually be useful to the chemist. This is in contrast to Boström’s similar approach for recommending reagents.

Lilly has implemented an interesting system called “Kernel.” Kernel is a virtual assistant for medicinal chemists; its goal is to deliver key information (a biochemical assay summary email) to a scientist as quickly as possible, usually within an hour of the data being generated. The email to the original submitter of the compound includes the assay results as well as some basic automated analysis. The email also includes information for other compounds, such as what was synthesized recently, which compounds were assayed, which were most active, etc.—a bit of a newsletter about what is happening in the lab.

Savage et al. developed a system for recommending candidate reactants for the synthesis of a given product. They approached the problem as a link-prediction problem over a network of chemical reactions (reactants linking to products) described in the patent literature. Their system is intended to aid chemists in designing an experiment to synthesize a proposed molecule, but it has not yet been deployed. Rather, they have concentrated on developing and comparing algorithms that take advantage of their chemical network in order to improve on a simple similarity-based approach. Their use case is to help a chemist who is actively searching for a product starting point. Although they are using recommender system algorithms in a very interesting way, they are not yet providing recommendations within the context of a chemist’s work.

Hall et al. describe using a graph database to efficiently search for close analogues of a given molecule. While their data structure could be used for experiment recommendations, their main concern was improving searches for an individual chemist who would like to get a better sense of the chemical space around a particular molecule or fragment.

Our work has similar motivations to these projects, but it has a different purpose: to augment the scientists’ normal work routines by making relevant suggestions and recommendations without the scientists needing to do any additional work or spend time working with a new system.

RECOMMENDERS AT THE NOVARTIS INSTITUTES FOR BIOMEDICAL RESEARCH

At the Novartis Institutes for BioMedical Research we have been working on five different chemistry recommenders, each of which is in various stages of development and deployment. They are summarized in Table 1. Since recommenders in medicinal chemistry are relatively new, our approach has been to focus on high-value use cases such as those that provide immediate

<table>
<thead>
<tr>
<th>System</th>
<th>Trigger</th>
<th>Analysis</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chem Recommender</td>
<td>Lab notebook entry</td>
<td>Similarity search based on chemical reaction scheme</td>
<td>Information about one similar reaction, prioritizing older experiments and reactions with higher yields</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Automatic exclusion of purification reactions and reactions that have already been recommended to the recipient</td>
<td></td>
</tr>
<tr>
<td>NIBR Hops</td>
<td>Deposition of a new in-house macromolecular crystal structure</td>
<td>3D similarity search from bound ligand; Top 3D hits that have physical sample and come from both in-house and external sources</td>
<td>Up to eight molecules containing novel 2D scaffolds that are new to originating program and different from each other</td>
</tr>
<tr>
<td>Assay Sommeller</td>
<td>Scientist enters molecule in ideation environment</td>
<td>Cross-target analysis based on z-scaled potency measurements or predictions driven by 2D nearest neighbor search of query structure</td>
<td>List of most relevant assays for query structure based on combined activity profile of nearest neighbors</td>
</tr>
<tr>
<td>Transformer</td>
<td>User-tagged compound optimization challenge</td>
<td>Matched Molecular Pair analysis to find transformations likely to address the challenge; Selection of transformations that have not yet been tried in the originating program</td>
<td>Up to n transformations prioritizing those transformations with higher confidence and limited/no negative impact on other properties</td>
</tr>
<tr>
<td>Every Well Counts</td>
<td>Partially filled plate request for dose-response</td>
<td>Various computational models to identify most informative compounds; Automatic exclusion of compounds already tested</td>
<td>Compounds to completely fill the plate with prioritizing compounds meeting one of several available profiles: find potent hits, explore expanded chemical space, expand model domain of applicability</td>
</tr>
</tbody>
</table>

Table 1. A Summary of Recommenders at NIBR
benefit to our scientists and are well-supported by our existing systems. This enables us to deploy our recommendation engines, assess their impact, and get feedback.

“Chem Recommender” is our oldest recommender in production, operational since July 2017.8 It mines our chemistry electronic lab notebook and suggests related chemical syntheses based upon the work a chemist has recently completed. It also provides literature recommendations for compounds that are proposed for synthesis.

“NIBR Hops” was the first recommender, piloted across the company in January 2017 and launched in production in March 2019. Upon deposition of a new protein co-crystal structure with a bound project compound, NIBR Hops recommends compounds that have 3D similarity to the bioactive conformation of the original ligand by searching both internal and external small molecule databases. Furthermore, the proposed compounds have 2D scaffolds that are new to the project and therefore represent potentially high-value scaffold changes.

A proof-of-concept for an assay recommendation engine called “Assay Sommelier” was developed in early 2019. Based on information on portfolio project compounds, the system recommends the most relevant assays that are not part of the project flowchart. The proposed assays are followed up with either to identify and screen for potential off-targets early on or to identify relevant new chemical starting points.

Two other recommenders are being developed. The first, “Transformer”, will send recommendations to medicinal chemists who are addressing compound optimization challenges (e.g., reducing microsomal clearance). The second, “Every Well Counts”, seeks to augment the productivity of bioassay runs by leveraging empty wells in routine assay plates in order to generate additional meaningful data for both organization-wide assays and project-specific assays.

Data: The Common Thread. As can be seen in the short descriptions above, our recommenders inherently draw on historical data. They benefit from a variety of internally developed systems that take advantage of Novartis’ long history, wide range of projects, and historical experimental data. The chemistry electronic lab notebook is a treasure trove of data regarding historical synthetic work, as well as works in progress, and is used within Chem Recommender. NIBR Hops, in turn, relies on an in-house crystallography database and searches through millions of compounds in the Novartis small molecule database.

Profile-QSAR, which is used by Assay Sommelier and Every Well Counts, is a massive multitask machine learning method that has produced experiment-quality activity predictions for over 8000 historical assays.9 It is trained on 18 million IC50s from 1.8 million compounds tested on 12,000 assays. In addition to on-the-fly predictions, the models are updated every month and 50 billion IC50s are recomputed for 5.5 million Novartis compounds on the over 8000 models. These predictions are stored in a special indexed file system, which allows for the extraction of all active compounds for a series of assays or of all assays hit by a series of compounds in seconds. This allows for very fast recommendations for on-target and off-target activities. The Matched Molecular Pair Data Base (mmpdb), used by Transformer, enables the automated and systematic compilation of medicinal chemistry rules from compound/property data sets. At Novartis, we have used mmpdb to build a database of rules and their impact on more than 10 properties, such as hERG binding, which are commonly monitored in medicinal chemistry optimization. The Compound Series and Favorites (CSF)10 program, also employed by Transformer, provides scientists with a shared space to store scaffold definitions and most significant compounds (Favorites). Compounds are annotated with tags, such as why the compound was made and rgg problems that need to be solved.

All of our recommendation systems take these data, which are derived and stored for other purposes, in order to combine it all in interesting and new ways and then generate their recommendations. Details are provided below.

Chem Recommender. In a company the size of Novartis, it is quite impossible to tell if scientists at different sites, or even at the same site but different benches, have done similar work in the past. For example, scientists with a short tenure with the company may not have developed the contacts that would inform them of related work. Chem Recommender was developed in order to alert chemists to the related synthetic work of their colleagues, allowing them to accomplish their goals more quickly and saving both time and money. Chem Recommender analyzes the work that chemists have recently completed in their electronic lab notebook (ELN) and then suggests prior, related work to them.

Although our ideal goal is to insert recommendations into our ELN as scientists are conducting their work, in the current implementation the recommendations are sent via email. Since July 2017 we have sent nearly 25,000 recommendations to over 1000 unique chemists. At this time, only the top hit is suggested to the chemist—the goal being to produce a recommendation that can be consumed quickly, without scrolling or clicking. The similarity search also includes a time component that favors older experiments—when faced with a choice of several recommendation candidates, we choose the oldest experiment on the assumption that a chemist is less likely to already know about older work.

Figure 1 contains a sample recommendation email. Each recommendation includes an image of the experiment being recommended, links to the other scientist and the recommended lab notebook entry, and a link to the lab notebook entry that was the inspiration for the recommendation. Other buttons allow a one-click rating as well as allow the scientist to increase or decrease the frequency of the recommendations. In order to avoid overwhelming the chemists, they receive a recommendation no more frequently than once a week. In addition, we only send an email when we have something to recommend; being a noninteractive system, there is no pressure to generate a

---

**Did you know?**

**Maggie Pancost** conducted work similar to yours in Mar 2009.

How useful was this update? You can also reply to this email to provide more detailed comments.

1 | 2 | 3 | 4 | 5

Would you like to receive these updates more or less frequently (never more than once a week)?

4 --

Figure 1. An example recommendation email (the reaction shown is from the synthesis of penicillin11).
recommendation for the user if we have nothing useful to suggest.

Evaluation of enterprise recommendation systems is difficult; we are not looking for click conversions or increased sales such as in commercial recommendation systems. In fact, our recommendation emails are intentionally designed to provide as much information as possible without requiring any other action on the part of the recipient. It is possible that a chemist will be inspired just by seeing the other chemist’s name, or an image of the reaction being recommended; the lack of explicit user action on the email does not indicate that the recommendation was not useful. As a result, we have taken a multipronged approach to determine how well Chem Recommender is performing. We monitor click-throughs, allow scientists to rate the recommendations, track email frequency changes, and record comments through both online surveys and interviews.

Click-throughs. We have instrumented the links in the recommendations so that we are able to record click-throughs. A typical email marketing campaign will have a click-through rate of less than 4%; ours is around 11%, which is nearly three times the average. Several factors contribute to this increased rate, one of which is that the email is sent from an internal corporate address and is clearly not spam. However, we can also attribute some of that rate to a genuine interest in the content of the recommendations.

Ratings. As seen in Figure 1, we also provide a simple way for users to rate a recommendation on a 1 (low) to 5 (high) scale. Four percent of the recommendations we have sent have been rated, and the distribution of ratings is shown in Figure 2.

![Figure 2. Distribution of recommendation ratings.](image)

We consider this distribution a success, particularly if we add together the 1 and 2 ratings (419) and the 4 and 5 ratings (564). This shows that the positive ratings slightly outnumber the negative ratings. One problem with our simple numeric rating system is that, especially for negative ratings, it connotes a bad recommendation that is not relevant (e.g., the recommended experiment is not similar to the chemist’s experiment) with a good recommendation that is relevant but not useful (e.g., the recommended experiment is similar but the chemist was already aware of the work). For this reason, whether we get a negative (1 or 2) or a positive (4 or 5) rating, we give the chemists the opportunity to fill out a simple survey to further explain their rating. In the survey, we ask if the recommendation was similar to the work they were doing, whether or not the experiment or the product is of interest, and whether the scientist already knew about the recommended work—multiple choices are permitted. The top reason for a low rating, accounting for nearly half of all responses, was that the scientist already knew of the recommended work. The second reason, for approximately 30% of responses, was that the recommendation was not similar to their work. In comparison, for highly rated recommendations 75% of the respondents said that the recommendation was similar to their own work. In other words, recommendations were rated highly if they were, indeed, similar, whereas recommendations were rated lower if the work was already familiar. We hypothesize two reasons for the familiarity: (1) the work is routine chemistry known to all of the chemists, and (2) the scientist actually already knows the other chemist or the experiment that was recommended.

Frequency Changes. We allow users to change the frequency of how often they receive a recommendation email, from a maximum of once per week and at least 7 days from the last recommendation (the default) to a minimum of once every 6 weeks. We do not provide a mechanism to entirely opt out of the recommendation emails. To date, we have had 440 frequency increases as compared to 88 decreases.

As can be seen in Figure 3, the vast majority of users have indicated a recommendation frequency of once per week. On more than one occasion, we have observed chemists providing a low rating for a recommendation and then immediately choosing to increase their email frequency! Clearly, they see the potential in our system even if one recommendation did not provide them with useful information.

Comments. The scientists are always welcome to submit additional comments. The simple survey form described above allows them to type a free-form comment; some comments were also received by email. Some comments validate our work as they are indicative of scientists who were able to make progress on their own projects due to our recommendations:

- “The exact transformation on the exact substrate I wanted. I tried another method that did not work. Thanks!”
- “The system has found multiple interesting hits for me that I never would have found otherwise.”
- “… about a month ago Chem Recommender pointed at 2 patents similar to my current scaffold that we did not know about. It was a great help!”

Perhaps our biggest surprise was the reaction to recommendations for syntheses that had low yields. We imagined at the start of the project that informing scientists about experiments that did not succeed would help them avoid similar problems. In practice, however, negative results received low ratings:

- “The shown reaction did not yield any product.”

![Figure 3. Distribution of recommendation email frequency.](image)
As a result of these comments, we modified our algorithm to prioritize reactions with higher yields. In some cases, the recommendation may have been useful, but the chemist had already solved their problems:

- “It is exactly the reaction I seek, but I already saw it myself by searching in the ELN.”

In other cases, our recommendations were good but recommended people who were already working closely together. In one case, one chemist was performing separations and the other was purifying those separations—of course they were familiar with each other’s work!

- “We did this experiment together on the same project.”

However, we also have evidence that people in the same organization on larger projects are not always familiar with their colleagues’ work, so we do not use organization or project data to limit our recommendations.

**NIBR Hops.** NIBR Hops automatically recommends compounds with chemical scaffolds that are novel to a given drug discovery program. Such new “cores” can be highly valuable to programs if there are existing or anticipated issues with the current chemical matter. New cores are able to move away from liabilities, open up different SAR paths for optimization, and improve intellectual property positions.

NIBR Hops proposals are derived from new internal structures of small molecule protein complexes. On average, one or two new crystal structures are deposited across NIBR per day into an internal database. While these events are less frequent than the new chemical reactions underpinning ChemRecommender, they are of particularly high value. Crystal structures represent both a large R&D investment and a high information content. For scaffold hopping, the bioactive conformations of the small molecules in them powerfully focus the search space.

The NIBR Hops system saves time for scientists and provides 3D design inspiration, even when no computational chemist has been assigned to the program yet or they are not routinely looking for new starting points. The various steps involved in scaffold hopping (e.g., 3D query preparation and similarity searching) are automated, as are the advanced analytics that choose the most relevant results.

**Figure 4.** Sample body of a NIBR Hops message. In a production NIBR Hops message within Novartis, Novartis small molecule compound codes would replace the hyperlinks currently referring to compound names from their associated PDB entries.
NIBR Hops has three primary drivers for choosing which molecules to recommend. The first is a high 3D similarity, which increases the likelihood of retaining biological activity. The second is the novelty of a 2D chemical scaffold relative to a project’s prior body of chemical space, as well as among the proposed molecules themselves. The third is a low barrier to action and follow up on the recommended compounds. This is important given the objective for novelty in NIBR Hops; the more novel the scaffold hop, the more likely that the initial attempt could reduce potency. The cost of potential activity loss is mitigated by proposing molecules with a readily available sample from either internal or commercial sources.

NIBR Hops recommendations are sent as emails to select team members. The computational protocol that leads to the message is triggered overnight after new entries are detected in the internal crystal structure database. The recommendation messages are therefore timed soon after new data-enabling scaffold hops becomes available, and coinciding with a time when teams are more open to considering new ideas.

**NIBR Hops Message Exemplified by Carbonic Anhydrase.** An example of a message sent by NIBR Hops is illustrated in Figure 4. By default, the recipients are the crystallographer who solved the structure, the chemists who made the molecule, the project team leader, and the computational chemist, if one was assisting the effort. The recipients are identified from different internal data systems by cross-referencing with project codes. The automatically generated email contains a hyperlink to the new entry in the internal crystal structure database as well as the internal Novartis registration number for the small molecule bound in it. The potential scaffold hops are the central element of the message and are delivered as a 2D structure grid embedded in the message and as 3D overlays via the “3D similar” hyperlink; both deliveries are discussed further below. Further general details about NIBR Hops are also available through a link in the email that addresses frequently asked questions on an organizational and impact level as well as the high-level technical details of the process.

In Figure 4 we have illustrated an NIBR Hops message using public domain carbonic anhydrase inhibitors. A 2D depiction of the bound small molecule is placed in the middle of the grid and highlighted with a box. Around it, the proposed molecules with different scaffolds are rendered, originating from either Novartis in-house or vendor compound databases. Novartis compound numbers are hyperlinked to a web view of corporate compound details, and the vendor compounds are hyperlinked to an external web view of compound details that is integrated from public domain data sources. The system attempts to produce an even mix of internal and external compounds, with a minimum of five total answers required to produce a grid and send a recommendation. All proposed compounds have a physical sample available, either from the Novartis compound inventory or from a verified vendor inventory maintained at MolPort. The 2D depictions are for the parent form from the data warehouses and are in the tautomeric and ionization form as registered.

The same compounds are also delivered as 3D overlays to the bound conformation of the small molecule in the crystal structure, as illustrated in Figure 5 using the same carbonic anhydrase inhibitors as above. A recipient simply clicks a hyperlink to launch a separate web browser session with the overlays; no additional software installation is needed. Within the session, it is possible to choose which molecules are displayed in the overlay on the right-hand side by clicking the table rows. The overlay can be rotated and zoomed to examine the recommended compounds in detail and compare them with the original project compound. No protein structure is displayed, as the search is ligand based. In the “3D similar” web view, it is possible to interactivly assess the nature of the 3D overlay for each proposed answer and also to examine the ionization and tautomerization form used in the 3D comparison. The “3D similar” view can be a starting point for further design ideas beyond the exact proposed molecules. The recipient can, for example, decide to design a compound with the substituents of their choice or to use only a particular moiety from the proposed molecule as a 3D isostere.

For push messages like those from NIBR Hops, it is important that the recipients can quickly assess if the message is relevant to them. Therefore, the scaffold hop message is designed to be as
simple as possible: the content is short, few molecules are proposed, and details are available through links. Based on feedback, recipients can glance at the 2D structure grid and decide in a few seconds if the chemical content is potentially interesting to them and warrants further investigation in 3D. This further assessment through links also triggers engagement monitoring, as discussed later.

From a technology perspective, the innovation of NIBR Hops is in its integration of established technologies, data from multiple systems, and analytics based on past practices in scaffold hopping. In order to deliver simple, actionable proposals to medicinal chemists. In fact, practically all technological details are hidden in the output, which allows for switching to more sophisticated technologies or different providers with minimal output design changes. In a routine NIBR Hops run, approximately 10 million candidate molecules are considered in order to produce up to eight molecules with potential scaffold hops. In its current form, the NIBR Hops protocol predicts the protomer and tautomer form of the bound ligand in the context of the protein–ligand complex and then removes the protein, using just the bound ligand as a query in 3D searching (Supporting Information). Technologies previously stated to be successful in scaffold hopping have also been employed, such as ROCS. The final recommended molecules lack undesirable substructural flags and have corresponding samples available at Novartis or through a verified vendor inventory.

In an illustrative study with public domain crystal structures, we tested the parameters and criteria used in the NIBR Hops protocol. When we initiated our protocol with a publicly available crystal structure of a small molecule inhibitor bound to carbonic anhydrase and searched in a database derived from 2D structures of other small molecules crystallized in carbonic anhydrase, at least four answers were obtained which did not share the same 2D Murcko scaffold as the query molecule (representing a known body of chemistry instead of project compounds in the production version at Novartis). The answers were included in a NIBR Hops answer grid (Figure 4) among vendor molecule hits. This suggests that configuration and analytics in our automated protocol can pass through known scaffold hops. In its practical application, NIBR Hops is driven by both 3D similarity and novelty—which is hard to quantify and becomes a more dominating factor as the body of chemical space increases in a project. We did not carry out a pure biological-activity enrichment study using 3D similarity criteria alone.

Results. For a recommender to have impact in an R&D organization, scientists must first engage with the recommendation and then take action. The production version of NIBR Hops was designed for measuring engagement, as key details about the proposed molecules are accessible via web links in the message body (Figure 4). This design allows for a great deal of information in a relatively small email body. The web links include 3D alignments, the most popular link, which scientists typically access after glancing at the 2D structure grid and becoming interested in the recommendations. By storing the clicks in a database and associating them with users, messages, teams, and unique recommendations, it has been possible to analyze engagement at different levels.

Engagement analytics were carried out 8 months after the production launch of NIBR Hops in March 2019. In that time frame, over 200 crystal structures with a bound small molecule were deposited into an internal protein structure database at NIBR, averaging 1–2 depositions per workday (Figure 6). Consistent with the multiple criteria involving novelty and quality discussed previously, a third of the depositions resulted in a message being sent to team members. We considered at least one click of any hyperlink in a message as an indication of engagement. At the level of an individual email, approximately 14% of recipients clicked the links in the messages. However, since each recommendation was sent to multiple recipients who sometimes forwarded the emails to additional recipients, we also evaluate the engagement at the recommendation level across all recipients. The results suggested that around 49% of unique recommendations received measurable interaction from at least one scientist. Finally, we estimate that 77% of our unique drug discovery programs engaged at least once with at least one of the several recommendations they received. These are high engagement levels, potentially reflecting the high value of crystallographic data and a significant investment in it.

We received less engagement from ratings and survey responses, with about a 3% rating response rate (12 out of 385 emails). Interestingly, the rating response rate and click-through rate at NIBR Hops are very similar to those observed with Chem Recommender. An early pattern of “love it or hate it” emerged, where a few recipients rated the recommendations either with one or two stars or with four or five stars. Comments included “quality of suggestions is good”, “suggestions are novel”, “quality of the suggestions is not good”, and “I am not working on this project”. From the NIBR Hops pilot, comments given in person included “I now look forward to uploading my data because I get something back” and “I can look at the...
structure grid and make a decision in seconds10 if the structures are worth a further assessment. At this point, the small volume of ratings and comments makes it hard to draw conclusions from them. Consequently, we have decided to follow up with the scientists in person with specific questions separating quality, novelty, and relevance.

The relatively high percentage of recommendations clicked suggests a good opportunity for impact. Importantly, this method of measuring engagement did not require any extra action from scientists. Our impact analytics will follow the same principle. We plan to use NIBR’s data systems to analyze which compounds and associated scaffolds were assayed after the recommendations were sent, as well as whether the compound was from the internal compound archive, purchased, or synthesized. While it may not capture the cases where 3D overlays have acted as inspiration, this approach does not add any extra burden and can be carried out at the recommendation, individual, and team levels.

**Assay Sommelier.** Safety is an important reason for the failure of drug candidates in clinical trials.16 On-target as well as off-target effects can cause corresponding adverse events. While on-target effects are typically within the focus of drug discovery teams, off-target effects are often identified only late (if at all) in the life cycle of a project. The initial assessment of likely off-target activity at the start of the optimization of a compound series is followed up only in the final stages of the clinical candidate optimization. For this, selected compounds are submitted to panel screens to assess their potential for adverse events based on their interactions with relevant off-target proteins.

One way to mitigate this issue would be a more frequent measurement of those panels; however, this is very expensive in terms of time and cost, especially if done for many compounds during the optimization phase. A shortcut to indicate whether or not a set of compounds is likely to show effects on a specific target is the similarity principle, by which similar compounds have similar properties.17 In practice, this allows the scientists to make decisions about whether or not to test a compound based on summary statistics of previous measurements for related compounds.

The objective of the Assay Sommelier recommender is to proactively indicate relevant assays, based on outlying activity data for similar molecules as compared to their peers. For this, we take advantage of the recent advances in the performance of chemical similarity searches.18 This allows us to search vast amounts of compounds interactively and to combine the results with the corresponding biological activity data.

To aggregate the similarity and bioactivity data, we combine a nearest-neighbor search with z-scaled potency measurements. A heuristic retrieves k compounds and ranks each relevant assay according to their average absolute z-scaled responses (see Figure 7). This approach allows for the ranking of either experimental data only or a combination of experimental and in-silico readouts. Due to the ease and efficiency of the ranking, the Sommelier can suggest results in almost real time to the chemist. So far, this system was applied to only a selection of projects.

**Transformer.** In a case under development, “Transformer,” we will provide recommendations to medicinal chemists who are addressing compound optimization challenges, e.g., reducing microsomal clearance. Here, internal knowledge of past transformations that have beneficially changed the property of a compound is transferred between programs and generalized from one chemical scaffold to another via matched molecular pairs. The trigger for recommendations will be a scientist tagging a molecule with a particular challenge in the CSF tool,19 an internal company-wide system for tracking medicinal chemistry progress in the various compound series. As a series evolves, the team records the key molecule(s) and tags the properties that are the focus of their optimization efforts. These tags will launch a search in our in-house matched molecular pair database for transformations that can be applied to the key molecule and are predicted to improve the property of interest.

Our vision is to process the results of that search and to select a small set of high-quality and novel (transformed) molecules for the team to consider. Quality considerations include confidence in the prediction that the transformation will have the desired impact on the tagged property, as well as an assessment of the impact on other properties. For the sake of novelty, we need to check that the exact compound has not already been made and tested for that property. We may also want to check that the same transformation has not already been assessed in this team on a related series. In its first incarnation, this recommender will use the same email platform as described above for Chem Recommender and NIBR Hops. When the searches and analyses become fast enough, the recommendations can become more instantaneous and, for example, pop up as soon as a molecule gets tagged, during data analysis, or during ideation.

Current development efforts for this recommender focus on creating a workflow that will take a key molecule with a tagged challenge and create an annotated set of transformed molecules for team assessment. The computational engine we are using is mmpdb, an open-source matched molecular pair (MMP) platform to create, compile, store, retrieve, and use MMP rules.18 We have used mmpdb to build a database of MMP rules and their impact on greater than 10 properties that are
commonly monitored in medicinal chemistry optimization, such as hERG binding, CYP inhibition (multiple isoforms), permeability, solubility, microsomal clearance (multiple species), plasma-protein binding, and distribution coefficients. The mmpdb platform includes a code for creating transformed molecules that are optimized for given properties. In the case of this recommender, we will first develop an automated workflow that expert users can apply with a number of teams to develop the criteria that will allow selection of the high-quality, novel molecules. Capturing and coding these criteria will allow for the implementation of “Transformer” as an autonomous recommender.

The business case for this recommender is clear when considering the fact that the top challenges across our medicinal chemistry portfolio are also the challenges that have been addressed by many successful optimization efforts. In 2019, the top-two challenges in the CSF system were solubility and microsomal clearance. Together, they account for almost 40% of all tagged challenges. When we built our in-house transformation database, the two properties with the most MMP rules were solubility and rat microsomal clearance.

**Every Well Counts.** In a case under planning, “Every Well Counts,” we seek to optimize the productivity and per-plate return on the investment of bioassay runs in a way that is impactful for project teams, and NIBR at large. Routine assay plates are often only partially filled, as assay requests rarely contain a number of compounds that is an exact multiple of the plate capacity (e.g., if 20 compounds are to be assayed but there is room for 44 dose–response curves, 24 will be left empty). In some cases, assay automation results in reagent addition to the entire plate even if some wells do not contain test compounds. For these cases, filling those empty wells with relevant compounds would generate additional data without any change in the number of plates being run or in the amount of biological reagent being used. In our estimate, we could generate many thousands of additional dose responses per year at NIBR, with minimal additional effort, if the empty wells on plates were filled with additional compounds. “Every Well Counts” endeavors to do this in an automated and AI-informed fashion.

We envision a system where projects can opt in on a per-assay basis to have partial plates filled with additional compounds, using a predictive model or other automated selection criteria chosen by the project team. This system would leverage our existing compound management system that automates dose–response plate creation, such as compound cherry picking, dilution, and spotting. The trigger in this case would be a partially filled dose–response plate following the deadline for project teams to submit compounds to an assay. The data generated from this approach could serve various purposes for both project teams and the organization. For a given project, wells could be filled with compounds informed by computational models such as Profile-QSAR, or recommendations from systems such as NIBR Hops and Assay Sommelier. This could potentially identify compounds with improved potency by exploring additional chemical space and investigating potential off-target activity. Alternatively, compounds could be selected with the intent to improve predictive models, which could benefit specific projects (e.g., a potency model) and the broader organization (e.g., an ADME model used by many projects or off-target predictions from panel assays). In another variation, a defined set of compounds could be run across multiple participating assays. This could be performed to assess compound promiscuity and refine frequent-hitter classification or to build out additional matched molecular pairs for all plate-based ADME assays. These data that would feed directly back to the “Transformer” system discussed above. Metrics for measuring success include the number of additional dose–response curves generated, the novel SAR established, the per-plate return on investment (average cost per dose–response curve), and the improvement in predictive models. Safeguards would need to be established to avoid depleting key compounds unintentionally, which is an extension of existing systems.

This approach of automatically filling in partial plates using AI expands on the recommender concept, allowing teams or the organization to select a recommendation system based on their goals and to generate data for those recommendations with minimal additional time and resource requirements.

**DISCUSSION**

Deploying recommenders in any enterprise, especially in a scientific setting like ours, is fundamentally different than deploying consumer-oriented recommendation systems. Our goal is to implement systems that inform medicinal chemists of relevant work in order to speed the drug discovery process. While much has been published about consumer systems (e.g., algorithms, evaluation), little has been written about enterprise recommendation systems. Below are some of the challenges we have observed that we feel are relevant to such scientifically focused enterprise recommendation systems.

First, what is the best way to deliver recommendations to our scientists? Ideally, like commercial shopping, music, or movie systems, recommendations should be delivered in context. It makes sense to show related products when a shopper is looking to make a purchase; however, this is not always possible with production systems that are often written by third parties, such as electronic laboratory notebooks. We do not always control the underlying data or the delivery user interface. For Chem Recommender and NIBR Hops, we have used email to send our recommendations, but this removes them from the scientists’ work context. In the future, we may consider using intelligent speaker systems within the lab to alert chemists of relevant information in a laboratory-friendly manner.

Second, how do we evaluate scientific recommender systems? In commercial recommenders, evaluation criteria are relatively straightforward. Did an email campaign result in a click-through to the company website? Did a user purchase a suggested item? Did the user watch the entire recommended movie or add the recommended song to a playlist? In an enterprise, this is more difficult. For example, Chem Recommender emails were designed to provide as much information to the user as possible; users only needed to click to get additional information. If a chemist does not click through, we have no information on whether the recommendation was good or not. And, if an otherwise good recommendation is rated poorly, was it because the chemist already knew about it or because it came too late? One approach we are investigating for NIBR Hops is to see if a recommended compound was ordered for the project after a recommendation email was sent. But, even in this case, was the compound ordered because of the recommendation, or was the project team already considering that compound (which is also an issue for commercial recommenders)? Unfortunately, our recommendation systems do not have insight into these external considerations.

Third, how do you determine what to recommend to a user? Commercial recommendation systems tend to use variants of collaborative filtering algorithms. In these systems, a user is
Recommendations for prior syntheses, readily available compounds, relevant assays, molecular transformations, and additional compounds to assay enable scientists to learn about relevant work, optimize time in the laboratory, and ultimately speed the drug discovery process.

In many ways, our approach is the “low-hanging fruit” of AI in drug discovery. Rather than a wholesale change in tactics, we seek to build upon existing data and systems in order to nudge our medicinal chemists in fruitful directions. We are convinced that our drug discovery cases will inspire additional investment in recommender systems by others and lead to the evolution of the field. We also believe that our real-world examples will prompt consideration of additional dimensions of impact, along with the quality of predictions. There are certainly challenges to our approach, but favorable reviews of our deployed systems have given us the confidence to proceed.

**ASSOCIATED CONTENT**

**Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jmedchem.9b02130.

High-level flow of the NIBR Hops process and associated software (PDF)

**AUTHOR INFORMATION**

**Corresponding Author**

Mika Lindvall — Novartis Institutes for BioMedical Research, Emeryville, California 94608, United States; orcid.org/0000-0002-5637-8948; Email: mika.lindvall@outlook.com

Authors

Steven L. Rohall — Novartis Institutes for BioMedical Research, Cambridge, Massachusetts 02139, United States
Lydia Auch — Novartis Institutes for BioMedical Research, Emeryville, California 94608, United States
Jonathan Gable — Novartis Institutes for BioMedical Research, Emeryville, California 94608, United States
Jacob Gora — Novartis Institutes for BioMedical Research, Cambridge, Massachusetts 02139, United States
Johanna Jansen — Novartis Institutes for BioMedical Research, Emeryville, California 94608, United States; orcid.org/0000-0003-3937-6243
Yipin Lu — Novartis Institutes for BioMedical Research, Emeryville, California 94608, United States; orcid.org/0000-0002-3400-4400
Eric Martin — Novartis Institutes for BioMedical Research, Emeryville, California 94608, United States; orcid.org/0000-0001-7040-5108
Margaret Pancost-Heidebrecht — Novartis Institutes for BioMedical Research, Cambridge, Massachusetts 02139, United States
Bill Shirley — Novartis Institutes for BioMedical Research, Emeryville, California 94608, United States
Nikolaus SteiI — Novartis Institutes for BioMedical Research, CH-4056 Basel, Switzerland

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jmedchem.9b02130

Notes

The authors declare no competing financial interest.
**ACKNOWLEDGMENTS**

The authors thank Peter Skewes-Cox for insights on many of these recommender systems. We also thank all of the chemists at NIBR who took action based on our reaction and scaffold hopping recommendations.

**ABBREVIATIONS USED**

ADME, absorption, distribution, metabolism, and excretion; AI, artificial intelligence; CSF, compound series and favorites; CYP, cytochromes P450; ELN, electronic lab notebook; hERG, human Ether-á-go-go-Related Gene; IC$_{50}$, half-maximal inhibition concentration; ML, machine learning; MMP, matched molecular pair; NIBR, Novartis Institutes for BioMedical Research; PDB, Protein Data Bank; QSAR, quantitative structure–activity relationship; ROCS, rapid overlay of chemical structures; SAR, structure–activity relationship

**REFERENCES**


