Building Admiral, an Automated Molecular Dynamics and Analysis Platform

Matthew P. Baumgartner and Hongzhou Zhang*

Cite This: *ACS Med. Chem. Lett.* 2020, 11, 2331–2335

Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: We present Admiral (Automated Docking and Molecular dynamics InfoRmatics and AnaLysis), a platform which automates the process of running molecular docking and molecular dynamics on compound designs for medicinal chemistry project teams. In addition to running the simulations, Admiral analyzes the simulation and automatically generates a PowerPoint slide, with the goal of having all the information required to decide whether to synthesize the compound in one place. This information includes results and analyses from the MD simulation, predicted ADME and physical-chemical properties, information on similar compounds in the SAR, and an animated GIF of the simulation. This report is then emailed to the compound designer, generally within the same day. Within Eli Lilly and Co., we have developed and deployed Admiral on an internal discovery project where it has been heavily used by the project team. Several additional discovery projects have adopted the platform in recent months.

KEYWORDS: molecular dynamics, molecular docking, automation, computational chemistry, medicinal chemistry

Admiral: Automated High Throughput Molecular Dynamics Platform

Drug discovery is an iterative process based on empiricism, meaning we design and make compounds and then test them in biological assays to see how well they work. From this new knowledge, we then design the next set of compounds and repeat the process. At a high level, there are two ways to improve drug discovery learning cycles: shortening the learning cycles (going faster) or increasing the learning in each iteration (making better compounds). In this work, we will discuss our efforts to develop and apply computational methods to the pre-clinical discovery space with the goal of improving the both the speed and quality of experimental and computational learning cycles.

This work builds on our previous work in bringing process automation to the pre-clinical drug discovery space within Eli Lilly and Co.¹ Previously, we presented Kernel, which is an automated virtual assistant for medicinal chemistry teams. The guiding principle behind that work and the work presented here is the idea of moving from reactive user interfaces to proactive ones. Each week, when new assay data comes out for a project, Kernel automatically summarizes the results and sends a notification to the team members. In addition to the assay summaries, it identifies compounds that have “surprising” assay results (i.e., are more potent than an activity model would have predicted). It then automatically performs *de novo* design around these compounds and suggests the designs to the medicinal chemistry team. All of this is done automatically in such a way that analyses are completed almost before the chemists know to ask for them.

We have taken this principle and applied it to an internal, structurally enabled oncology project. In addition to other,

more traditional project support activities (creating docking or QSAR models, interactive designs, etc.), we decided to apply molecular dynamics (MD) simulations to this project to a degree that has not been done previously within Lilly. Our goal is to run an MD simulation on *every* designed compound and use information from the simulation and other sources to prioritize designs prior to synthesis.

Initial Effort. With the goal of running an MD simulation on every compound designed by project team members, the computational chemistry group quickly automated setting up and running the MD simulations. However, after the MD simulations had run, analyzing the results and recommending which compound designs should be prioritized remained a large manual effort. Primarily, it consisted of viewing and analyzing the MD trajectories, computing physical-chemical properties, and looking up the local SAR space. Once this information had been compiled, it was put into a PowerPoint slide deck along with a recommendation from the modeler of whether the compound should be made or not. This slide deck was then presented to the team at the next project meeting. As one might imagine, this was a labor-intensive and time-consuming process.

Received: August 19, 2020

Accepted: September 28, 2020

Published: September 28, 2020

ADMIRAL: Automated Docking and Molecular dynamics InfoRmatics and AnaLysis

Figure 1. Admiral process workflow.

Figure 2. Example simulation report of marketed drug imatinib in cAbl (PDB ID: 2PL0) with annotations. Included in the report are images of the designed compound and the most similar compound in the project SAR with its related assay data, a table of physical-chemical properties and predicted ADME properties, an animated GIF of the simulation, the simulation interaction diagram, and a time-series view of the protein–ligand contacts. Additionally, quantification of contact stability and classifications of the protein and ligand RMSD and RMSF are included, as well as links to the full Simulation Interaction Diagram report and to frames extracted from the simulation.

Admiral Overview. To reduce the amount of manual work involved in setting up, running, and analyzing the MD simulations, we built a tool called Admiral (Automated Docking and Molecular dynamics InfoRmatics and AnaLysis); the main steps of the workflow are shown in Figure 1. Admiral is a fully automated tool that detects when compound designs are added to a centralized project tracker. It then performs molecular docking, selects a pose, and sets up and runs an MD simulation. When the simulation is finished, it performs several analyses of the simulation. It then *automatically creates a PowerPoint slide* with all the information needed to evaluate the compound design. Finally, it emails the report to the person who designed the compound.

Docking and Simulation. Compounds from the centralized internal project tracker software are docked to the protein using Glide 2019.2.^{2–5} A docked pose is selected either by taking the top scoring pose or by some project-specific or scaffold-specific criteria. For certain scaffolds in a project, we have implemented some automated RMSD-based pose-filtering to ensure that the core of the docked pose is in the expected position. In the case where the input compound is not one of the specified scaffolds, the top scoring pose is

simulated by default. Additionally, we have included the ability to run manually docked poses when the automatically selected pose is identified to be inadequate or incorrect. The docked protein–ligand complex is then prepared using the “prepwizard” tool^{6,7} and solvated in an orthorhombic water box using the “multisim” command.^{8,9} The MD simulations are then run for 100 ns using the hydrogen mass repartitioning protocol, which allows us to run at a 4 fs time step (twice as fast as the standard 2 fs time step).¹⁰

Automated Analysis. When the simulation is complete, the Simulation Interactions Diagram report from Desmond is generated as a PDF file using the ‘event_analysis.py’ and the ‘analyze_simulation.py’ scripts.⁹ This report contains several standard analyses of the simulation, including ligand and protein RMSD and RMSF calculations, protein secondary structure analysis, protein–ligand interaction analysis, and ligand torsion profiles.

Given that within Lilly, and we suspect that within many other companies as well, the *de facto* medium of exchange for scientific ideas or analysis is PowerPoint slides (whether we like it or not), we sought to streamline the process of creating PowerPoint reports. This will allow chemists to be able to

MD3 On SPime 11.1

File Edit View Design Analysis Projects Admin Help Manager

Current Project: Cancer / Project A

Compound Sets

Quick Find

Search

Project A

Project Tracker

			Authors		Project A MD Report			Project A Docking			clogP (Chemaxon)	HTSA pH6	Lilly PSA
	id	structure	EntryTime	Hypothesis	md_report	has_report	best_score	pose	F Docking Classification	clogP	F pH6 HTSA Class	lly-psa	
14	CHEMBL1623510_		5/24/2020	Add HBond	Show Details	Y	-13.485	Show 3D Image	Potentially good	3.504	<0.3 mg/ml	149.6	
15	CHEMBL4117886_		4/23/2020	Interact with resid...	Show Details	Y	-12.207	Show 3D Image	Potentially good	2.23	<0.3 mg/ml	122.1	
16	CHEMBL4117662_		6/4/2020	scaffold 1	Show Details	Y	-11.449	Show 3D Image	Potentially good	2.996	<0.3 mg/ml	116.4	
17	CHEMBL4116226_		6/15/2020	Improve solubility	Show Details	Y	-11.81	Show 3D Image	Potentially good	4.549	<0.3 mg/ml	108.4	
18	CHEMBL4116480_		6/15/2020	FEP set quinc5	Show Details	Y	-11.272	Show 3D Image	Likely less good	5.192	0.01 - 0.1 mg/ml	95.87	

1 of 186 Selected186 of 186 Filtered

Figure 3. Central project tracker. Team members add their designs to the tracker, and Admiral detects them and starts the simulations. When the simulations are complete, a link to the report is added back into the tracker.

quickly paste the generated analysis slide into their own presentations. To do this, we used the Python package “python-pptx”,¹¹ which enables the programmatic creation of PowerPoint presentations and slides. Using the python-pptx library, adding an image to a slide involves only calling the appropriate function (add_picture), giving it the image file, and telling it where on the slide the image should go. In this way, slides can be built with images, tables, text boxes, and animated GIFs.

Admiral MD Simulation Reports. Upon completion of the MD simulation and generation of the Simulation Interaction Diagram report, information from several different sources is collected or computed, including information from the simulation, calculated physical-chemical and ADME properties, near neighbors in the SAR and their assay data, as well as an animated GIF of the simulation. All of the images and data are automatically inserted into a single PowerPoint slide, an example of which is shown in Figure 2.

Included on the left side of the slide (see Figure 2) is a 2D image of the designed compound and a table of properties, including the name of the person who designed the compound and when they designed it, as well as the hypothesis that the design is addressing. Also in the table are computed physical-chemical properties and predicted ADME properties as well as the docking score and a project-specific classification of the docking score was generated by comparing docking scores to experimental assay data and determining a score cutoff below which nearly all compounds had poor binding. In the top middle of the slide is the most similar compound in the project SAR (by 2D similarity) and its key assay data. In the bottom middle of the slide is an animated GIF, created using PyMOL,¹² of the simulation, which allows one to visually see what is happening in the simulation. On the bottom left of the slide is an image showing the protein–ligand contacts over time. At each time step, if a residue is making a contact with the protein (hydrogen bond, π -stacking, hydrophobic, etc.), an orange line is drawn. This image can be used to get a visual view of the

stability of the contacts over time. Contacts that are consistent over the simulation show up as (mostly) solid orange bars, and conversely, inconsistent contacts show up as broken bars. This visual interpretation has been quantified by calculating the average percent contact in 10 ns windows of the simulation. If the standard deviation of the means is above a certain threshold, the contact is considered “unstable”. Finally, in the top right corner of the slide is the “MD Health report”, which is a classification (Good, Questionable, or Bad) of the protein and ligand RMSDs and RMSFs. Simulations are classified as having “Questionable” or “Bad” ligand or protein RMSDs if they are more than 2 Å away from their starting positions (ligand heavy atoms and protein *Ca*’s) for more than a certain percentage of the simulation.

User Interaction. Admiral was designed with several goals in mind: first, to make it as easy as possible for users (primarily medicinal and computational chemists and structural biologists) to interact with it; second, for the results to be as robust and meaningful as possible; and finally, for the results to be returned to the users as quickly as possible. The project team uses an in-house desktop application called “SPime” to design molecules using *in silico* models (physicochemical, ADME, and docking models). Figure 3 provides a snapshot of the SPime platform showing the project tracker. All users need to do is to add their compounds to the project tracker in SPime, which the team already uses to track and prioritize design ideas. From there, Admiral automatically gets the compounds, simulates them, and generates the PowerPoint report. The user who designed the compound then gets an email alert informing them that the simulation is complete, which includes a link to the report. The link to the report also gets added back into the central project tracker, where all users can access it.

The reports are designed to include all the information needed to decide whether to synthesize the compound. To make the report as robust and meaningful as possible, the reports have gone through (and continue to go through) multiple rounds of changes in response to feedback from the team, including adding desired data to the report and removing

data deemed to be not helpful. Additionally, as the project evolved, the assays that the team focused on changed, and so the report was updated to reflect this.

The third design goal is timeliness. If a theoretical simulation could give a perfect answer as to whether to make a compound, but it takes 5 years to give the answer, it would not be useful in the week-to-week design cycle in which biotech and pharmaceutical companies. In a more realistic situation, if the simulation takes even 2 weeks to return a result, the medicinal chemists are likely going to just make the compound, assuming the synthesis is not overly complex. To reduce the turnaround time from when the user adds the compound the project tracker to when they get the results, we have moved to running on the fastest machines available on Amazon AWS and to using hydrogen mass repartitioning. This has improved the turnaround time from around 18–24 h to 4–5 h.

Impact. Within Lilly, we deployed Admiral on one internal pre-clinical discovery project on a structurally enabled oncology target in early 2020; in more recent months, it has been adopted by four additional projects. As previously mentioned, before creating this automation system, the computational chemists on the project would semi-manually run the MD simulations and then manually analyze the simulations and create PowerPoint slides (by copy-pasting information from three or four different sources) to present to the project team. In one design cycle before the automation platform had been built, approximately 20 man-hours were spent to evaluate around 50 compounds and create a presentation. After the Admiral was built and deployed, the computational chemists were able to evaluate 115 designs in about 1.5 h. This represents an approximately 30-fold increase in productivity and allows the computational chemists to focus on simply analyzing the results rather than all of the manual labor involved in opening simulation files and doing routine calculations.

In addition to the reduction in workload for the computational chemists on the team, Admiral also enabled the medicinal chemists to utilize MD simulations essentially by themselves, without the involvement of the computational chemists. To fully enable this, the project team was trained on several occasions on what MD simulations are, what their limitations are, and how to (and how not to) interpret the data. When chemists receive reports that they struggle to interpret or have other questions, the computational chemists on the project provide support and guidance. Admiral has been heavily used by the project team, with a large portion of the team using the generated report summary slides in their own presentations to the team.

Considerations. The Admiral tool is likely to be applicable to most structurally enabled pre-clinical discovery projects. There are several factors that are required or recommended to be successful. These are not necessarily specific to Admiral but are general requirements for any docking and MD workflow. The first requirement is to have an X-ray structure or a high-quality homology model of the protein of interest, preferably bound to a ligand. Next, the amount of flexibility that the protein experiences upon binding different small molecules should be relatively small. If upon binding a small molecule the protein goes through a large-scale rearrangement, it will be necessary to run the MD simulations much longer to capture this, if indeed the MD simulations can be done at all.

Conclusions. In this work, we present Admiral (Automated Docking and Molecular dynamics InfoRmatics and AnaLysis), which is an automation platform for running and analyzing molecular dynamics simulations designed to work for Lilly's internal discovery project teams. It works by taking compound design ideas that the project team enters into a centralized project tracker and sets up and runs MD simulations. These MD simulations are analyzed, and a report is generated in the form of a PowerPoint slide which contains information from the MD simulation as well as from predicted properties and SAR near neighbors. This report is designed to have all the information needed to evaluate the design and decide whether it should be prioritized for synthesis. This has allowed computational and medicinal chemists on the project team to focus on interpreting the data and making decisions based on it rather than spending laborious hours simply collecting the information. Admiral started as a prototype from one oncology project in early 2020. Given the success that we have had with it, this platform has been expanded to additional discovery projects recently. We are standardizing the process to make it more easily deployable to the future projects.

AUTHOR INFORMATION

Corresponding Author

Hongzhou Zhang – Research Information and Digital Solution, Eli Lilly and Company Inc., Indianapolis, Indiana 46285, United States; orcid.org/0000-0003-1160-1170; Email: zhang_hongzhou@lilly.com

Author

Matthew P. Baumgartner – Computational Chemistry and Cheminformatics, Eli Lilly and Company Inc., San Diego, California 92121, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsmmedchemlett.0c00458>

Notes

The authors declare no competing financial interest.

REFERENCES

- (1) Vidler, L. R.; Baumgartner, M. P. Creating a Virtual Assistant for Medicinal Chemistry. *ACS Med. Chem. Lett.* **2019**, *10* (7), 1051–1055.
- (2) Friesner, R. A.; et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–49.
- (3) Friesner, R. A.; et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–96.
- (4) Halgren, T. A.; et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47* (7), 1750–9.
- (5) *Glide*, Schrödinger Release 2019-2; Schrödinger LLC, New York, 2019.
- (6) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W.; et al. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27* (3), 221–234.
- (7) *Protein Preparation Wizard*, Schrödinger Release 2019-2; Schrödinger, LLC, New York, 2019. *Epik*; Schrödinger, LLC, New York, 2016. *Impact*; Schrödinger, LLC, New York, 2016. *Prime*; Schrödinger, LLC, New York, 2020.
- (8) Bowers, K. J.; et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, Tampa, FL, 2006.

- (9) *Desmond Molecular Dynamics System*, Schrödinger Release 2020-3; D. E. Shaw Research, New York, 2020. *Maestro-Desmond Interoperability Tools*; Schrödinger, LLC, New York, 2020.
- (10) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **1999**, 20 (8), 786–798.
- (11) Canny, S. Python library for creating and updating PowerPoint (.pptx) files, <https://github.com/scanny/python-pptx>.
- (12) *PyMOL Molecular Graphics System, Version 1.8*; Schrödinger, LLC, New York, 2015.