



“The bioisosteres generation we can do using the CDD Vault AI Module is great for triggering new ideas and structural insights.”

- Dr. Lori Ferrins, Associate Professor, Pharmaceutical Sciences, Lab for Neglected Tropical Disease Drug Discovery, Northeastern University

Customer Case Study



Northeastern University Lab for Neglected Tropical Disease Drug Discovery Collaboration Integrates CDD Vault AI Module Generative Bioisosteres with NVIDIA BioNeMo NIM microservices for AlphaFold2 and DiffDock

Situation

Northeastern University Lab for Neglected Tropical Disease Drug Discovery, jointly run by Professor Michael Pollastri and Professor Lori Ferrins, focuses on diseases that do not attract sizable research investment in the private sector since they affect the poorest parts of the world, making it difficult for pharmaceutical companies to recoup research and development costs.

The Neglected Disease Lab focuses on Trypanosomatid parasites (*Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania spp.*), fungal

pathogens (*Candida albicans* and *Candida auris*), as well as pathogenic free-living amoebae (including brain-eating amoebae) which lead to a range of brain, skin, eye, and disseminated diseases in humans and animals which are typically fatal.

The Lab for Neglected Tropical Disease Drug Discovery, which collaborates with other labs domestically and internationally, needed to find a better way to manage data, which was often stored on spreadsheets and in other non-centralized documents.

Solution

Northeastern University's Lab for Neglected Tropical Disease Drug Discovery deployed Collaborative Drug Discovery's CDD Vault, the hosted drug discovery informatics platform that securely manages both internal and external biological and chemical data.

The deployment includes the CDD Vault AI Module, which supports researchers managing and interpreting complex datasets, predicting compound behavior, and identifying potential leads with greater precision, and is directly integrated with the CDD Visualization Module for multiparameter optimization.

The Lab augments the CDD AI Module with use of two AI models from NVIDIA's NIM microservices —AlphaFold2, the deep learning model that reduces the time it takes to determine a protein's structure; and DiffDock, which predicts the 3D orientation and docking interactions of a small molecule ligand to a protein. Generative Bioisosteres and ultrafast deep learning similarity search within CDD Vault suggests new molecules which the NVIDIA NIMs evaluate *in silico* with Northeastern providing experimental validation in the laboratory.

"We use CDD Vault as a secure central repository to support our own lab work and as a collaboration point for our industry, academic, public-private partnerships, and government collaborations," says Dr. Lori Ferrins, Associate Professor, Pharmaceutical Sciences, Lab for Neglected Tropical Disease Drug Discovery at Northeastern University.

Benefits

The Lab for Neglected Tropical Disease Drug Discovery has found a number of benefits since adopting the combination of CDD Vault and NVIDIA technologies, including a short list:

- CDD Vault securely handles private data and collaborative data
- CDD Vault Visualization and Search for plotting and identifying trends
- AI Module to generate bioisosteres
- In-Vault access to ChEMBL, SureChEMBL, and Enamine
- NVIDIA's powerful AlphaFold2 NIM
- DiffDock integrated with AlphaFold2

A Trusted Repository: "CDD Vault is Central to Everything that We Do"

The Lab for Neglected Tropical Disease Drug Discovery found the trusted central repository it needed with its deployment of CDD Vault.

"CDD Vault is central to everything that we do," Dr. Ferrins says. "We use CDD Vault to store more than 8,000 small molecules. As data comes in from our own lab and from our collaborators, it goes immediately into the Vault, and everyone is alerted so they can see the new data that has been uploaded. NVIDIA NIMs allows us to add powerful modeling capabilities. So our *in silico* workflows now complement our experimental workflows."

The Lab subdivides CDD Vault into secure partitions so labs can have a separate data store for different projects.

CDD Vault also helps for quality control and in identifying variations. "We can track differences in compounds based on batch variations," Dr. Ferrins says. "Whenever we are scaling up a compound and it enters into the system as a new batch, we're able to track the *in vitro* readings of the different assays that we're using. If we see a difference, we can look to see if the change can be attributed to something that's happened in terms of that batch, or if there were other factors." NVIDIA adds

modeling for targets like kinases, even when there are no known crystal structures.

CDD Vault Visualization: “Great for Plotting and Identifying Trends”

The Lab uses CDD Vault Visualization in exploring data, including plotting and identifying trends.

“The CDD Vault data visualization tools enable a different way of exploration,” Dr. Ferrins says. “Our Ph.D. student just defended her thesis a couple of weeks ago, and she used CDD Vault Visualization quite a lot during her dissertation writing. She used it to plot biochemical versus in vitro data, thinking about it in the context of lipophilicity.”

Lipophilicity is important in the Lab’s work as potential drugs would need to enter cell membranes, as well as the parasite membrane.

“We were trying to see if there was a key lipophilicity range where our compounds tended to work well enough to get into the parasite,” Dr. Ferrins says. “We use the data visualization tool because it’s great for plotting and identifying trends in our data. We want to better understand things like does lipophilicity correlate with aqueous solubility? Typically, we would think a compound that’s really lipophilic would be not very soluble. But sometimes there’s an exception, and the tools within CDD Vault help us to identify and understand that better.”

The Lab has sectioned part of its CDD Vault as a “sandbox” where scientists can explore ideas before deciding whether they want to register compounds into their project Vault.

“One of the challenges that we have in general, is we think that there’s a very narrow range in

“The CDD Vault data visualization tools enable a different way of exploration. We were trying to hybridize molecules. If we hadn’t had all of that data there together in the Vault, it would have been much more difficult to actually identify the overlap.”

- Dr. Lori Ferrins, Associate Professor, Pharmaceutical Sciences, Lab for Neglected Tropical Disease Drug Discovery at Northeastern University

terms of lipophilicity and polar surface area in which we have to be operate to actually have compounds that translate from a biochemical assay into a phenotypic assay,” Dr. Ferrins says. “We can use the sandbox to start to ideate, to see how a chemical modification translates in terms of these predictive properties. When something looks promising, we move it to our real Vault.”

CDD Vault Searchability Provides A Powerful Tool, NVIDIA AlphaFold2 and DiffDock NIMs Provide Next Generation Analytics


The Lab makes good use of the searchability of CDD Vault.

“The Vault searchability provides us with a powerful tool,” Dr. Ferrins says. “The number of ways in which we can search the vault are quite incredible. For example, we can search data in terms of readouts in a particular assay. So for

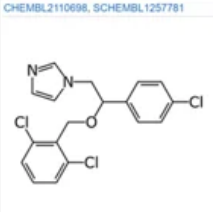
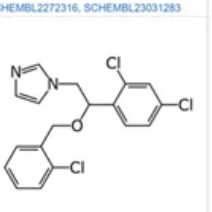
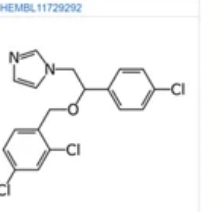
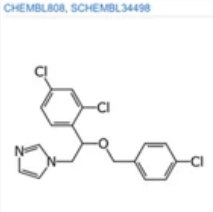
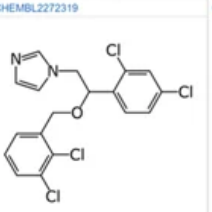
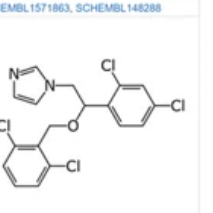
CDD.VAULT · CDD New Demo Vault ▾

Explore Data ELN Import Data Reports Settings 4 Calculate properties for structures

DEMO-0001007

Similar Structures  ChEMBL SureChEMBL

25 hits with this scaffold

 1.000 MW: 381.690, LogD: 4.0, LogP: 5.5, LogS: -6.6, HBA: 3, RB: 6, TPSA: 27.1, CNS-APO: 3.2	 0.892 MW: 381.690, LogD: 3.6, LogP: 5.3, LogS: -6.4, HBA: 3, RB: 6, TPSA: 27.1, CNS-APO: 3.4	 0.882 MW: 381.690, LogD: 4.2, LogP: 5.6, LogS: -6.9, HBA: 3, RB: 6, TPSA: 27.1, CNS-APO: 3.2
 CHEMBL808, SCHEMBL34498	 CHEMBL2272319	 CHEMBL1571863, SCHEMBL148288

CDD-3346008 (2 molecules with this structure)

- Find molecules with this structure
- Suggest biosisteres using deep learning similarity
- Find ChEMBL and patented molecules using deep learning similarity

The CDD Vault AI Module includes AI-powered biosister generation.

example, when we're thinking about trying to develop a new project, we might be looking for compounds that meet a certain threshold in terms of their activity, have a certain level of cytotoxicity, selectivity, maybe solubility is a factor, or lipophilicity, and we can add all of those features into a search and get an idea of what are our best compounds are across our entire vault. This helps identify opportunities to start an optimization campaign.”

Dr. Ferrins speaks of recent work with kinase inhibitors in which searching the Vault provided interesting finds.

“We have a lot of kinase inhibitor chemotypes, which tend to have common structural features,” Dr. Ferrins says. “And so what belongs to one series might closely overlap with another series, which we might be able to learn from. Fairly recently we identified two series of

compounds that had a difference in a key piece to their structure. This led us to see if we could effectively cross pollinate. We were trying to hybridize molecules. If we hadn't had all of that data there together in the Vault, it would have been much more difficult to actually identify the overlap.”

Supports Collaboration

CDD Vault and NVIDIA NIMs are powerful collaboration tools, which is especially important because of its domestic and international collaborations.

“We have a lot of collaborations,” Dr. Ferrins says. “We do all the chemistry here, but we collaborate very widely for the biology. We collaborate here in the US, and internationally. The large volume of data produced across all of our collaborations is fed into CDD Vault and

NVIDIA Alphafold2 and DiffDock NIMs allow us to do modeling. Our biological collaborators have seats in the vault, so they're able to look at the data themselves as it's being produced and handled."

Using CDD Vault as a central repository removes one of the challenges the Lab previously faced when dealing with spreadsheet data from within its own lab, as well as all of the collaborators.

"Storing our data in CDD Vault is much safer than in a spreadsheet of data, where if one person makes a typo or deletes a line, all of a sudden everything is chaos," Dr. Ferrins says.

Using the CDD Vault AI Module to Generate Bioisosteres

The Lab values the CDD Vault AI Module as an idea generator—including the ease with which researchers can use the AI Module to generate bioisosteres.

"The bioisosteres function is a tool to foster creativity," says Dr. Michael Pollastri, Senior Vice Provost and Academic Lead, Roux Institute at Northeastern University. "Often when we are heads down in a project it isn't always easy to see what lateral moves could be made in the structural space. The generative aspect of the bioisosteres function helps foster creativity and supports us asking 'what if...' questions."

Dr. Ferrins agrees.

"The bioisosteres generation we can do using the CDD Vault AI Module is great for triggering new ideas and structural insights," Dr. Ferrins says. "Each bioisostere includes the predicted properties and degree of similarity, so it provides a quick snapshot of if you were to replace a particular functional group in one

place, how it will impact things like molecular weight and lipophilicity and other factors we're thinking about as we're trying to design our molecules." Bioisosteres putatively have bioactivity, with the NVIDIA Alphafold2 and DiffDock NIMs they can be tested in silico and then validated in the laboratory.

Dr. Ferrins points to the lab's work with kinetoplastid parasites *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania spp.* The Lab has biochemical and in vitro data on the

"Often when we are heads down in a project it isn't always easy to see what lateral moves could be made in the structural space. The generative aspect of the bioisosteres function helps foster creativity and supports us asking 'what if...' questions."

- Dr. Michael Pollastri, Senior Vice Provost and Academic Lead, Roux Institute at Northeastern University

kinetoplastids, and Dr. Ferrins used bioisosteres to look at variations in ring structure.

"There is an animopyrimidine ring that is part of the hinge binding interaction," Dr. Ferrins says. "For the kinase binding pocket you need a hydrogen bond donor-acceptor motif. We've made compounds that have a fluoro or chloro that might cause a steric clash with the gatekeeper from a selectivity perspective. And so it was interesting to have a look at what other options there might be for this part of the molecule. Bioisosteres provide a way to see what other structures might be considered."

In-Vault Access to ChEMBL, SureChEMBL, and Enamine

CDD Vault includes in-vault access to the ChEMBL and SureChEMBL databases and the Enamine catalog, providing secure and seamless access to these key resources.

“Having these resources in the AI Module is beneficial because we can quickly see what's commercially available from within the vault, without having to take our compounds out and put them into another search engine, or onto another platform,” Dr. Ferrins says. “One of the things we're doing at the moment is looking through all of the data that we have in our pathogenic free-living amoeba work, and trying to identify commercially available derivatives of some of the top compounds. We can explore what's available in the commercial space, and start to generate preliminary structure activity relationships. Having these resources all centralized in one location is really beneficial.”

NVIDIA AlphaFold2: “Such a Powerful Tool”

Dr. Ferrins sees NVIDIA AlphaFold2 as an important drug discovery tool because of its ability to model proteins in 3D.

“AlphaFold is such a powerful tool for our research,” Dr. Ferrins says. “When you think about the diseases that we work on, a lot of these proteins have never been crystallized. We don't know what the crystal structure is. We don't know how they exist in their 3D folded state.”

“Our research often targets parasite proteins for which X-ray crystal structures do not exist; this can limit our modeling approaches to targets where homology models can be built,”

Dr. Pollastri says. “AlphaFold affords the opportunity to expand the range of protein structures that we can explore as we seek putative targets for the compounds (generally via phenotypic screening).”

Seeing 3D structures should help with targeting.

“We've had a number of collaborative projects over the last few years, which are focused on target-based work,” Dr. Ferrins says. “That means getting crystal structures of compounds bound to a specific parasitic (where possible),

“AlphaFold is such a powerful tool for our research,” Dr. Ferrins says. “When you think about the diseases that we work on, a lot of these proteins have never been crystallized. We don't know what the crystal structure is. We don't know how they exist in their 3D folded state.”

- Dr. Lori Ferrins, Associate Professor, Pharmaceutical Sciences, Lab for Neglected Tropical Disease Drug Discovery at Northeastern University

predominantly kinase targets, and using that information to guide our optimization processes—such as what might be the solvent exposed region of the molecule and how could we target for selectivity.”

Dr. Ferrins continues: “Tools like AlphaFold will be hugely beneficial because we can actually

start to make more informed hypotheses as to how to generate compounds that are active against a particular parasitic target, and to expand the variety of targets we can pursue.”

DiffDock a Great Fit with AlphaFold2

Dr. Ferrins sees NVIDIA’s DiffDock as being a natural companion tool to AlphaFold.

“One of the things that we want to do is to have crystal structures of compounds bound to a specific parasitic kinase,” Dr. Ferrins says. “We use AlphaFold to generate the model and then use DiffDock to actually dock our compounds into that molecule. We can drive structure-based drug discovery programs on kinase targets that have never been reported and that we don't know the structure for.”

CDD Vault As a Teaching Tool

Northeastern University places a heavy emphasis on using AI to help drive and develop programs and teaching, in addition to supporting research.

“In the Fall I’m teaching a class on the principles of drug design,” Dr. Ferrins says. “I’m already loading our sandbox within CDD Vault with real data on pathogenic free-living amoeba. I'm using the sandbox because I don't want the students to be able to do anything in our real vault, but they will experience the power of running searches and exploring data with the Vault. With NVIDIA AlphaFold2 and DiffDock NIMs, we can share with the students the cutting edge modeling capabilities.”

The class will also use the CDD Vault AI Module.

“We want our students to know how to use resources like ChEMBL, which historically, we've had to leave CDD to go and run the search

ourselves, rather than having it all located in one platform,” Dr. Ferrins says. “We want them to start thinking about: Here's a compound, here's what it's known for in humans, now go and see how you might try to optimize that against particular organisms.”

About Collaborative Drug Discovery

Collaborative Drug Discovery provides a modern approach to drug discovery informatics that is trusted globally by thousands of leading researchers. Our CDD Vault is a hosted informatics platform that securely manages both private and external biological and chemical data. It provides core functionality including chemical registration, structure activity relationship, inventory, visualization, and electronic lab notebook capabilities. For more information, visit us at www.collaborativedrug.com.

About NVIDIA

NVIDIA (NASDAQ: NVDA) is the world leader in accelerated computing.

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, availability, and performance of NVIDIA’s products, services, and technologies, including NVIDIA AI Foundry, NVIDIA NIM microservices, NVIDIA Blueprints, NVIDIA RAPIDS, NVIDIA AI Enterprise software platform, NVIDIA BioNeMo, NVIDIA MONAI, NVIDIA DGX B200 systems, NVIDIA Blackwell architecture, and NVIDIA DGX Cloud; NVIDIA’s partnership and collaboration with third parties, and the benefit and impact thereof; third parties adopting NVIDIA’s products and technologies, the benefits and impact thereof,

and the features and performance of their offerings; and the combination of NVIDIA's AI and accelerated computing capabilities with the expertise of industry leaders being poised to usher in a new era of medical and biological innovation and improve patient outcomes worldwide are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

Many of the products and features described herein remain in various stages and will be offered on a when-and-if-available basis. The

statements above are not intended to be, and should not be interpreted as a commitment, promise, or legal obligation, and the development, release, and timing of any features or functionalities described for our products is subject to change and remains at the sole discretion of NVIDIA. NVIDIA will have no liability for failure to deliver or delay in the delivery of any of the products, features or functions set forth herein.

All rights reserved. NVIDIA, the NVIDIA logo, BioNeMo, DGX, NVIDIA NIM and NVIDIA RAPIDS are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. Features, pricing, availability and specifications are subject to change without notice.