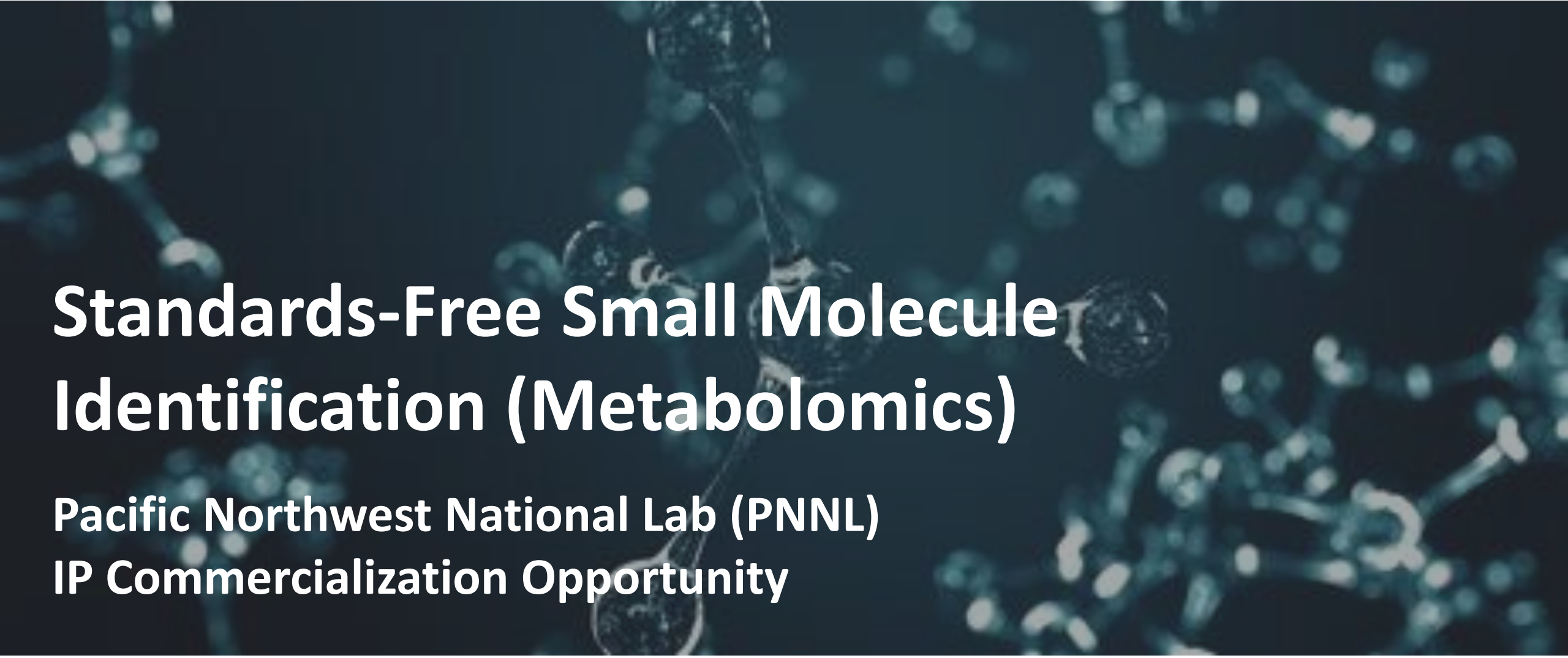


Standards-Free Small Molecule Identification (Metabolomics)

Pacific Northwest National Lab (PNNL)
IP Commercialization Opportunity

tradespace

Pacific Northwest
NATIONAL LABORATORY

Pacific Northwest National Lab (PNNL) Licensing & Commercialization Opportunity

Opportunity: PNNL has engaged Tradespace to approach select partners to commercialize **DarkChem** - a set of software tools for small molecule identification and discovery

Technology: End-to-end set of specialized machine learning tools for small molecules. Uses variational autoencoder (generative neural network) to learn and generate molecular structures based on desired chemical properties

Key Benefits of Darkchem:

- **Substance Identification:** Can identify over 90% of molecules in a sample
  - Traditional approaches only identify 20%
- **Property Prediction:** Based on molecular structure
- **New Molecule Discovery:** Two orders of magnitude faster than first-principles simulation (targeted on discovery for desired properties)

Opportunity Snapshot:

Available IP

- Patent Application (US20200176087A1)
- Expertise & Data Package
- Python / Tensor Flow code / tools
- Access to Research Team

Maturity Level

TRL 6: Software tools demonstrated in a relevant environment

Potential Partnership Models

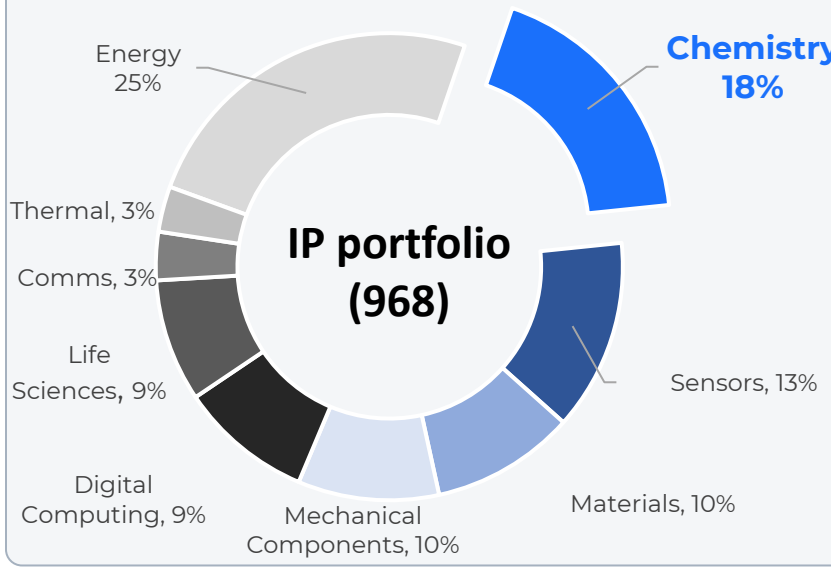
Patent License | Collaborative R&D | Sponsored Research

Pacific Northwest National Lab Snapshot

Leading Department of Energy lab with strengths in chemistry, data analytics, life science, and technological innovation in energy resiliency and national security. Based in Richland, WA

Annual R&D Spend: $1.01B **Staff:** 4,722

PNNL Computational Chemistry Expertise



- **Physics-Informed Learning Machines Center (PhILMS):** Collaboration with Sandia, MIT, Stanford, and others on Applied Deep Learning
- **Energy Sciences Center:** $100M Research facility dedicated to experimental and computational chemistry Research

National Labs with Computational Chemistry IP

Rank	Organization	IP Strength
#1	**PNNL**	**80%**
#1	Argonne National Lab	80%
#3	Army Research Lab	79%
#4	Naval Research Lab	61%
#5	NASA	55%
#6	HHS	51%
#7	Los Alamos National Lab	46%
#8	Dept. of Agriculture	41%
#9	Savannah River National Lab	30%
#10	Lawrence Livermore Nat'l Lab	29%

PNNL Toolkit Overview

Pacific Northwest NATIONAL LABORATORY

PNNL's approach includes four key tools to enable identification of molecules in biological and environmental samples using high-power computing:

- **Data Extraction for Integrated Multidimensional Spectrometry (DEIMoS):** Modular software tool that can extract features from data collected on multi-dimensional analytical platforms.
- **In-silico chemical library engine (ISiCLE):** Molecular dynamics and quantum chemistry-based work-flow used to automate chemical property predictions
- **Multi Attribute Matching Example (MAME):** Modular Python package that matches properties based on various chemical attributes.
- **DarkChem:** Variational autoencoder that learns a continuous numerical or latent representation of molecular structure, which can characterize and expand reference libraries

tradespace

Technology Overview: Standards-Free Small Molecule Identification

Pacific Northwest NATIONAL LABORATORY

IP Snapshot

Patents: 1 Patent Application

Research Team Available: Yes

Computing Specs: Current version written in Python using Keras with Tensorflow backend. Darkchem 2 in developing using Google JAX

IP Availability: IP License; Collaborative R&D; Sponsored Research

Technology Maturity Level

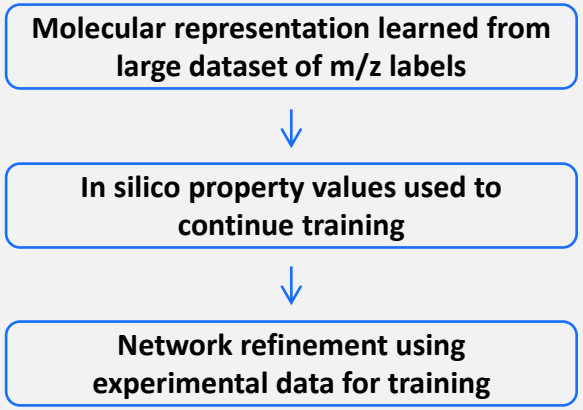
TRL 6: Software tools demonstrated in a relevant environment

Technology Description

US2020176087A1

SW tools for state-of-the-art first-principles simulation, distinguished by use of molecular dynamics, quantum chemistry, and ion mobility calculations to generate predictions of chemicals that may be in the sample or to generate novel molecular structures. They rely on a variational autoencoder to learn a continuous representation of molecular structure

Model Training Workflow



Key Benefits

- **High-Power Computing:** Leverages Variational Autoencoder (form of generative neural network) that has been adapted for metabolomics and small molecule identification
- **Highly Accurate:** Can predict CCS (key feature of molecular structure) to within 2.5%
- **Novel Molecule Discovery:** Unlike traditional approaches, Darkchem can generalize to novel molecules outside of the chemical classes represented by a training set
- **Extensible Network:** Network can be extended to other training data / molecular representations, AND for use with other analytical platforms

tradespace

Engagement Opportunities

PNNL has engaged Tradespace to approach select partners to commercialize DarkChem - a set of software tools for small molecule identification and discovery

Potential collaboration models include:

- Patent License
- Cooperative Research and Development Agreements (CRADAs)
- Sponsored Research
- Collaborative Research

Please contact Alec Sorensen, CEO of Tradespace, for further information regarding the opportunity

Contact:
Alec Sorensen
CEO Tradespace
alec@tradespace.io
804-836-7938

tradespace

Appendix

Case Study:
NMDA Receptor Antagonist Molecule Discovery

Pacific Northwest NATIONAL LABORATORY

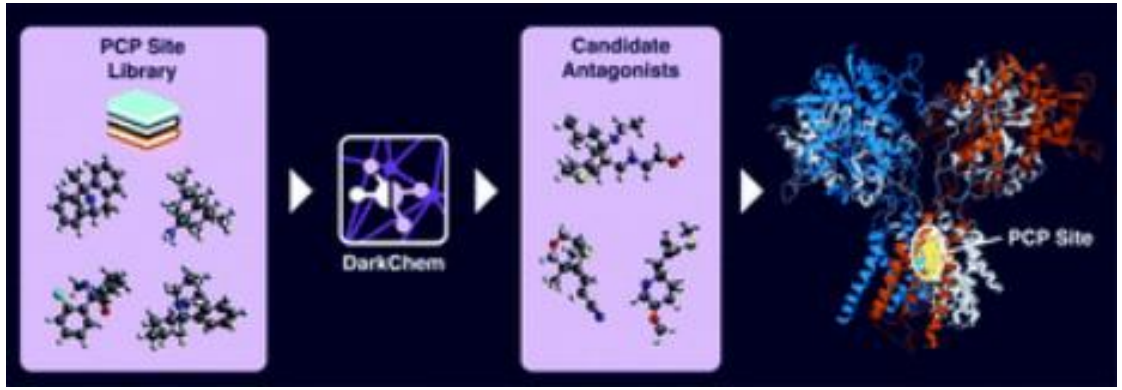
Desired Property: Uncompetitive antagonists of the N-methyl D-aspartate receptor (NMDAR) have therapeutic benefit in the treatment of neurological diseases such as Parkinson's and Alzheimer's

Objectives

- Creation and release of comprehensive library of experimentally validated NMDAR phencyclidine (PCP) site antagonists

Outcomes

- Generation of twelve candidate antagonists that are not available in existing chemical databases



tradespace

Darkchem Training Datasets

Pacific Northwest NATIONAL LABORATORY

- **PubChem:** Used to pretrain the variational autoencoder on a large set of SMILES strings (N=53,335,670) with calculated m/z

- **In Silico Dataset:** The union of the Human Metabolome Database (HMDB), the Universal Natural Products Database (UNPD), and the Distributed Structure-Searchable Toxicity (DSSTox) database with in silico predicted CCS values
  - The in silico data set additionally included CCS, calculated using ISiCLE. The in silico dataset is a larger (N=608,691) proxy to actual experimental CCS values (N=403, 486, and 371 for [M+H]+, [M−H]−, and [M+Na]+ adducts, respectively; a combined 756 unique parent molecules)

- **Curated library of molecules with experimental CCS values (metabolomics.pnnl.gov), which span a representative subset of known chemical space**

tradespace