JSTO DTRA.mil in the News

August 2024 | Vol. 14 No. 4

SPECIAL EDITION ISSUE

Research

DTRA MISSION

DTRA provides cross-cutting solutions to enable the Department of Defense, the United States Government, and international partners to Deter strategic attack against the United States and its allies; Prevent, reduce, and counter Weapons of Mass Destruction (WMD) and emerging threats; and Prevail against WMD-armed adversaries in crisis and conflict.

CHEMICAL AND BIOLOGICAL TECHNOLOGIES DEPARTMENT MISSION

Lead DoD science and technology to enable the Joint Force, nation, and our allies to anticipate, safeguard, and defend against chemical and biological threats.

Inside cover: A research scientist with the Naval Medical Research Command's Biological Defense Research Directorate observes trends during an antibody purification run on the Next Generation Chromatography System. (U.S. Navy photo by Mike Wilson)

Back cover: A research scientist with the Naval Medical Research Command's Biological Defense Research Directorate performs tests of a recent lateral flow immunoassay production. (U.S. Navy photo by Mike Wilson)

DEFENSE THREAT REDUCTION AGENCY

Research and Development Directorate
Chemical and Biological Technologies Department
Joint Science and Technology Office
for Chemical and Biological Defense

8725 John J. Kingman Road, Stop 6201, Fort Belvoir, VA 22060

www.dtra.mil

Find us on social media: @doddtra

SPECIAL EDITION ISSUE

Basic Research IN THE Spotlight

Basic research continues to play a critical role in creating the foundational knowledge necessary to advance disruptive capabilities to detect, deter, and defeat chemical and biological threats for the Joint Force, our nation, and allies.

This issue highlights four major accomplishments in JSTO's artificial intelligence/machine learning (AI/ML) portfolio:

- Accelerating drug discovery of antimicrobial peptides
- Advancing an automated, user-friendly AI/ML platform intended for life science researchers
- Creating computational molecular-docking approaches to elucidate drug/target protein binding leading to accelerated development of more effective medical countermeasures
- Organizing a data science workshop to inform creating a reliable metadata framework for supporting DTRA JSTO's AI/ML development capabilities

The final article highlights recent advances made in developing an ultrasensitive, hand-held detection device based on earlier basic research involving integrating electronics into biology.

From Years to Hours: Accelerating Drug Discovery with Advanced ML Techniques

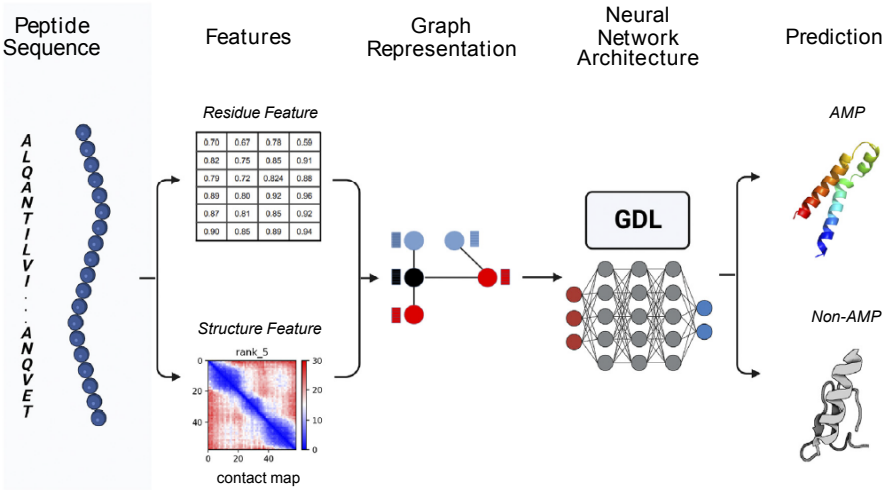
To help remedy the cumbersome, time-consuming, and expense of creating new drugs against biological pathogens, research into artificial intelligence (AI), machine learning (ML), and deep learning (DL) technologies have allowed scientists to focus on rapid drug design and development by combining DL with the principles of geometry to reduce failure rates, length of schedule, and cost-associated risks of drug development.

It is increasingly difficult to develop medical countermeasures (MCMs) against the biological pathogens available to our adversaries. To deter these biological pathogens against the Joint Force and maintain a scientific and technological edge, the Defense Threat Reduction Agency's (DTRA) Chemical and Biological Technologies Department in its role as the Joint Science and Technology Office (JSTO) for Chemical and Biological Defense, an integral component of the Chemical and Biological Defense Program, invested in basic research with the University of Pennsylvania (UPenn) to rapidly create MCMs using aspects of machine learning.

The traditional drug discovery, design, and development process uses taxing laboratory tests and experiments to evaluate the safety and effectiveness of new drugs. Those with the desired traits are further evaluated in multiple clinical studies before achieving regulatory approval.

To help speed up the process, researchers at UPenn developed a model system that uses DL to identify new antimicrobial peptides (AMPs), which are a class of natural peptides that exhibit antimicrobial activity characterized by broad-spectrum actions including antibacterial, antifungal, antiviral, and antiparasitic effects. In addition, AMPs have low molecular masses and often possess high antimicrobial, antibiofilm, and anti-inflammatory properties.

The UPenn researchers used several DL methods to predict, identify, design, and develop AMPs that offer protection against bacterial pathogens which cause human diseases. Even though they were promising while needing further development, these first-generation computational tools did not include the precise 3D properties of AMPs



General rational pipeline for antimicrobial peptide (AMP) prediction using GDL: From an initial AMP amino acid sequence, the relevant physicochemical characteristics are extracted, and the 3D structure of the sequence is predicted. Once the sequence and spatial relationships are obtained, they are converted into graphs in which the structural information is represented by the edges, while the amino acid residue information is represented by the nodes. The graph-based data is presented to a GDL network to predict whether the candidate is likely to have antimicrobial activity. (UPenn image created with BioRender.com)

. . . UPenn researchers harnessed Geometric Deep Learning (GDL) that combines DL with the principles of 3D geometry to predict AMPs, which possess high degrees of target specificity and affinity . . .

to improve the accuracy of peptide binding properties, target affinity, and drug-like efficacy.

To overcome this hurdle, the UPenn researchers harnessed Geometric Deep Learning (GDL) that combines DL with the principles of 3D geometry to predict AMPs, which possess high degrees of target specificity and affinity, which translates into greater antimicrobial potency. Also, GDL can capture the complex relationships between geometric properties and function of these molecules, as it more adequately represents the 3D non-Euclidean nature of peptide structures than traditional machine learning methods that operate in the 2D Euclidean spaces.

In contrast to traditional DL methods, GDL considers the fundamental structural features in three ways:

1. Manifolds, used to represent the conformational space of AMPs, which is the space of all possible 3D structures that an AMP can adopt
2. Sequences, used to represent the amino acid sequences of AMPs that dictate their structural and functional properties
3. Graphs representing the interactions between AMPs and their target molecules

This comprehensive approach enables GDL to identify the key features essential for antimicrobial activity and is an efficient DL approach for drug design, development, and biological molecule prediction. Candidates identified through GDL that exhibit the desired drug-like characteristics against a wide array of biological pathogens hold promise as effective MCMs, and GDL represents a rapid method to defeat the ever-evolving biological pathogens facing the Joint Force. ●

AI/ML

The general meaning of AI, in the context of scientific research, is the ability of a digital computer to think, learn, and execute simple and complex tasks, and the ability to learn from data and improve performance over time without specific programming, usually associated with human directions. ML, a subfield of AI, is a powerful tool that is trained on large data sets to identify specific patterns resulting in an improved performance of specific tasks. DL is a subfield of ML that uses so called "artificial neural networks" to process complex data. In this regard it mimics the interconnection of the neurons in a brain. It can learn from a huge amount of data and can capture fundamental relationships between different data types.

Non-Euclidean Geometry

In mathematics and computer science, non-Euclidean geometry focuses on the characteristics of spherical and hyperbolic objects and their associated curvature properties, such as angles, coordinates, points, and surfaces. In contrast, Euclidean geometry focuses on planar surfaces with no curvature. Computational tools and ML algorithms for predicting the 3D properties of AMPs must include all these non-Euclidean factors to account for the molecular structural curvature due to the peptide bonds formation, spatial angles of individual amino acids, and other molecular characteristics.

To Learn More

Frontiers in Bioinformatics, July 2023, Vol. 3, "Geometric deep learning as a potential tool for antimicrobial peptide prediction."

Click here to read the article

BioAutoMATED: An End-to-End Machine Learning Tool for Biomedical Science and Drug Development

It is increasingly challenging to rapidly develop medical countermeasures (MCMs) against the expansive growth of chemical and biological (CB) threats to the Joint Force. Drug discovery and development are time-consuming and do not always provide the expected results. Basic research into developing automated machine learning (AutoML) specifically for biological sequence data created a promising tool for these challenges: an AutoML framework called BioAutoMATED to analyze biological sequences, interpret biological sequence data, and offer functionalities for designing new biological sequences with the desired properties.

The Defense Threat Reduction Agency's (DTRA) Chemical and Biological Technologies Department in its role as the Joint Science and Technology Office (JSTO) for Chemical and Biological Defense, an integral component of the Chemical and Biological Defense Program, is investing with Massachusetts Institute of Technology (MIT) researchers toward the development of rapid MCMs for CB threats using BioAutoMATED.

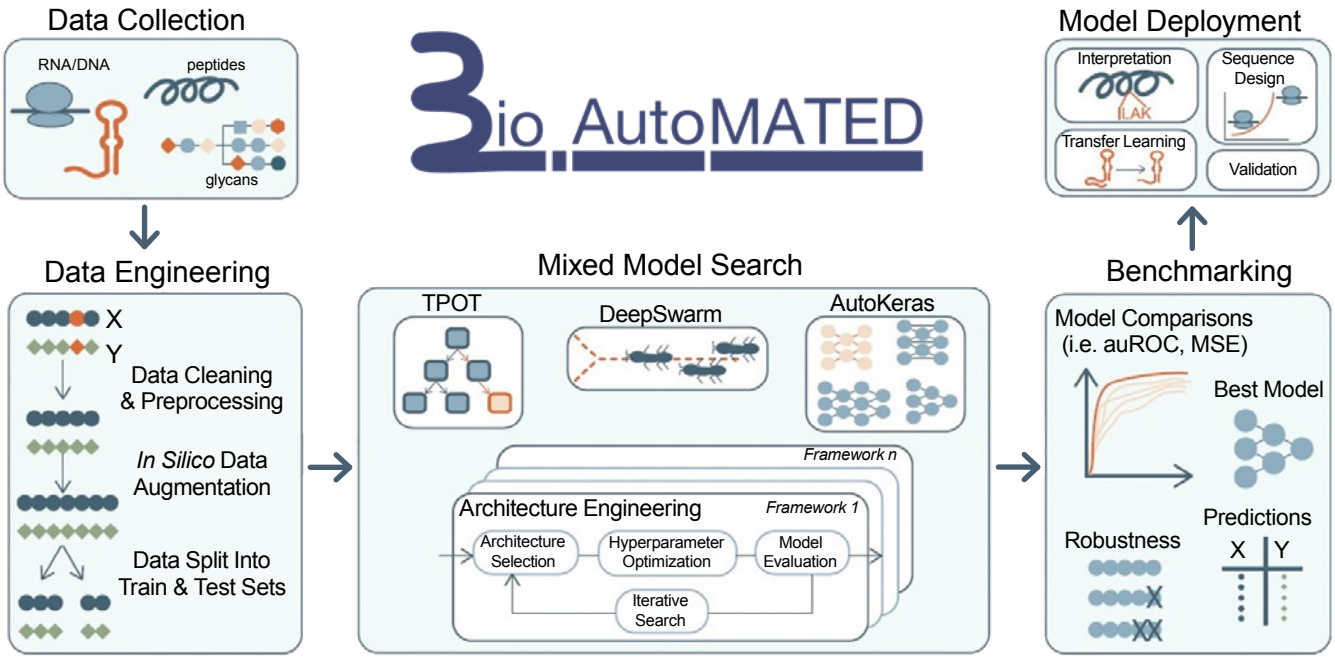
The rapid growth of biological datasets has created opportunities for using ML as a new tool in biomedical research. These datasets often contain complex sequences, like DNA or protein sequences, and analyzing them with ML provides valuable insights and accelerates the development of new biological tools and treatments. Building, training, and deploying ML models requires expertise and involves many critical decisions that affect the outcome. For life science researchers with limited computational biological experience in ML, AutoML can handle tasks like data-preprocessing, selecting the right type of ML model, tuning the model parameters, and evaluating the model's performance. Despite the availability of various AutoML tools, few are optimized for biological sequence data.

For life science researchers with limited computational biological experience in ML, AutoML can handle tasks like data-preprocessing, selecting the right type of ML model, tuning the model parameters, and evaluating the model's performance.

To Learn More

Cell Systems, June 2023, Vol. 14, No. 6, "BioAutoMATED: An end-to-end automated machine learning tool for explanation and design of biological sequences."

Click here to read the article



The BioAutoMATED framework automates the identification of predictive ML architectures for any set of nucleic acid, peptide, or glycan sequences. The integrated pipeline includes standard data processing, a mixed model search via three open-source AutoML libraries, and general benchmarking of models. Trained models are automatically deployed for interpretation and sequence design. BioAutoMATED also includes features to externally validate models with additional datasets and retrain models on new data using transfer learning. (MIT image)

MIT researchers developed BioAutoMATED to be specifically designed for analyzing biological sequences and optimized for building models with nucleic acid, peptide, and glycan sequence inputs. For instance, it can identify potential regulatory sequences that are involved in gene regulation and analyze sequence features of peptides that are relevant for peptide-drug interactions. BioAutoMATED can also classify glycans (sugars, for example found on cell membrane or bacteria cell wall surfaces) based on their sequence and predict their immunogenicity in humans, which could influence the development of vaccines and therapeutics. BioAutoMATED also has the capability to design new biological sequences with specific functions to guide the development of new and more effective drugs.

BioAutoMATED provides not just predictions but also insights into the underlying biological sequence features driving those predictions. A key feature of BioAutoMATED is interpretability by offering tools to analyze the best models automatically. These tools can assist in identifying specific locations and patterns within sequences that contribute most to the model's predictions. This transparency allows researchers to gain a deeper understanding of their data and to build more robust models.

Although BioAutoMATED is not a complete substitute for human expertise in ML, it is useful as a starting point in merging computational results with experimental work, which not only simplifies the integration of ML into accelerated drug discovery against new biological threats, but also presents new possibilities for scientists in drug development who lack extensive computational biology techniques.

Beyond its role in data analysis and synthetic biology, BioAutoMATED allows for accessible and interpretable ML applications and is a valuable tool in the field of drug discovery and drug development for emerging threats against the Joint Force. ●

On Demand Solutions: How Innovations in Artificial Intelligence and Data Science are Delivering Biological Treatments

An innovative deep learning model called DIFFDOCK offers a promising new approach to accelerate drug discovery and develop medical countermeasures (MCMs) to protect the Joint Force against new and emerging biological threats.

DIFFDOCK is a new molecular docking approach model for drug discovery, which has led to an even more advanced method—DIFFDOCK-L—that uses increased data-generating techniques and a larger model size with higher accuracy in predicting ligand binding poses. Overall, DIFFDOCK and DIFFDOCK-L represent a significant step forward in molecular docking, offering better accuracy, efficiency, and flexibility compared to traditional search-based molecular docking methods.

The Defense Threat Reduction Agency's (DTRA) Chemical and Biological Technologies Department in its role as the Joint Science and Technology Office (JSTO) for Chemical and Biological Defense, an integral component of the Chemical and Biological Defense Program, is investing in basic research at the Massachusetts Institute of Technology (MIT) that integrates computer-aided drug development (CADD) techniques, such as computational modeling, complex algorithms, computer software, and molecular docking approaches, to enhance the rapid development of effective MCMs.

CADD approaches bypass the slow, traditional drug discovery and development process by shortening the timeline to improve existing or develop new MCMs against biological pathogens. Molecular docking predicts how well a drug molecule (ligand) will fit and bind to its target protein (e.g., protein of a pathogen). This is crucial in drug discovery as it shows key interactions between a drug and its target protein, revealing the molecular basis of its activity.

Molecular Docking

Molecular docking is a computational method in drug discovery to predict the preferred orientation, conformation, and binding affinity of two interacting molecules, typically ligand (a drug candidate) and target (a biological protein). In addition, molecular docking can guide optimizing the ligand's structure to improve its binding properties to the protein. Molecular docking plays a crucial role in identifying potential drug candidates and is often used with experimental techniques to lead the design and optimization of new drugs.

DIFFDOCK & DIFFDOCK-L

Unlike traditional molecular docking methods, which often use a search-based and scoring function approach, DIFFDOCK is a new approach to molecular docking that uses diffusion generative modeling, aimed at to learn the entire range of possible poses (positions, orientations, and conformations) that a ligand might adopt when bound to a target protein. DIFFDOCK provides faster and more precise computer-aided design of small molecules and will contribute to the discovery of new promising candidates against biological threats.

To Learn More

Conference Paper at ICLR 2023, "DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking."

Click here to read the article

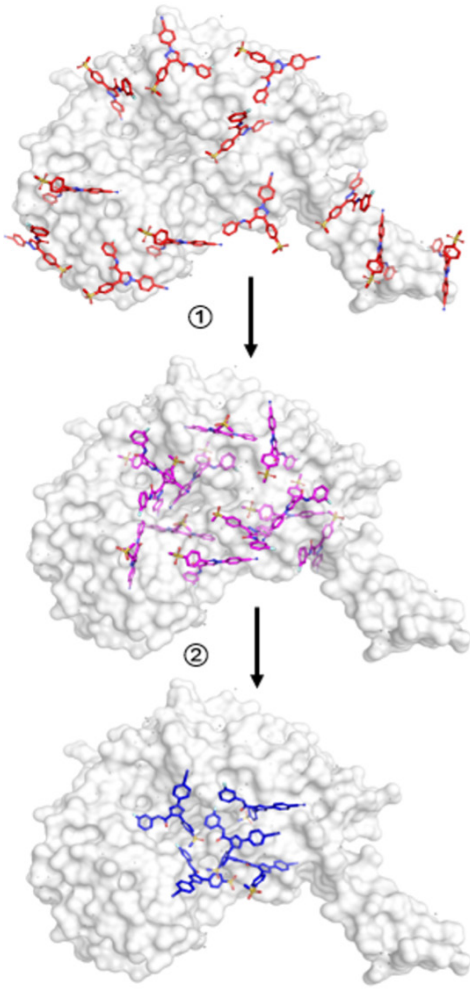
. . . DIFFDOCK and DIFFDOCK-L learn the entire range of possible poses that a ligand might adopt when bound to a protein rather than predict a single best one . . .

Traditional molecular docking methods predict the optimal binding pose (orientation) of a ligand and estimate its binding affinity with the protein. These methods involve searching through many possible ligands poses until the best one is found. However, search-based methods can be expensive and difficult to scale to large datasets, plus the accuracy can be limited by the scoring function and search algorithm.

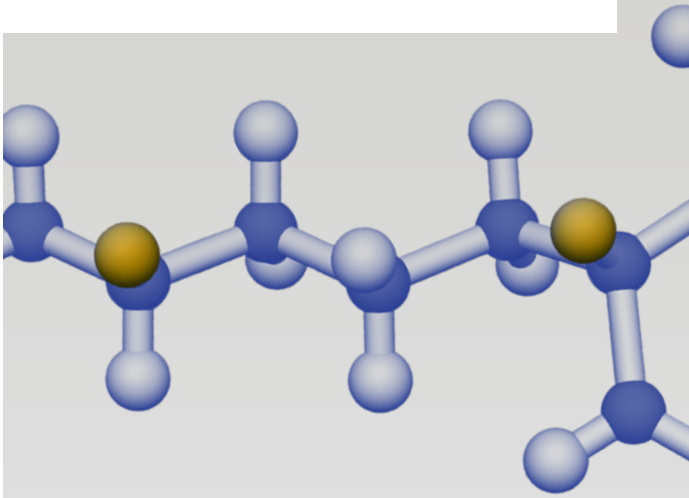
Through DTRA JSTO's investments in basic research, researchers at MIT's Computer Science & Artificial Intelligence Laboratory generated this new approach that differs from previous regression-based frameworks to use a generative modeling approach that is better aligned with the objective of molecular docking. At the heart of its modelling strategy, DIFFDOCK and DIFFDOCK-L learn the entire range of possible poses that a ligand might adopt when bound to a protein rather than predict a single best one through a process that refines random poses according to their compatibility with the protein structure.

By using a learned score, DIFFDOCK guides the refinement process to ensure compatibility with the protein. Over time, the random initial pose of the ligand transforms into a pose that better fits the protein. This diffusion process allows DIFFDOCK and the improved and advanced DIFFDOCK-L to explore a wider range of possibilities and ultimately achieve high accuracy in predicting ligand poses and allow them to generate new poses that a ligand might adopt when bound to the protein of a pathogen. The ability to learn from a distribution of the ligand's poses and the ability to guide the refinement process using a learned score allows DIFFDOCK and DIFFDOCK-L to explore a wide range of possibilities and achieve high accuracy in predicting ligand poses and providing additional options for new drugs against biological threats.

Although the choice between traditional search-based molecular docking approaches and DIFFDOCK or DIFFDOCK-L depends on the specific research question, computational resources, and available data, DIFFDOCK and DIFFDOCK-L offer a promising new approach that can accelerate drug discovery and help to develop new treatments to protect the Joint Force against biological pathogens. ●



DIFFDOCK employs a diffusion process to iteratively refine a ligand's position within a protein binding site. By exploring various ligand orientations and conformations within the binding site, a variety of optimal ligand poses can be identified. (DTRA JSTO image)



Data Science for Chemical and Biological Defense

Chemical and biological (CB) data is a strategic asset that can be used to inform decision making within the Department of Defense's (DoD) Chemical and Biological Defense Program (CBDP). However, reliably applying CB datasets to rapidly develop data-informed decision-support capabilities that allow the Joint Force to understand and mitigate CB threats remains a challenge.

Data science as a strategic asset has gained increased traction across the government with the release of policy, including Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (October 30, 2023) and the DoD Data, Analytics, and Artificial Intelligence Adoption Strategy (June 27, 2023). Therefore, the Defense Threat Reduction Agency's (DTRA) Chemical and Biological Technologies Department in its role as the Joint Science and Technology Office (JSTO) for Chemical and Biological Defense, an integral component of the CBDP, has invested in leveraging artificial intelligence and machine learning (AI/ML) to explore and characterize CB threat data. Among a myriad of other use cases, some of the data will be used to more accurately predict new biomarkers of CB exposure for early detection and diagnostics and to optimize developing new medical countermeasures.

While these initial investments in AI/ML have enabled better understanding of the opportunity areas, they have also illuminated some of the fundamental challenges associated with applying AI/ML technologies to CB data as well as larger fundamental data science challenges across technology areas. Although there is no one-size-fits-all approach, creating a strong data science framework for CB data within DTRA JSTO is a key component of a successful CBDP-wide data science enterprise.

There is an urgent need for DTRA JSTO, along with the broader data science community, to establish a standardized framework to ensure that reliable, reproducible, and re-usable data is available to develop capabilities to treat and protect the Joint Force.

To better understand the current data science landscape, DTRA JSTO held a Data Science Workshop that included subject matter experts in both computational and CB science fields from across the government, national labs, federally funded research and development centers, and academia.

The overall objectives of this workshop were to:

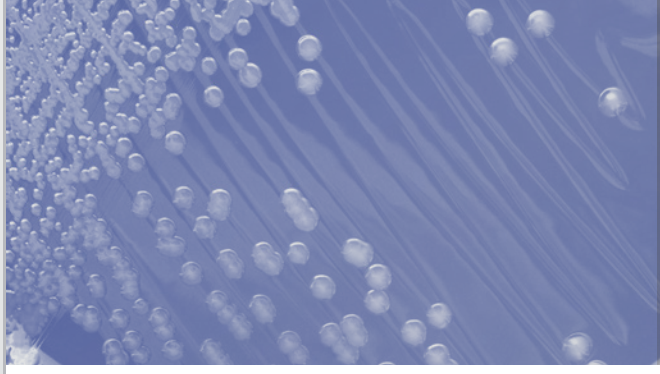
- Explore new, state-of-the-art technology areas within the field of data science
- Highlight gaps and challenges remaining in this evolving, rapidly changing data science field
- Build an informed path for FY25 DTRA JSTO basic research topics in data science

The workshop included presentations by invited speakers and group discussions. Four speaker presentations covered data standardization, provenance, harmonization, integration, data management and security, as well as community sharing of data. The workshop focused on discussing the regulatory requirements for the quality of data and the challenges faced with the compliance guidelines for rapidly developing medical countermeasures.

There was a shared understanding among participants that the complexity and variability of CB data types is extensive and the current frameworks for data formatting are too varied to harness it for maximum benefit and value across DTRA JSTO. There is an urgent need for DTRA JSTO, along with the broader data science community, to establish a standardized framework to ensure that reliable, reproducible, and re-usable data is available to develop capabilities to treat and protect the Joint Force. Ideally, standardizing specific data and datasets would allow it to be more efficiently validated, analyzed, and re-used.

Other recommendations from the workshop included employing incentive structures and acquisition vehicles to facilitate re-using and sharing data among the CB community and promoting data standards to build trustworthy partnerships between the DoD and industry. Developing a standardized framework for DTRA JSTO's current efforts will help build a roadmap to inform future research and development investments in data science to produce technology that protects the Joint Force, nation, and our allies. ●

POC: Sweta Batni, Ph.D., sweta.r.batni.civ@mail.mil



Basic Research Wins by a Nose: Unleashing the Power of Integrating Microelectronics into Biology

Basic research from 2014 has led to an effort to create a biological-to-electronic platform that rapidly senses information using bioengineered canine olfactory (smell) receptors and transferring it into actionable electronic forms—in other words, an "electronic nose."

The Defense Threat Reduction Agency's (DTRA) Chemical and Biological Technologies Department in its role as the Joint Science and Technology Office (JSTO) for Chemical and Biological Defense, an integral component of the Chemical and Biological Defense Program, invested with the University of Maryland (UMd) to envision live, "rewired" bacteria that independently survey warfighters (for example, their gastrointestinal tracts), hone in on disease-causing agents, and synthesize and deliver therapeutics for effective treatment.

The initial studies focused on engineering bacteria to sense and kill pathogens, or swim to cancer cells and activate gene expression processes that alter gene information and ensuing cellular biochemistry to initiate cell death. The goal was to design and build new bacteria combining the tools of synthetic biology with biofabrication using nature's existing ability of self-locomotion and gene expression to create bacteria programmed to stand on guard for harmful bacteria and viruses. Biofabrication extended the reach of synthetic biology by providing means for integrating biological constructs into microelectronic-fabricated systems to test their function and provide for electronic control.

The goal was to design and build new bacteria combining the tools of synthetic biology with biofabrication using nature's existing ability of self-locomotion and gene expression to create bacteria programmed to stand on guard for harmful bacteria and viruses.

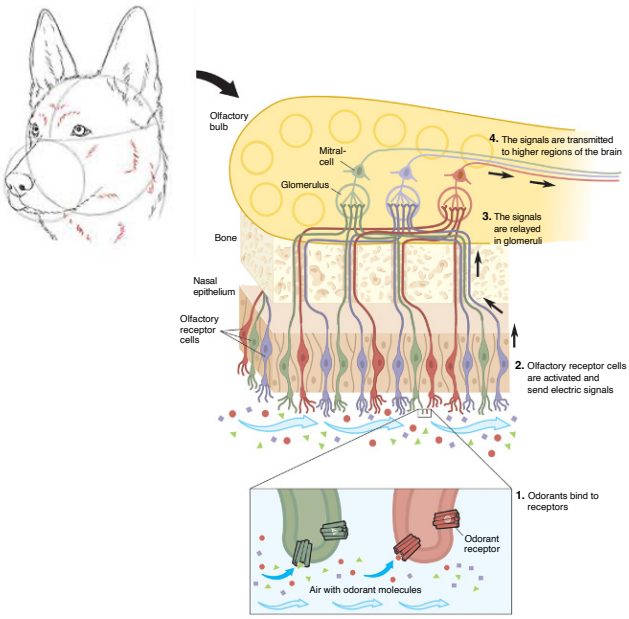
The DTRA JSTO program catalyzed a new field that integrates electronics with biology and the direct transfer of information between two mediums of information flow in electronics through the electron and in biology through the molecule that typically do not intersect. The researchers found they could convey information that is rich with content, rapid, easily accessed, and bidirectional.

They applied their discovery in several areas:

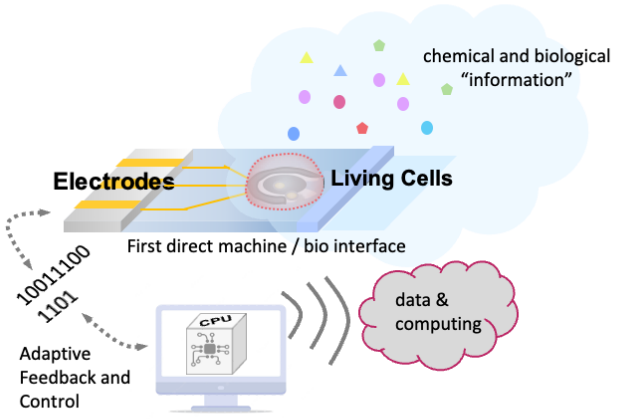
- Reprogramming microbes as a biosensor for *Staphylococcus aureus* (Staph) detection or preventing *Pseudomonas aeruginosa* gut infection or colorectal cancer
- Assessing melanin's ability to scavenge oxygen radicals and its potential links to mental health
- Quantifying antibody quality
- Providing a new way of measuring oxidative stress that correlated with progression of disease

An endeavor is now underway to tackle one of the greatest challenges in sensing: bioengineering a dog's olfaction system to create Living Microelectronics—A Smarter Nose. A joint effort between the U.S. Army Research Laboratory and University of Maryland seeks to construct a benchtop prototype device that will be transitioned into a working hand-held unit designed to function in the battlefield. ●

"Living" Microelectronic Materials – An Electronic Nose: The dog's nose has electronic access to the chemical or biological information transmitted to the dog's brain. This "living" microelectronics system is the first integration of a living cell with a machine interface. It will provide both the sensitive and specific identification of odors and the chemical and biological information derived from those odors. (UMd image)



Adapted from Richard Axel and Linda Buck Nobel Lecture, 2004



CBD S+T

CONFERENCE 2024

FOCUS / FORWARD

2024 Chemical and Biological Defense Science & Technology Conference
December 2-5, 2024
Fort Lauderdale, Florida

- @cbdstconference
- @dtrajstoengage
- @DoDDTRA
- CBD S&T Conference
- DTRA JSTO Engagement

Join thought leaders and experts from around the globe who are committed to making the world safer by confronting chem-bio defense challenges.

Register today at

CBDSTCONFERENCE.COM

Registration deadline: November 8, 2024

Approved for public release; distribution is unlimited

Within the Defense Threat Reduction Agency's Research and Development Directorate resides the Chemical and Biological Technologies Department performing the role of Joint Science and Technology Office for Chemical and Biological Defense, an integral component of the Chemical and Biological Defense Program. This publication highlights the department's advancements in protecting the Joint Force, our nation, and allies from chemical and biological threats through the innovative application of science and technology.

FOR MORE INFORMATION VISIT ONLINE

DTRA.MIL