THE USE OF AI AND ADVANCED COMPUTER SYSTEMS TO DEVELOP DRUGS AGAINST NEW EMERGING THREATS

Discovery Of Novel Small-molecule Compounds Targeting The Immune Homeostasis Protein TIPE2 Using Machine Learning And High Performance Computing

Bradley Kearney National Defense Medical College / US Army Japan ESEP **Hiroyuki Nakashima** National Defense Medical College **Masahiro Nakashima** National Defense Medical College **Takeshi Ono** National Defense Medical College **Azusa Katou** National Defense Medical College **Kohei Yamada** National Defense Medical College **Ryohei Suematsu** National Defense Medical College **Hiroyasu Goto** National Defense Medical College **Keiko Tanoue** National Defense Medical College **Manabu Kinoshita** National Defense Medical College

Sepsis is a medical condition where the body overreacts to an infection resulting in tissue damage and organ failure and is a leading cause of death in hospitals. Additionally, casualty projections suggest that sepsis rates will increase during large-scale combat operations due to prolonged field care. Immune homeostasis is a critical property of the immune system that balances responding to pathogens while minimizing damage to healthy tissue. Sepsis develops when immune homeostasis is impaired, resulting in a systemic hyperinflammatory state. While sepsis remains a global health concern, no effective therapy currently exists.

The tumor necrosis factor-alpha (TNF-α)-induced protein 8 (TNFAIP8/TIPE) family, including the novel TNFAIP8 like-protein 2 (TNFAIP8L2/TIPE2), plays a vital role in regulating inflammatory responses, immune homeostasis, and cancer development. Previous studies show that knockout of TIPE2 protein leads to accelerated and exacerbated immune response to endotoxin which mimics sepsis without actual infection. Likewise, inducing TIPE2 overexpression through adenovirus vectors results in a decrease in inflammation and sepsis-induced tissue damage. These results make TIPE2 a lead target for therapeutics to reduce septic morbidity by enhanced maintenance of immune homeostasis. However, there are two structurally homologous proteins in the TIPE family that perform other vital roles in cells. Therefore, any drugs targets of TIPE2 must be screened against other TIPE-family proteins to minimize the chance of off-target effects.

Computer-aided drug discovery (CADD) is a valuable tool that complements the expertise of medicinal chemists and structural biologists. There are between 1020 to 1023 synthesizable drug-like molecules based on current methods. CADD enables in silico experiments at scales much larger than are possible in vitro and in vivo and can identify candidate ligands that warrant further testing.

Here we present a computational drug design pipeline that incorporates traditional molecular docking for lead identification with machine learning (ML) tools for both ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) property prediction and lead optimization.
Using the atomic models for TIPE1, TIPE2, and TIPE3, we screened 301,003 small molecule compounds from three public databases using traditional docking software tools that were optimized to run on high performance computing resources. We identified the lead compounds that showed high binding affinity for TIPE2 and compared to the binding affinities for TIPE1 and TIPE3 to reduce the chance of off-target binding. We then took the lead compounds for TIPE2, predicted ADMET properties, and eliminated any compounds that showed poor ADMET scores. The remaining lead compounds were then optimized using a genetic algorithm to evolve predicted ligands on demand.

We show that this pipeline can generate unique lead compounds that target an immune homeostasis protein while minimizing off-target binding to two structurally homologous proteins. By using an in silico methodology, we have rapidly identified and optimized lead compounds. This pipeline, which utilizes traditional docking with machine learning methods for ADMET prediction and lead optimization, can be used for other protein targets and may lead to more effective and efficient vaccine and therapeutic development.

This work used Expanse at the San Diego Supercomputer Center through allocation BIO220173 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.