Presentation no. P172
ECC-2024

Advancing Vaccine Development through Precise AI-driven Prediction of Protective Antigens

EVAXION AI-Immunology™ Powered Vaccines

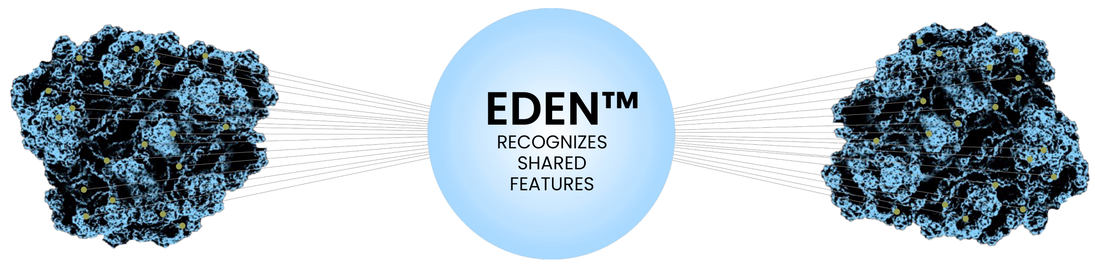
Christian Garde, Michael Schantz Klausen, Rasmus Villebro, Nikolas Thuesen, Jakob Guldberg Aes, Gry Persson, Sophie Schussek, Andreas Holm Mattsson, Thomas Trolle, Evaxion Biotech A/S, Hørsholm, Denmark

Abstract

Antimicrobial resistance (AMR) is on the rise and projected to cause up to 10 million deaths yearly by 2050 (Tang et al. 2023, Br J Biomed Sci). This underscores the urgent medical need for new and effective therapies, like antimicrobial vaccines, to tackle the resistant pathogens. Traditional vaccine development approaches, such as Reverse Vaccinology, is time-consuming and resource-intensive, highlighting the demand for a more efficient target discovery. To address this, we developed EDEN™, an AI model for fast and precise identification of broadly protective antigens. EDEN™ has proof-of-concept for bacterial vaccine development (Gulati et al. 2023, mBio) and is applied in ongoing vaccine programs targeting WHO prioritized AMR threats to global health. We here present the improved next generation of EDEN™ which leverages state-of-the-art deep learning methodologies. First, a deep transformer model was trained for comprehensive feature representation of protein sequences based on a dataset of more than 300,000 annotated proteins. This deep transformer was then applied to encode features for training the EDEN™ model to predict protective antigens. The training set of protective antigens was curated by screening scientific literature with Large Language Models using engineered prompts followed by manual confirmation by field experts. The resulting improved EDEN™ model achieved state-of-the-art performance in predicting protective antigens and can be used for efficient vaccine development to prevent AMR infections.

Highlights

1. LLM assists with identifying highly protective bacterial antigens
2. Models comprising Deep Transformers and GANs provides comprehensive protein and immunological features
3. EDEN™ predicts highly protective bacterial antigens with state-of-the-art predictive performance
4. EDEN™ predictions correlate with level of protection
5. Toxins constitute a distinct class of protective bacterial antigens
6. EDEN™ facilitate vaccine programs to address AMR infections



1. Large Language Models Facilitate Dataset Generation of Highly Protective Antigens

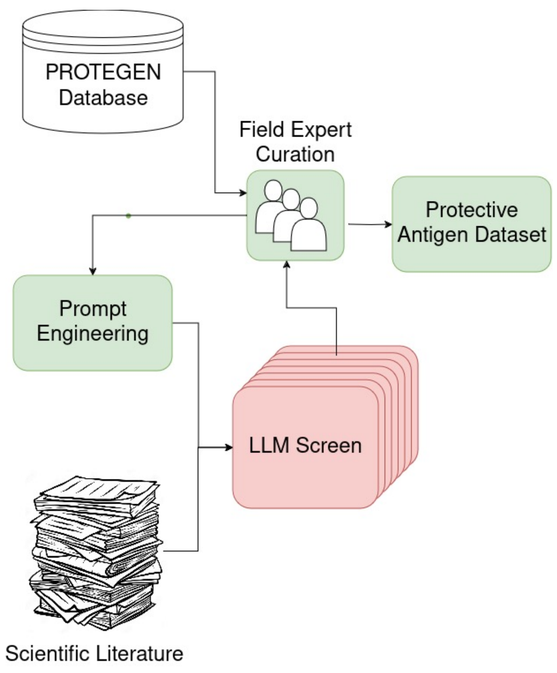


Figure 1. LLM assisted curation of protective bacterial antigens. Expert curation of protective antigens sourced from the PROTEGEN database was used to engineer prompts for LLM facilitated screening of >20.000 scientific papers and patents.. The literature labelled to describe protective antigens were subjected to field expert curation. Highly protective antigens were included in the final dataset, covering five bacterial species. The dataset was partitioned based on sequence similarity using K-means clustering. Negative data points was represented by the E. coli proteome and assigned the partition of the nearest positive neighbor.

3. EDEN™ Demonstrates State-of-the-art Performance with Predictions Correlates with Level of Protection

Figure 3.1 Challenge studies to assess efficacy of potential protective antigens. Potential protective antigens for two pathogens, *K. pneumonia* and *S. aureus*, were tested in murine lethal challenge models. Level of protection was assessed using Kaplan-Meier survival analyses. These studies establish a difficult test set comprising 7/26 *K. pneumoniae* and 11/38 S. aureus protective antigens (p-value < 0.05).

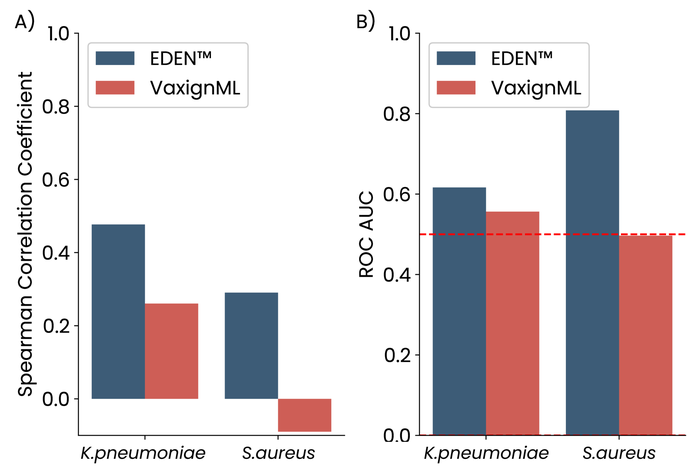


Figure 3.2 Performance evaluation on test set. (A) Spearman correlation analysis between level of protection and prediction scores from EDEN™ and the public gold standard VaxignML (Ong et al. 2020, Bioinformatics). (B) ROC analysis using a significance level 0.05 to the define positive and negative sets from the tested putative antigens.

2. Development of the EDEN™ Model

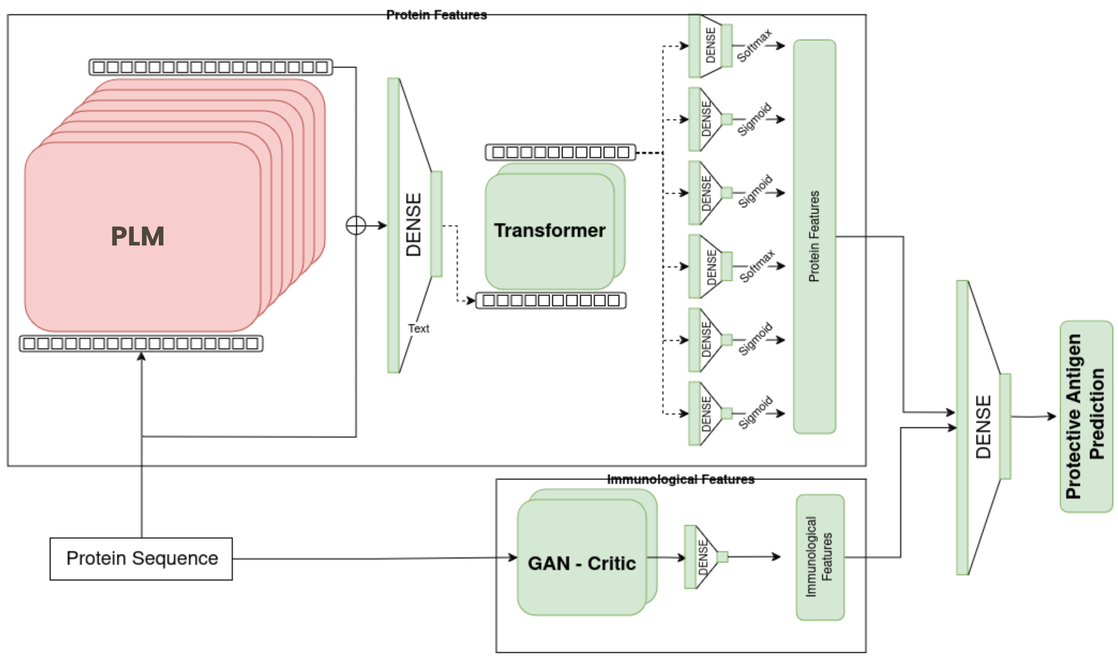


Figure 2.1 From protein sequence to antigen prediction. Protein sequences are transformed to features relating to surface exposure, immune recognition and others. The features are used as input to predict protective bacterial antigens.

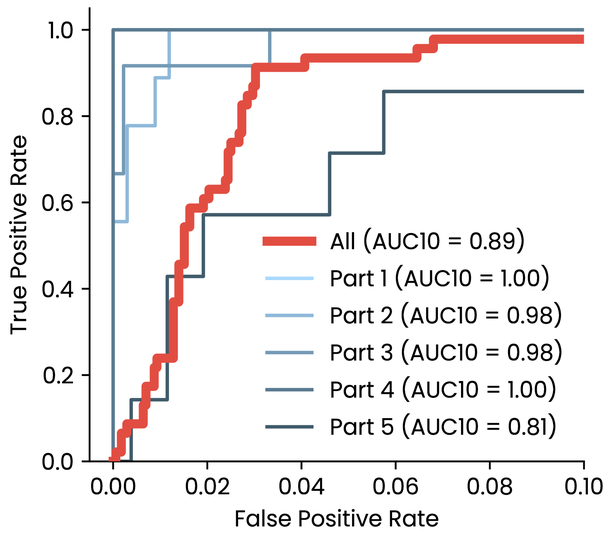


Figure 2.2 Receiver Operating Characteristics (ROC) for cross-validation analysis. ROC analysis demonstrating ability to generalize across distinct sequences. The focus was on the top 10% of the predictions to emphasize performance in the highest predicted set of proteins. AU-ROC values are reported for each CV partition and the full ensemble of partitions.

4. Toxins Constitute a Distinct Class of Highly Protective Antigens

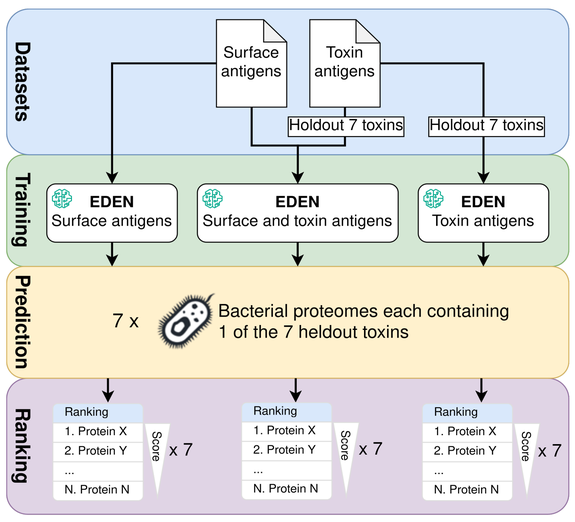


Figure 4.1 Toxin evaluation approach. Seven commercial protective toxin antigens and their associated proteomes were used as test set to evaluate ability to predict toxin antigens.

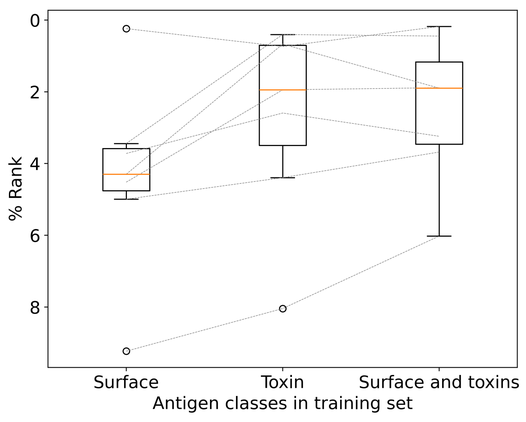
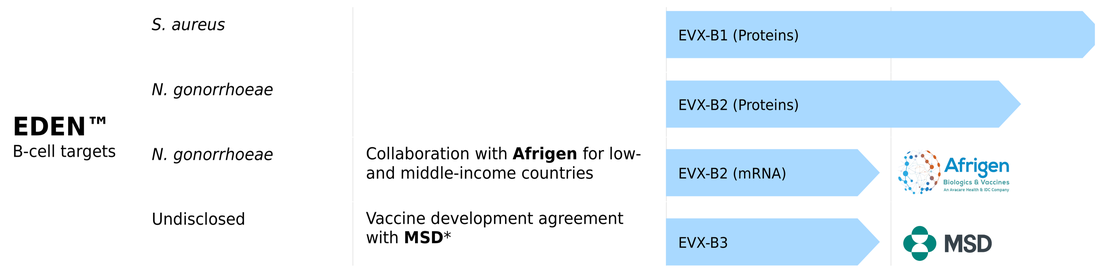


Figure 4.2 Prediction of protective toxin antigens. The predicted rank of the seven commercial protective toxins in their respective proteomes improves when EDEN™ is trained on datasets containing toxins (labelled along x-axis).

5. EDEN™ Facilitates Multiple Vaccine Programs to Address Bacterial Infections with High Medical Need



Contact

Christian Garde, PhD *Director of Bioinformatics* *cg@evaxion-biotech.com*