In silico methods or computer models

Bioinformatics is an accurate and robust tool in constant improvement to understand the basic principles of biology which can help to identify safe and novel drugs and vaccines. *In silico* simulations can predict possible toxic effects and/or efficacy of a potential drug/vaccine candidate without the use of animals. Also, computer predictions are less expensive and faster compared to animal models, since they allow for the mass screening of hundreds or thousands of candidates in parallel. This document gives insights into various resources available for computational identification and designing of new vaccine candidates.

Computer programs or tools:

- Computer Aided Drug Design (CADD) is used to predict the receptor binding site for a potential candidate molecule. Thus, CADD identifies probable binding site to avoid testing of drugs with non-biological activity. Also, by using such program new drugs and/or vaccine candidates for the specific binding site can be predicted and developed.
- Structure Activity Relationship (SARs) programs. This program predicts the biological activity of a drug candidate.
- Quantitative Structure Activity Relationship (QSARs) is the mathematical description of the relationship between physicochemical properties of a drug molecule and its biological activity.
- AlphaFold 2 is an AI-driven program that can predict the 3D structure of a protein from the DNA sequence. Although it is still in development, its promising results and exceptional accuracy can make it a key player among the *in silico* models in the future.
- VaxiJen, the first server to predict protective antigens without using alignment. It was created to allow antigen classification based solely on the physicochemical properties of proteins, specifically on auto-cross covariance (ACC) transformation of protein sequences into uniform vectors of major amino acid properties.
- ANTIGENpro uses two-stage architecture, multiple representations of the primary sequence, and five machine learning algorithms to predict the antigenicity of proteins. ANTIGENpro is a sequence-based, alignment-free and pathogen-independent predictor.
- AllergenFP and AllerTOP are the best allergen prediction tools for sequencing compared to the other analysis tools and servers.
- T- and B-cell epitope identification softwares are described and classified in an extensive review by Sánchez-Trincado et al. (Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA. Fundamentals and Methods for T- and B-Cell Epitope Prediction. J Immunol Res. 2017;2017:2680160. doi: 10.1155/2017/2680160. Epub 2017 Dec 28. PMID: 29445754; PMCID: PMC5763123.)
- The Immune Epitope Database and Analysis Resource (IEDB) is an unreservedly accessible asset that contains a broad assortment of tentatively estimated invulnerable epitopes and a setup of apparatuses for anticipating and dissecting epitopes.

This document was produced with the support of EC H2020 Project 951668: TRANSVAC-DS, and is a project coordinated by the European Vaccine Initiative (EVI).
UniversitätsKlinikum Heidelberg • Voßstraße 2, Geb. 4040 • 69115 Heidelberg, Germany
• e-mail: transvacinfo@euvaccine.eu

- A continuous B-cell epitope prediction (BCPred) method uses support vector machine (SVM) classifiers that were trained on a homology-reduced dataset of 701 linear B-cell epitopes recovered from the Bcipep database and 701 non-epitopes randomly retrieved from Swiss-Prot sequences using five different kernel approaches and five-fold cross-validation.
- The ABCpred server uses an artificial neural network to anticipate linear B-cell epitope areas in an antigen sequence. This server will aid in the identification of epitope regions that can be used to choose synthetic vaccine candidates, diagnose diseases and conduct allergy research.
- BepiPred is based on a random forest algorithm that is trained using epitopes from antibody–antigen protein structures. It is a new method based on known 3D structures and the large number of linear epitopes available from the IEDB database hence remains to out-perform compared to the other tools.
- LBtope is developed based on the experimentally validated B-cell epitopes and non-B cell epitopes from IEDB. Two types of datasets were derived as LBtope variable with 14,876 and 23,321 B-cell epitopes and non-epitopes of variable lengths, whereas LBtope fixed length has datasets with 12,063 B-cell epitopes and 20,589 non-epitopes of fixed lengths.
- The DiscoTope server uses three-dimensional (3D) protein structures to anticipate discontinuous B-cell epitopes. Surface accessibility (measured in terms of contact counts) and a unique epitope propensity amino acid score are used in the method.
- DiscoTope detects 15.5 percent of residues in discontinuous epitopes with a 95% specificity. The predictions can guide experimental epitope mapping in both rational vaccine design and the development of diagnostic tools, potentially leading to more efficient epitope identification.
- ElliPro predicts linear and discontinuous antibody epitopes based on the 3D structure of a protein antigen. ElliPro is based on the geometrical features of protein structure and requires no training. It could be used to predict many forms of protein–protein interactions.
- EpiPred is a program that predicts structural epitopes unique to a given antibody. Epitope predictions from EpiPred can be utilized to increase antibody–antigen docking performance. The approach can be utilized using an antibody homology model as input.

This document was produced with the support of EC H2020 Project 951668: TRANSVAC-DS, and is a project coordinated by the European Vaccine Initiative (EVI).
UniversitätsKlinikum Heidelberg • Voßstraße 2, Geb. 4040 • 69115 Heidelberg, Germany
• e-mail: transvacinfo@euvaccine.eu