Bioinformatics Tools for the Prediction of T-Cell Epitopes

T-cell responses are activated by specific peptides, called epitopes, presented on the cell surface by MHC molecules. Binding of peptides to the MHC is the most selective step in T-cell antigen presentation and therefore an essential factor in the selection of potential epitopes. Several in-vitro methods have been developed for the determination of peptide binding to MHC molecules, but these are all costly and time-consuming. In consequence, significant effort has been dedicated to the development of in-silico methods to model this event. Here, we describe two such tools, NetMHCcons and NetMHCIIpan, for the prediction of peptide binding to MHC class I and class II molecules, respectively, involved in the activation pathways of CD8+ and CD4+ T cells.
Recent advances in proteomics and mass-spectrometry have widely expanded the detectable peptide repertoire presented by major histocompatibility complex (MHC) molecules on the cell surface, collectively known as the immunopeptidome. Finely characterizing the immunopeptidome brings about important basic insights into the mechanisms of antigen presentation, but can also reveal promising targets for vaccine development and cancer immunotherapy. This report describes a number of practical and efficient approaches to analyze immunopeptidomics data, discussing the identification of meaningful sequence motifs in various scenarios and considering current limitations. Guidelines are provided for the filtering of false hits and contaminants, and to address the problem of motif deconvolution in cell lines expressing multiple MHC alleles, both for the MHC class I and class II systems. Finally, it is demonstrated how machine learning can be readily employed by non-expert users to generate accurate prediction models directly from mass-spectrometry eluted ligand data sets.
Identification of the cognate peptide-MHC target of T cell receptors using molecular modeling and force field scoring

Interactions of T cell receptors (TCR) to peptides in complex with MHC (p:MHC) are key features that mediate cellular immune responses. While MHC binding is required for a peptide to be presented to T cells, not all MHC binders are immunogenic. The interaction of a TCR to the p:MHC complex holds a key, but currently poorly comprehended, component for our understanding of this variation in the immunogenicity of MHC binding peptides. Here, we demonstrate that identification of the cognate target of a TCR from a set of p:MHC complexes to a high degree is achievable using simple force-field energy terms. Building a benchmark of TCR:p:MHC complexes where epitopes and non-epitopes are modelled using state-of-the-art molecular modelling tools, scoring p:MHC to a given TCR using force-fields, optimized in a cross-validation setup to evaluate TCR inter atomic interactions involved with each p:MHC, we demonstrate that this approach can successfully be used to distinguish between epitopes and non-epitopes. A detailed analysis of the performance of this force-field-based approach demonstrate that its predictive performance depend on the ability to both accurately predict the binding of the peptide to the MHC and model the TCR:p:MHC complex structure. In summary, we conclude that it is possible to identify the TCR cognate target among different candidate peptides by using a force-field based model, and believe this works could lay the foundation for future work within prediction of TCR:p:MHC interactions.
Improved methods for predicting peptide binding affinity to MHC class II molecules

Major histocompatibility complex class II (MHC-II) molecules are expressed on the surface of professional antigen presenting cells where they display peptides to T helper cells, which orchestrate the onset and outcome of many host immune responses. Understanding which peptides will be presented by the MHC-II molecule is therefore important for understanding the activation of T helper cells and can be used to identify T-cell epitopes. We here present updated versions of two MHC class II peptide binding affinity prediction methods, NetMHCII and NetMHCIIpan. These were constructed using an extended data set of quantitative MHC-peptide binding affinity data obtained from the Immune Epitope Database covering HLA-DR, HLA-DQ, HLA-DP and H-2 mouse molecules. We show that training with this extended data set improved the performance for peptide binding predictions for both methods. Both methods are publicly available at www.cbs.dtu.dk/services/NetMHCII-2.3 and www.cbs.dtu.dk/services/NetMHCIIpan-3.2. This article is protected by copyright. All rights reserved.
Peptide binding to MHC class I molecules is the single most selective step in antigen presentation and the strongest single correlate to peptide cellular immunogenicity. The cost of experimentally characterizing the rules of peptide presentation for a given MHC-I molecule is extensive, and predictors of peptide-MHC interactions constitute an attractive alternative.

Recently, an increasing amount of MHC presented peptides identified by mass spectrometry (MS ligands) has been published. Handling and interpretation of MS ligand data is, in general, challenging due to the polyspecificity nature of the data. We here outline a general pipeline for dealing with this challenge and accurately annotate ligands to the relevant MHC-I molecule they were eluted from by use of GibbsClustering and binding motif information inferred from in silico models. We illustrate the approach here in the context of MHC-I molecules (BoLA) of cattle. Next, we demonstrate how such annotated BoLA MS ligand data can readily be integrated with in vitro binding affinity data in a prediction model with very high and unprecedented performance for identification of BoLA-I restricted T-cell epitopes. The prediction model is freely available at http://www.cbs.dtu.dk/services/NetMHCpan/NetBoLApan. The approach has here been applied to the BoLA-I system, but the pipeline is readily applicable to MHC systems in other species.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, The Roslin Institute, Target Discovery Institute Mass Spectrometry Laboratory
Authors: Nielsen, M. (Intern), Connelley, T. (Ekstern), Ternette, N. (Ekstern)
Pages: 559-567
Publication date: 2018
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Proteome Research
Volume: 17
Issue number: 1
ISSN (Print): 1535-3893
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): SNIP 0.982 SJR 1.818
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 4.34 SJR 1.76 SNIP 1.018
Isolation and characterization of bacteriophages with therapeutic potential

The concerning spread of antibiotic resistant bacteria has directed the spotlight upon bacteriophages, in short phages, as potential candidates for therapeutic purposes. Far for being a novelty, phage therapy has been widely used in the 20s and 30s in western countries until the discovery of antibiotics, which, coupled with a lack of knowledge of phage biology at that time, let to the replacement of phage therapy by antibiotics. On the other side of the planet, the Georgian Eliava Institute has been using phages for treating bacterial diseases since short after phage discovery a century ago. Georgian pharmacies commonly sell phage cocktails from the Institute without the need of a doctor’s prescription. A thorough characterisation of the cocktail is though required for it to be accepted as pharmaceutical in the European Union. The potential to investigate the genetic material of microbial communities directly from the environment through metagenomics, allows for genomic characterisation of these cocktail. Furthermore, metagenomics analyses may lead to the discovery of novel phages with therapeutic potential, opening up a promising new horizon for phage therapy.

This thesis is divided into five parts, each assigned a chapter. Chapter 1 provides the reader with an introduction to phage biology, history and metagenomics. Here, the main bioinformatics methods used throughout the studies of the following chapters are also presented and briefly described. Chapter 2 presents the paper "HostPhinder: A Phage Host Prediction
Tool” published in May 2016. The tool predicts the bacterial host of a given phage based on co-occurrent k-mers between a query sequence and reference phage genomes with known host. HostPhinder’s accuracy in predicting the host species and genus of an evaluation set was higher than 74% and 81%, respectively. The tool can be applied to identify the host of phage sequences found for instance in metagenomes allowing for a first step characterisation. Chapter 3 presents the paper “Metagenomic analysis of therapeutic PYO phage cocktails from 1997 to 2014” submitted in October 2017 and currently under peer-revision. In this study, the compositions of 3 batches of a Georgian cocktail from 1997 to 2014 was compared by means of Next Generation Sequencing (NGS) and metagenomic analysis. Thirty and 29 phage draft genomes were found in the cocktails from 1997 and 2014, respectively. One of them was present in both sample and did not resemble any known phage genomes, strongly suggesting its novelty. Phage representatives of all bacterial targets supposedly targeted by the cocktail’s were found, as predicted using HostPhinder. A comparison between cocktails from 1997, 2000, and 2014 showed a closer composition between the first two cocktails. Chapter 4 presents the characterisation of historical S. aureus phages, once used for phage typing. Finally, the conclusive Chapter 5, recapitulates the main findings of this thesis and frame them into the perspective of potential future investigations.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Center for Biological Sequence Analysis, Department of Biotechnology and Biomedicine, Metabolic Signaling and Regulation
Authors: Villarroel, J. (Intern), Nielsen, M. (Intern), Larsen, M. V. (Intern), Kilstrup, M. (Intern)
Number of pages: 98
Publication date: 2018

Publication information
Publisher: Technical University of Denmark (DTU)
Original language: English
Main Research Area: Technical/natural sciences
Electronic versions: Julia_Villarroel_PhD_thesis_18October2017.pdf
Publication: Research › Ph.D. thesis – Annual report year: 2018

NetH2pan: A Computational Tool to Guide MHC peptide prediction on Murine Tumors
With the advancement of personalized cancer immunotherapies, new tools are needed to identify tumor antigens and evaluate T-cell responses in model systems, specifically those that exhibit clinically relevant tumor progression. Key transgenic mouse models of breast cancer are generated and maintained on the FVB genetic background, and one such model is the mouse mammary tumor virus-polyomavirus middle T antigen (MMTV-PyMT) mouse - an immunocompetent transgenic mouse that exhibits spontaneous mammary tumor development and metastasis with high penetrance. Backcrossing the MMTV-PyMT mouse from the FVB strain onto a C57BL/6 genetic background, in order to leverage well-developed C57BL/6 immunological tools, results in delayed tumor development and variable metastatic phenotypes. Therefore, we initiated characterization of the FVB MHC Class I H-2K haplotype to establish useful immunological tools for evaluating antigen specificity in the murine FVB strain. Our study provides the first detailed molecular and immunoproteomic characterization of the FVB H-2K MHC Class I alleles, including >8500 unique peptide ligands, a multi-allele murine MHC peptide prediction tool, and in vivo validation of these data using MMTV-PyMT primary tumors. This work allows researchers to rapidly predict H-2 peptide ligands for immune testing, including, but not limited to, the MMTV-PyMT model for metastatic breast cancer.

General information
State: Accepted/In press
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, University of Oklahoma Health Sciences Center, Universidad Nacional de San Martin, Utah State University
Number of pages: 29
Publication date: 2018
Main Research Area: Technical/natural sciences

Publication information
Journal: Cancer Immunology Research
ISSN (Print): 2326-6066
Ratings:
Web of Science (2018): Indexed yes
Scopus rating (2017): SNIP 1.584 SJR 4.956
Web of Science (2017): Indexed Yes
Scopus rating (2016): CiteScore 7.13 SJR 4.518 SNIP 1.469
Web of Science (2016): Indexed yes
Predicted MHC peptide binding promiscuity explains MHC class I 'hotspots' of antigen presentation defined by mass spectrometry eluted ligand data

Peptides that bind to and are presented by MHC class I and class II molecules collectively make up the immunopeptidome. In the context of vaccine development, an understanding of the immunopeptidome is essential, and much effort has been dedicated to its accurate and cost-effective identification. Current state-of-the-art methods mainly comprise in silico tools for predicting MHC binding, which is strongly correlated with peptide immunogenicity. However, only a small proportion of the peptides that bind to MHC molecules are, in fact, immunogenic, and substantial work has been dedicated to uncovering additional determinants of peptide immunogenicity. In this context, and in light of recent advancements in mass spectrometry (MS), the existence of immunological hotspots has been given new life, inciting the hypothesis that hotspots are associated with MHC class I peptide immunogenicity. We here introduce a precise terminology for defining these hotspots and carry out a systematic analysis of MS and in silico predicted hotspots. We find that hotspots defined from MS data are largely captured by peptide binding predictions, enabling their replication in silico. This leads us to conclude that hotspots, to a great degree, are simply a result of promiscuous HLA binding, which disproves the hypothesis that the identification of hotspots provides novel information in the context of immunogenic peptide prediction. Furthermore, our analyses demonstrate that the signal of ligand processing, although present in the MS data, has very low predictive power to discriminate between MS and in silico defined hotspots.
Mass spectrometry (MS)-based immunopeptidomics investigates the repertoire of peptides presented at the cell surface by major histocompatibility complex (MHC) molecules. The broad clinical relevance of MHC-associated peptides, e.g. in precision medicine, provides a strong rationale for the large-scale generation of immunopeptidomic datasets and recent developments in MS-based peptide analysis technologies now support the generation of the required data. Importantly, the availability of diverse immunopeptidomic datasets has resulted in an increasing need to standardize, store and exchange this type of data to enable better collaborations among researchers, to advance the field more efficiently and to establish quality measures required for the meaningful comparison of datasets. Here we present the SysteMHC Atlas (https://systemhcatlas.org), a public database that aims at collecting, organizing, sharing, visualizing and exploring immunopeptidomic information to the public domain and serves as a community resource toward the generation of a high-quality comprehensive map of the human immunopeptidome and the support of consistent measurement of immunopeptidomic sample cohorts.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning
Use of a Regression Model to Study Host-Genomic Determinants of Phage Susceptibility in MRSA

Staphylococcus aureus is a major agent of nosocomial infections. Especially in methicillin-resistant strains, conventional treatment options are limited and expensive, which has fueled a growing interest in phage therapy approaches. We have tested the susceptibility of 207 clinical S. aureus strains to 12 (nine monovalent) different therapeutic phage preparations and subsequently employed linear regression models to estimate the influence of individual host gene families on resistance to phages. Specifically, we used a two-step regression model setup with a preselection step based on gene family enrichment. We show that our models are robust and capture the data’s underlying signal by comparing their performance to that of models build on randomized data. In doing so, we have identified 167 gene families that govern phage resistance in our strain set and performed functional analysis on them. This revealed genes of possible prophage or mobile genetic element origin, along with genes involved in restriction-modification and transcription regulators, though the majority were genes of unknown function. This study is a step in the direction of understanding the intricate host-phage relationship in this important pathogen with the outlook to targeted phage therapy applications.

General information
State: Published
Organisations: Immunoinformatics and Machine Learning, Department of Bio and Health Informatics, GoSeqIt ApS, State Serum Institute, Hvidovre Hospital, Polish Academy of Sciences
Number of pages: 16
Publication date: 2018
Main Research Area: Technical/natural sciences

Publication information
Journal: The Journal of Antibiotics
Volume: 7
Issue number: 1
ISSN (Print): 0021-8820
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.331 SJR 0.182
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.16 SJR 0.198 SNIP 0.187
Web of Science (2016): Indexed yes
An Analysis of Natural T Cell Responses to Predicted Tumor Neoepitopes

Personalization of cancer immunotherapies such as therapeutic vaccines and adoptive T-cell therapy may benefit from efficient identification and targeting of patient-specific neoepitopes. However, current neoepitope prediction methods based on sequencing and predictions of epitope processing and presentation result in a low rate of validation, suggesting that the determinants of peptide immunogenicity are not well understood. We gathered published data on human neoepitopes originating from single amino acid substitutions for which T cell reactivity had been experimentally tested, including both immunogenic and non-immunogenic neoepitopes. Out of 1,948 neoepitope-HLA (human leukocyte antigen) combinations from 13 publications, 53 were reported to elicit a T cell response. From these data, we found an enrichment for responses among peptides of length 9. Even though the peptides had been pre-selected based on presumed likelihood of being immunogenic, we found using NetMHCpan-4.0 that immunogenic neoepitopes were predicted to bind significantly more strongly to HLA compared to non-immunogenic peptides. Investigation of the HLA binding strength of the
immunogenic peptides revealed that the vast majority (96%) shared very strong predicted binding to HLA and that the binding strength was comparable to that observed for pathogen-derived epitopes. Finally, we found that neopeptide dissimilarity to self is a predictor of immunogenicity in situations where neo- and normal peptides share comparable predicted binding strength. In conclusion, these results suggest new strategies for prioritization of mutated peptides, but new data will be needed to confirm their value.

**General information**

State: Published
Organisations: Department of Bio and Health Informatics, Cancer Genomics, Immunoinformatics and Machine Learning, T-cells & Cancer, National Veterinary Institute, Universidad Nacional de San Martin
Authors: Bjerregaard, A. (Intern), Nielsen, M. (Intern), Jurtz, V. I. (Intern), Barra, C. M. (Ekstern), Hadrup, S. R. (Intern), Szallasi, Z. I. (Intern), Eklund, A. C. (Intern)
Publication date: 2017
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Frontiers in Immunology
Volume: 8
Article number: 1566
ISSN (Print): 1664-3224
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.484 SJR 2.803
Web of Science (2017): Indexed yes
Scopus rating (2016): CiteScore 5.37 SJR 3.034 SNIP 1.476
Web of Science (2016): Indexed yes
Scopus rating (2015): SJR 2.827 SNIP 1.277 CiteScore 5.09
Web of Science (2015): Indexed yes
Scopus rating (2014): SJR 2.389 SNIP 1.057 CiteScore 4.24
Web of Science (2014): Indexed yes
Scopus rating (2013): SJR 1.908 SNIP 0.855 CiteScore 3.55
ISI indexed (2013): ISI indexed no
Scopus rating (2012): SJR 0.809 SNIP 0.193 CiteScore 1.38
ISI indexed (2012): ISI indexed no
Scopus rating (2011): SJR 0.121
Web of Science (2011): Indexed yes
Original language: English
Neopitopes, Neoantigens, Prediction, Immunogenicity, Mutations, MHC binding
Electronic versions:
fimmu_08_01566.pdf
DOIs:
10.3389/fimmu.2017.01566

**Bibliographical note**

This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).
Source: FindIt
Source-ID: 2393361022
Publication: Research - peer-review › Journal article – Annual report year: 2017

**ArrayPitope: Automated Analysis of Amino Acid Substitutions for Peptide Microarray-Based Antibody Epitope Mapping**

Identification of epitopes targeted by antibodies (B cell epitopes) is of critical importance for the development of many diagnostic and therapeutic tools. For clinical usage, such epitopes must be extensively characterized in order to validate specificity and to document potential cross-reactivity. B cell epitopes are typically classified as either linear epitopes, i.e. short consecutive segments from the protein sequence or conformational epitopes adapted through native protein folding. Recent advances in high-density peptide microarrays enable high-throughput, high-resolution identification and characterization of linear B cell epitopes. Using exhaustive amino acid substitution analysis of peptides originating from target antigens, these microarrays can be used to address the specificity of polyclonal antibodies raised against such antigens containing hundreds of epitopes. However, the interpretation of the data provided in such large-scale screenings is far from trivial and in most cases it requires advanced computational and statistical skills. Here, we present an online application for automated identification of linear B cell epitopes, allowing the non-expert user to analyse peptide
microarray data. The application takes as input quantitative peptide data of fully or partially substituted overlapping peptides from a given antigen sequence and identifies epitope residues (residues that are significantly affected by substitutions) and visualize the selectivity towards each residue by sequence logo plots. Demonstrating utility, the application was used to identify and address the antibody specificity of 18 linear epitope regions in Human Serum Albumin (HSA), using peptide microarray data consisting of fully substituted peptides spanning the entire sequence of HSA and incubated with polyclonal rabbit anti-HSA (and mouse anti-rabbit-Cy3). The application is made available at: www.cbs.dtu.dk/services/ArrayPitope.

**General information**

State: Published
Organisations: Center for Biological Sequence Analysis, Department of Bio and Health Informatics, Department of Acoustic Technology, Center for Biological sequence analysis, Immunoinformatics and Machine Learning, University of Copenhagen
Authors: Hansen, C. S. (Intern), Østerbye, T. (Ekstern), Marcatili, P. (Intern), Lund, O. (Intern), Buus, S. (Intern), Nielsen, M. (Intern)
Number of pages: 14
Publication date: 2017
Main Research Area: Technical/natural sciences

**Publication information**

Journal: PLoS One
Volume: 12
Issue number: 1
Article number: e0168453
ISSN (Print): 1932-6203
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.111 SJR 1.164
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.11 SJR 1.236 SNIP 1.101
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.427 SNIP 1.136 CiteScore 3.32
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.559 SNIP 1.148 CiteScore 3.54
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.772 SNIP 1.153 CiteScore 3.94
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.982 SNIP 1.156 CiteScore 4.15
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 2.425 SNIP 1.233 CiteScore 4.58
ISI indexed (2011): ISI indexed no
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 2.705 SNIP 1.178
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 2.614 SNIP 1.046
Web of Science (2009): Indexed yes
BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes

Antibodies have become an indispensable tool for many biotechnological and clinical applications. They bind their molecular target (antigen) by recognizing a portion of its structure (epitope) in a highly specific manner. The ability to predict epitopes from antigen sequences alone is a complex task. Despite substantial effort, limited advancement has been achieved over the last decade in the accuracy of epitope prediction methods, especially for those that rely on the sequence of the antigen only. Here, we present BepiPred-2.0 (http://www.cbs.dtu.dk/services/BepiPred/), a web server for predicting B-cell epitopes from antigen sequences. BepiPred-2.0 is based on a random forest algorithm trained on epitopes annotated from antibody-antigen protein structures. This new method was found to outperform other available tools for sequence-based epitope prediction both on epitope data derived from solved 3D structures, and on a large collection of linear epitopes downloaded from the IEDB database. The method displays results in a user-friendly and informative way, both for computer-savvy and non-expert users. We believe that BepiPred-2.0 will be a valuable tool for the bioinformatics and immunology community.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, La Jolla Institute for Allergy & Immunology
Authors: Jespersen, M. C. (Intern), Peters, B. (Ekstern), Nielsen, M. (Intern), Marcatili, P. (Intern)
Number of pages: 6
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Nucleic Acids Research
Volume: 45
Issue number: W1
Article number: gkx346
ISSN (Print): 0305-1048
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 9.28 SJR 7.883 SNIP 2.744
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 7.358 SNIP 2.631 CiteScore 9.48
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 6.64 SNIP 2.552 CiteScore 8.74
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 6.801 SNIP 2.284 CiteScore 8.46
Peptide antigen-presentation by Major Histocompatibility Class (MHC) I proteins initiates CD8+ T cell mediated immunity against pathogens and cancers. MHC I molecules typically bind peptides with nine amino acids in length with both ends tucked inside the major A and F binding pocket. It has been known for a while that longer peptides can also bind by either bulging out of the groove in the middle of the peptide or by binding in a zig-zag fashion inside the groove. In a recent study, we identified an alternative binding conformation of naturally occurring peptides from Toxoplasma gondii bound by HLA-A*02:01. These peptides were extended at the C-terminus (PΩ) and contained charged amino acids not more than 3 residues after the anchor amino acid at PΩ, which enabled them to open the F pocket and expose their C-terminal extension into the solvent. Here, we show that the mechanism of F pocket opening is dictated by the charge of the first charged amino acid found within the extension. While positively charged amino acid result in the Tyr84 swing, amino acids that are negatively charged induce a not previously described Lys146 lift. Further, we demonstrate that the peptides with alternative binding modes have properties that fit very poorly to the conventional MHC class I pathway, and suggest they are presented via alternative means, potentially including cross-presentation via the MHC class II pathway.
Citrullination only infrequently impacts peptide binding to HLA class II MHC

It has been hypothesized that HLA class II alleles associated with rheumatoid arthritis (RA) preferentially present self-antigens altered by post-translational modification, such as citrullination. To understand the role of citrullination we tested four RA-associated citrullinated epitopes and their corresponding wild-type version for binding to 28 common HLA class II. Binding patterns were variable, and no consistent impact of citrullination was identified. Indeed, in one case citrullination significantly increased binding compared to the WT peptide, in another citrullination was associated with a reduction in promiscuity by 40%. For a more comprehensive analysis, we tested over 200 citrullinated peptides derived from vimentin and collagen II for their capacity to bind the RA-associated shared epitope alleles DRB1*01:01 and DRB1*04:01. The overall effect of citrullination on binding was found to be relatively minor, and only rarely associated with 3-fold increases or decreases in affinity. Previous studies have suggested that citrullination of MHC anchor residues, in particular P4, is associated with generation of novel RA-associated epitopes. However, analysis of the predicted MHC-binding cores of all peptides tested found that in modified peptides with increased binding affinity the citrullinated residue was predicted to occupy an anchor position in only a minority of cases. Finally, we also show that identification of citrullinated peptide binders could be facilitated by using the NetMHCIIpan 3.1 algorithm, representing citrullination as a wildcard. Our studies identify a total of 117 citrullinated peptides that bound RA-associated alleles with an affinity of 1000 nM or better.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, La Jolla Institute for Allergy & Immunology, Janssen Research and Development, Hospital Israelita Albert Einstein
Number of pages: 17
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: P L o S One
Volume: 12
Issue number: 5
Article number: e0177140
ISSN (Print): 1932-6203
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.111 SJR 1.164
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Genomics of phages with therapeutic potential

Bacteriophages, viruses that prey on bacteria, have been applied since the 1920’s to treat and prevent bacterial infection. After the discovery of antibiotics, this route was however largely abandoned. Now, with antimicrobial resistance in human-pathogenic bacteria on the rise and a dire need for alternatives, phage therapy once again takes center stage.

Phage therapy holds the promise of substantial benefits both from the economic as well as the public health perspective but also holds distinct challenges. The aim of this PhD was to address how bioinformatics tools, specifically genomics and mathematical modelling, can be applied to move the field towards a future of actual phage therapy in humans. It is composed of three related research projects.

The first part of this thesis is an introduction to various topics and methods relevant to the research projects that jointly make up this PhD. Chapters 1 - 3 deal with phages, their use in therapy and the nosocomial pathogen Staphylococcus aureus. Following that, Chapter 4 and 5 provide an overview of Next Generation Sequencing as well as commonly employed genomics tools, while Chapter 6 details basics of Machine Learning.
The second part, divided into three chapters, presents the three research projects. In project 1, an important commercial phage cocktail with a long history was sequenced and its component phages analyzed. It was found that the cocktail is composed of at least 23 different phage types, which were present in differing abundances. Some of these phage types were successfully amplified on a collection of in-house bacteria corresponding to the cocktail’s stated bacterial targets. Further, no harmful genes were detected in the cocktail.

Project 2 deals with phage communities in sewage by comparing samples from around the world to each other as well as to databases of available phage genomes. It revealed a great diversity in the sequences, many of which were distant from all known phages. The phage content of the different sample locations exhibited a rather stable genomic distance that was not influenced by whether the locations were geographically close or not.

Project 3 had the goal of identifying gene families in the extensive accessory genome of the hospital pathogen Staphylococcus aureus that influence its susceptibility to clinical phage preparations. This was done by phage testing a set of patient-derived S. aureus isolates against a panel of phage preparations. We then sought to model the results using the bacteria’s genetic background as features. Doing so, we built nine models with sufficient explanatory power over the susceptibility outcome and from them identified a set of 167 gene families relevant for phage susceptibility.

The third part of the thesis consists of conclusive remarks and a critical reflection on how each of these projects has impacted the field and how they are connected as well as pointing out directions for future investigations.

In summary, the work included in this thesis focuses on applying genomics and mathematical modelling to questions related to phage therapy.

**General information**

State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Statens Serum Institut, GoSeqIt ApS
Authors: Zschach, H. (Intern), Nielsen, M. (Intern), Hasman, H. (Intern), Larsen, M. V. (Intern)
Number of pages: 131
Publication date: 2017

**Publication information**

Publisher: Technical University of Denmark (DTU)
Original language: English
Main Research Area: Technical/natural sciences
Electronic versions: genomics_phages_therapeutic.pdf

**Relations**

Projects:
Genomics of phages with therapeutic potential
Publication: Research › Ph.D. thesis – Annual report year: 2018

**GibbsCluster: unsupervised clustering and alignment of peptide sequences**

Receptor interactions with short linear peptide fragments (ligands) are at the base of many biological signaling processes. Conserved and information-rich amino acid patterns, commonly called sequence motifs, shape and regulate these interactions. Because of the properties of a receptor-ligand system or of the assay used to interrogate it, experimental data often contain multiple sequence motifs. GibbsCluster is a powerful tool for unsupervised motif discovery because it can simultaneously cluster and align peptide data. The GibbsCluster 2.0 presented here is an improved version incorporating insertion and deletions accounting for variations in motif length in the peptide input. In basic terms, the program takes as input a set of peptide sequences and clusters them into meaningful groups. It returns the optimal number of clusters it identified, together with the sequence alignment and sequence motif characterizing each cluster. Several parameters are available to customize cluster analysis, including adjustable penalties for small clusters and overlapping groups and a trash cluster to remove outliers. As an example application, we used the server to deconvolute multiple specificities in large-scale peptidome data generated by mass spectrometry. The server is available at http://www.cbs.dtu.dk/services/GibbsCluster-2.0.

**General information**

State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Universidad Nacional de San Martin
Authors: Andreatta, M. (Ekstern), Alvarez, B. (Ekstern), Nielsen, M. (Intern)
Number of pages: 6
Pages: W458-W463
Identification of immediate early gene products of bovine herpes virus 1 (BHV-1) as dominant antigens recognized by CD8 T cells in immune cattle

In common with other herpes viruses, bovine herpes virus 1 (BHV-1) induces strong virus-specific CD8 T-cell responses. However, there is a paucity of information on the antigenic specificity of the responding T-cells. The development of a system to generate virus-specific CD8 T-cell lines from BHV-1-immune cattle, employing Theileria-transformed cell lines for antigen presentation, has enabled us to address this issue. Use of this system allowed the study to screen for CD8 T-cell antigens that are efficiently presented on the surface of virus-infected cells. Screening of a panel of 16 candidate viral gene products with CD8 T-cell lines from 3 BHV-1-immune cattle of defined MHC genotypes identified 4 antigens, including 3 immediate early (IE) gene products (ICP4, ICP22 and Circ) and a tegument protein (UL49). Identification of the MHC restriction specificities revealed that the antigens were presented by two or three class I MHC alleles in each animal. Six CD8 T-cell epitopes were identified in the three IE proteins by screening of synthetic peptides. Use of an algorithm (NetMHCpan) that predicts the peptide-binding characteristics of restricting MHC alleles confirmed and, in some cases refined, the identity of the epitopes. Analyses of the epitope specificity of the CD8 T-cell lines showed that a large component of the response is directed against these IE epitopes. The results indicate that these IE gene products are dominant targets of the CD8 T-cell response in BHV-I-immune cattle and hence are prime-candidate antigens for the generation of a subunit vaccine.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, University of Edinburgh
Authors: Hart, J. (Ekstern), MacHugh, N. D. (Ekstern), Sheldrake, T. (Ekstern), Nielsen, M. (Intern), Morrison, W. I. (Ekstern)
Number of pages: 12
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of General Virology
ISSN (Print): 0022-1317
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.877 SJR 1.325
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.93 SJR 1.544 SNIP 0.891
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.738 SNIP 0.998 CiteScore 3.26
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.69 SNIP 1.057 CiteScore 3.25
Machine Learning Reveals a Non-Canonical Mode of Peptide Binding to MHC class II Molecules

MHC class II molecules play a fundamental role in the cellular immune system: they load short peptide fragments derived from extracellular proteins and present them on the cell surface. It is currently thought that the peptide binds lying more or less flat in the MHC groove, with a fixed distance of nine amino acids between the first and last residue in contact with the MHCII. While confirming that the great majority of peptides bind to the MHC using this canonical mode, we report evidence for an alternative, less common mode of interaction. A fraction of observed ligands were shown to have an unconventional spacing of the anchor residues that directly interact with the MHC, which could only be accommodated to the canonical MHC motif either by imposing a more stretched out peptide backbone (a 8mer core) or by the peptide bulging out of the MHC groove (a 10mer core). We estimated that on average 2% of peptides bind with a core deletion, and 0.45% with a core insertion, but the frequency of such non-canonical cores was as high as 10% for certain MHCII...
molecules. A mutational analysis and experimental validation of a number of these anomalous ligands demonstrated that they could only fit to their MHC binding motif with a non-canonical binding core of length different from nine. This previously undescribed mode of peptide binding to MHCII molecules gives a more complete picture of peptide presentation by MHCII and allows us to model more accurately this event. This article is protected by copyright. All rights reserved.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Universidad Nacional de San Martin, La Jolla Institute for Allergy & Immunology
Authors: Andreatta, M. (Ekstern), Jurtz, V. I. (Intern), Kaever, T. (Ekstern), Sette, A. (Ekstern), Peters, B. (Ekstern), Nielsen, M. (Intern)
Pages: 255-264
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Immunology
ISSN (Print): 0019-2805
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.938 SJR 1.69
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.74 SJR 1.964 SNIP 0.965
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 2.075 SNIP 0.965 CiteScore 3.83
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 2.048 SNIP 1.043 CiteScore 3.61
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 2.086 SNIP 1.084 CiteScore 3.97
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.941 SNIP 1.04 CiteScore 3.94
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.884 SNIP 0.992 CiteScore 3.75
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 0.121 SNIP 0.912
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 0.122 SNIP 0.924
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 0.111 SNIP 0.922
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 0.122 SNIP 0.965
Web of Science (2007): Indexed yes
Metagenomic Analysis of Therapeutic PYO Phage Cocktails from 1997 to 2014

Phage therapy has regained interest in recent years due to the alarming spread of antibiotic resistance. Whilst phage cocktails are commonly sold in pharmacies in countries such as Georgia and Russia, this is not the case in western countries due to western regulatory agencies requiring a thorough characterization of the drug. Here, DNA sequencing of constituent biological entities constitutes a first step. The pyophage (PYO) cocktail is one of the main commercial products of the Georgian Eliava Institute of Bacteriophage, Microbiology and Virology and is used to cure skin infections. Since its first production in the 1930s, the composition of the cocktail has been periodically modified to add phages effective against emerging pathogenic strains. In this paper, we compared the composition of three PYO cocktails from 1997 (PYO97), 2000 (PYO2000) and 2014 (PYO2014). Based on next generation sequencing, de novo assembly and binning of contigs into draft genomes based on tetranucleotide distance, thirty and twenty-nine phage draft genomes were predicted in PYO97 and PYO2014, respectively. Of these, thirteen and fifteen shared high similarity to known phages. Eleven draft genomes were found to be common in the two cocktails. One of these showed no similarity to publicly available phage genomes. Representatives of phages targeting E. faecalis, E. faecium, E. coli, Proteus, P. aeruginosa and S. aureus were found in both cocktails. Finally, we estimated larger overlap of the PYO2000 cocktail to PYO97 compared to PYO2014. Using next generation sequencing and metagenomics analysis, we were able to characterize and compare the content of PYO cocktails separated by 17 years in time. Even though the cocktail composition is upgraded every six months, we found it to remain relatively stable over the years.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Department of Biotechnology and Biomedicine, Metabolic Signaling and Regulation, GoSeqIt ApS
Authors: Villarroel, J. (Intern), Larsen, M. V. (Ekstern), Kilstrup, M. (Intern), Nielsen, M. (Intern)
Number of pages: 22
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Viruses
Volume: 9
Issue number: 11
Article number: 328
ISSN (Print): 1999-4915
Ratings:
Web of Science (2018): Indexed yes
Scopus rating (2017): SNIP 1.13 SJR 1.805
Web of Science (2017): Indexed Yes
Scopus rating (2016): CiteScore 3.6 SJR 1.747 SNIP 1.02
Web of Science (2016): Indexed yes
Scopus rating (2015): SJR 1.832 SNIP 1.034 CiteScore 3.74
Web of Science (2015): Indexed yes
Scopus rating (2014): SJR 1.906 SNIP 1.098 CiteScore 3.8
MHC-I Ligand Discovery Using Targeted Database Searches of Mass Spectrometry Data: Implications for T-Cell Immunotherapies

Class I major histocompatibility complex (MHC-I)-bound peptide ligands dictate the activation and specificity of CD8+ T cells and thus are important for devising T-cell immunotherapies. In recent times, advances in mass spectrometry (MS) have enabled the precise identification of these MHC-I peptides, wherein MS spectra are compared against a reference proteome. Unfortunately, matching these spectra to reference proteome databases is hindered by inflated search spaces attributed to a lack of enzyme restriction in the searches, limiting the efficiency with which MHC ligands are discovered. Here we offer a solution to this problem whereby we developed a targeted database search approach and accompanying tool SpectMHC, that is based on a priori-predicted MHC-I peptides. We first validated the approach using MS data from two different allotype-specific immunoprecipitates for the C57BL/6 mouse background. We then developed allotype-specific HLA databases to search previously published MS data sets of human peripheral blood mononuclear cells (PBMCs). This targeted search strategy improved peptide identifications for both mouse and human ligandomes by greater than 2-fold and is superior to traditional "no enzyme" searches of reference proteomes. Our targeted database search promises to uncover otherwise missed novel T-cell epitopes of therapeutic potential.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Dalhousie University, University of Tubingen
Authors: Murphy, J. P. (Ekstern), Konda, P. (Ekstern), Kowalewski, D. J. (Ekstern), Schuster, H. (Ekstern), Clements, D. (Ekstern), Kim, Y. (Ekstern), Cohen, A. M. (Ekstern), Sharif, T. (Ekstern), Nielsen, M. (Intern), Stevanovic, S. (Ekstern), Lee, P. W. (Ekstern), Gujar, S. (Ekstern)
Number of pages: 11
Pages: 1806-1816
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Proteome Research
Volume: 16
Issue number: 4
ISSN (Print): 1535-3893
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): SNIP 0.982 SJR 1.818
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 4.34 SJR 1.76 SNIP 1.018
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 1.933 SNIP 1.08 CiteScore 4.45
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 1.959 SNIP 1.174 CiteScore 4.64
MIToS.jl: mutual information tools for protein sequence analysis in the Julia language

Motivation: MIToS is an environment for mutual information analysis and a framework for protein multiple sequence alignments (MSAs) and protein structures (PDB) management in Julia language. It integrates sequence and structural information through SIFTS, making Pfam MSAs analysis straightforward. MIToS streamlines the implementation of any measure calculated from residue contingency tables and its optimization and testing in terms of protein contact prediction. As an example, we implemented and tested a BLOSUM62-based pseudo-count strategy in mutual information analysis.

Availability and Implementation: The software is totally implemented in Julia and supported for Linux, OS X and Windows. It’s freely available on GitHub under MIT license: http://mitos.leloir.org.ar.

Contacts: diegozea@gmail.com or.cmb@leloir.org.ar

Supplementary information:
Supplementary data are available at Bioinformatics online.

General information
State: Published
Organisations: Center for Biological sequence analysis, Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Fundación Instituto Leloir
Authors: Zea, D. J. (Ekstern), Anfossi, D. (Ekstern), Nielsen, M. (Intern), Marino-Buslje, C. (Ekstern)
MuPeXI: prediction of neo-epitopes from tumor sequencing data

Personalization of immunotherapies such as cancer vaccines and adoptive T cell therapy depends on identification of patient-specific neo-epitopes that can be specifically targeted. MuPeXI, the mutant peptide extractor and informer, is a program to identify tumor-specific peptides and assess their potential to be neo-epitopes. The program input is a file with somatic mutation calls, a list of HLA types, and optionally a gene expression profile. The output is a table with all tumor-specific peptides derived from nucleotide substitutions, insertions, and deletions, along with comprehensive annotation, including HLA binding and similarity to normal peptides. The peptides are sorted according to a priority score which is intended to roughly predict immunogenicity. We applied MuPeXI to three tumors for which predicted MHC-binding peptides had been screened for T cell reactivity, and found that MuPeXI was able to prioritize immunogenic peptides with an area under the curve of 0.63. Compared to other available tools, MuPeXI provides more information and is easier to use. MuPeXI is available as stand-alone software and as a web server at http://www.cbs.dtu.dk/services/MuPeXI.
NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data

Cytotoxic T cells are of central importance in the immune system's response to disease. They recognize defective cells by binding to peptides presented on the cell surface by MHC class I molecules. Peptide binding to MHC molecules is the single most selective step in the Ag-presentation pathway. Therefore, in the quest for T cell epitopes, the prediction of peptide binding to MHC molecules has attracted widespread attention. In the past, predictors of peptide-MHC interactions have primarily been trained on binding affinity data. Recently, an increasing number of MHC-presented peptides identified by mass spectrometry have been reported containing information about peptide-processing steps in the presentation pathway and the length distribution of naturally presented peptides. In this article, we present NetMHCpan-4.0, a method trained on binding affinity and eluted ligand data leveraging the information from both data types. Large-scale benchmarking of the method demonstrates an increase in predictive performance compared with state-of-the-art methods when it comes to identification of naturally processed ligands, cancer neoantigens, and T cell epitopes.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, La Jolla Institute for Allergy & Immunology, Universidad Nacional de San Martin
Authors: Jurtz, V. I. (Intern), Paul, S. (Ekstern), Andreatta, M. (Ekstern), Marcatili, P. (Intern), Peters, B. (Ekstern), Nielsen, M. (Intern)
Number of pages: 10
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Immunology
Volume: 199
Issue number: 8
ISSN (Print): 0022-1767
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): SNIP 1.112 SJR 2.837
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 4.79 SJR 3.474 SNIP 1.176
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 3.571 SNIP 1.26 CiteScore 5.05
Next-generation ELISA diagnostic assay for Chagas Disease based on the combination of short peptidic epitopes

Chagas Disease, caused by the protozoan Trypanosoma cruzi, is a major health and economic problem in Latin America for which no vaccine or appropriate drugs for large-scale public health interventions are yet available. Accurate diagnosis is essential for the early identification and follow up of vector-borne cases and to prevent transmission of the disease by way of blood transfusions and organ transplantation. Diagnosis is routinely performed using serological methods, some of which require the production of parasite lysates, parasite antigenic fractions or purified recombinant antigens. Although available serological tests give satisfactory results, the production of reliable reagents remains laborious and expensive. Short peptides spanning linear B-cell epitopes have proven ideal serodiagnostic reagents in a wide range of diseases. Recently, we have conducted a large-scale screening of T. cruzi linear B-cell epitopes using high-density peptide chips, leading to the identification of several hundred novel sequence signatures associated to chronic Chagas Disease. Here, we performed a serological assessment of 27 selected epitopes and of their use in a novel multipeptide-based diagnostic
method. A combination of 7 of these peptides were finally evaluated in ELISA format against a panel of 199 sera samples (Chagas-positive and negative, including sera from Leishmaniasis-positive subjects). The multipeptide formulation displayed a high diagnostic performance, with a sensitivity of 96.3% and a specificity of 99.15%. Therefore, the use of synthetic peptides as diagnostic tools are an attractive alternative in Chagas’ disease diagnosis.

**General information**

State: Published

Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Universidad Nacional de San Martin, Hospital de Ninos Ricardo Gutierrez, Universidad Nacional de Salta

Authors: Mucci, J. (Ekstern), Carmona, S. J. (Ekstern), Volcovich, R. (Ekstern), Altcheh, J. (Ekstern), Bracamonte, E. (Ekstern), Marco, J. D. (Ekstern), Nielsen, M. (Intern), Buscaglia, C. A. (Ekstern), Aguero, F. (Ekstern)

Number of pages: 19

Publication date: 2017

Main Research Area: Technical/natural sciences

**Publication information**

Journal: P L o S Neglected Tropical Diseases (Print)

Volume: 11

Issue number: 10

Article number: e0005972

ISSN (Print): 1935-2727

Ratings:

Web of Science (2018): Indexed yes

Web of Science (2017): Indexed yes

Scopus rating (2016): CiteScore 3.97

Scopus rating (2015): CiteScore 4.09

Scopus rating (2014): CiteScore 4.61

Scopus rating (2013): CiteScore 4.72

Scopus rating (2012): CiteScore 4.75

Scopus rating (2011): CiteScore 4.64

Web of Science (2011): Indexed yes

BFI (2010): BFI-level 1

BFI (2009): BFI-level 1

BFI (2008): BFI-level 1

Original language: English

Electronic versions:

pntd.0005972.pdf

DOIs:

10.1371/journal.pntd.0005972

**Bibliographical note**

Copyright: © 2017 Mucci et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Source: FindIt

Source-ID: 2391443111

Publication: Research - peer-review » Journal article – Annual report year: 2017

**NNAlign: a platform to construct and evaluate artificial neural network models of receptor-ligand interactions**

Peptides are extensively used to characterize functional or (linear) structural aspects of receptor-ligand interactions in biological systems, e.g. SH2, SH3, PDZ peptide-recognition domains, the MHC membrane receptors and enzymes such as kinases and phosphatases. NNAlign is a method for the identification of such linear motifs in biological sequences. The algorithm aligns the amino acid or nucleotide sequences provided as training set, and generates a model of the sequence motif detected in the data. The webserver allows setting up cross-validation experiments to estimate the performance of the model, as well as evaluations on independent data. Many features of the training sequences can be encoded as input, and the network architecture is highly customizable. The results returned by the server include a graphical representation of the motif identified by the method, performance values and a downloadable model that can be applied to scan protein sequences for occurrence of the motif. While its performance for the characterization of peptide-MHC interactions is widely documented, we extended NNAlign to be applicable to other receptor-ligand systems as well. Version 2.0 supports alignments with insertions and deletions, encoding of receptor pseudo-sequences, and custom alphabets for the training sequences. The server is available at http://www.cbs.dtu.dk/services/NNAlign-2.0.
Prediction and in vitro verification of potential CTL epitopes conserved among PRRSV-2 strains

Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) is the causative agent of one of the most important porcine diseases with a high impact on animal health, welfare, and production economy. PRRSV exhibits a multitude of immunoevasive strategies that, in combination with a very high mutation rate, has hampered the development of safe and broadly protective vaccines. Aiming at a vaccine inducing an effective cytotoxic T cell response, a bioinformatics approach was taken to identify conserved PRRSV-derived peptides predicted to react broadly with common swine leukocyte antigen (SLA) class I alleles. Briefly, all possible 9- and 10-mer peptides were generated from 104 complete PRRSV type 2 genomes of confirmed high quality, and peptides with high binding affinity to five common SLAs were identified combining the NetMHCpan and positional scanning combinatorial peptide libraries binding predictions. Predicted binders were prioritized according to genomic conservation and SLA coverage using the PopCover algorithm. From this, 53 peptides were acquired for further analysis. Binding affinity and stability of a subset of 101 peptide-SLA combinations were validated in vitro for 4 of the 5 SLAs. Eventually, 23% of the predicted peptide-SLA combinations showed to form complexes with a dissociation half-life ≥30 min. Additionally, combining the two prediction methods proved to be more robust across alleles than either method used alone in terms of predicted-to-observed correlations. In summary, our approach represents a finely tuned epitope prediction pipeline providing a rationally selected ensemble of peptides for future in vivo experiments with pigs expressing the included SLAs.
Protein features as determinants of wild-type glycoside hydrolase thermostability

Thermostable enzymes for conversion of lignocellulosic biomass into biofuels have significant advantages over enzymes with more moderate thermostability due to the challenging application conditions. Experimental discovery of thermostable enzymes is highly cost intensive, and the development of in-silico methods guiding the discovery process would be of high value. To develop such an in-silico method and provide the data foundation of it, we determined the melting temperatures of 602 fungal glycoside hydrolysases from the families GH5, 6, 7, 10, 11, 43 and AA9 (formerly GH61). We, then used sequence and homology modeled structure information of these enzymes to develop the ThermoP melting temperature prediction method. Furthermore, in the context of thermostability, we determined the relative importance of 160 molecular features, such as amino acid frequencies and spatial interactions, and exemplified their biological significance. The presented prediction method is made publicly available at http://www.cbs.dtu.dk/services/ThermoP. This article is protected by copyright. All rights reserved.
The Intergenic Recombinant HLA-B*46:01 Has a Distinctive Peptidome that Includes KIR2DL3 Ligands

HLA-B*46:01 was formed by an intergenic mini-conversion, between HLA-B*15:01 and HLA-C*01:02, in Southeast Asia during the last 50,000 years, and it has since become the most common HLA-B allele in the region. A functional effect of the mini-conversion was introduction of the C1 epitope into HLA-B*46:01, making it an exceptional HLA-B allotype that is recognized by the C1-specific natural killer (NK) cell receptor KIR2DL3. High-resolution mass spectrometry showed that HLA-B*46:01 has a low-diversity peptidome that is distinct from those of its parents. A minority (21%) of HLA-B*46:01 peptides, with common C-terminal characteristics, form ligands for KIR2DL3. The HLA-B*46:01 peptidome is predicted to be enriched for peptide antigens derived from Mycobacterium leprae. Overall, the results indicate that the distinctive peptidome and functions of HLA-B*46:01 provide carriers with resistance to leprosy, which drove its rapid rise in frequency in Southeast Asia.

General information

State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Stanford University, University of Oklahoma Health Sciences Center, Pure Protein LLC
Pages: 1394-1405
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Cell Reports
Volume: 19
Issue number: 7
ISSN (Print): 2211-1247
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): SNIP 1.648 SJR 7.552
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
A combined prediction strategy increases identification of peptides bound with high affinity and stability to porcine MHC class I molecules SLA-1*04:01, SLA-2*04:01, and SLA-3*04:01

Affinity and stability of peptides bound by major histocompatibility complex (MHC) class I molecules are important factors in presentation of peptides to cytotoxic T lymphocytes (CTLs). In silico prediction methods of peptide-MHC binding followed by experimental analysis of peptide-MHC interactions constitute an attractive protocol to select target peptides from the vast pool of viral proteome peptides. We have earlier reported the peptide binding motif of the porcine MHC-I molecules SLA-1*04:01 and SLA-2*04:01, identified by an ELISA affinity-based positional scanning combinatorial peptide library (PSCPL) approach. Here, we report the peptide binding motif of SLA-3*04:01 and combine two prediction methods and analysis of both peptide binding affinity and stability of peptide-MHC complexes to improve rational peptide selection.

Using a peptide prediction strategy combining PSCPL binding matrices and in silico prediction algorithms (NetMHCpan), peptide ligands from a repository of 8900 peptides were predicted for binding to SLA-1*04:01, SLA-2*04:01, and SLA-3*04:01 and validated by affinity and stability assays. From the pool of predicted peptides for SLA-1*04:01, SLA-2*04:01, and SLA-3*04:01, a total of 71, 28, and 38 % were binders with affinities below 500 nM, respectively. Comparison of peptide-SLA binding affinity and complex stability showed that peptides of high affinity generally, but not always, produce complexes of high stability. In conclusion, we demonstrate how state-of-the-art prediction and in vitro immunology tools in combination can be used for accurate selection of peptides for MHC class I binding, hence providing an expansion of the field of peptide-MHC analysis also to include pigs as a livestock experimental model.
A novel approach to probe host-pathogen interactions of bovine digital dermatitis, a model of a complex polymicrobial infection

Polymicrobial infections represent a great challenge for the clarification of disease etiology and the development of comprehensive diagnostic or therapeutic tools, particularly for fastidious and difficult-to-cultivate bacteria. Using bovine digital dermatitis (DD) as a disease model, we introduce a novel strategy to study the pathogenesis of complex infections. The strategy combines meta-transcriptomics with high-density peptide-microarray technology to screen for in vivo-expressed microbial genes and the host antibody response at the site of infection. Bacterial expression patterns supported the assumption that treponemes were the major DD pathogens but also indicated the active involvement of other phyla (primarily Bacteroidetes). Bacterial genes involved in chemotaxis, flagellar synthesis and protection against oxidative and acidic stress were among the major factors defining the disease. The extraordinary diversity observed in bacterial expression, antigens and host antibody responses between individual cows pointed toward microbial variability as a hallmark of DD. Persistence of infection and DD reinfection in the same individual is common; thus, high microbial diversity may undermine the host's capacity to mount an efficient immune response and maintain immunological memory towards DD. The common antigenic markers identified here using a high-density peptide microarray address this issue and may be useful for future preventive measures against DD.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, National Veterinary Institute, Section for Bacteriology, Pathology and Parasitology, Metagenomics, Hospital of Southern Jutland, Schafer-N ApS
Number of pages: 13
Publication date: 2016
Main Research Area: Technical/natural sciences

Publication information
Journal: BMC Genomics
Volume: 17
Issue number: 1
Article number: 987
ISSN (Print): 1471-2164
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.151 SJR 2.11
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 4.05 SJR 2.163 SNIP 1.096
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 2.348 SNIP 1.159 CiteScore 4.3
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 2.327 SNIP 1.199 CiteScore 4.18
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 2.195 SNIP 1.188 CiteScore 4.39
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 2.236 SNIP 1.243 CiteScore 4.61
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 2.307 SNIP 1.191 CiteScore 4.38
ISI indexed (2011): ISI indexed yes
Defining the HLA class I-associated viral antigen repertoire from HIV-1-infected human cells

Recognition and eradication of infected cells by cytotoxic T lymphocytes is a key defense mechanism against intracellular pathogens. High-throughput definition of HLA class I-associated immunopeptidomes by mass spectrometry is an increasingly important analytical tool to advance our understanding of the induction of T-cell responses against pathogens such as HIV-1. We utilized a liquid chromatography tandem mass spectrometry workflow including de novo-assisted database searching to define the HLA class I-associated immunopeptidome of HIV-1-infected human cells. We here report for the first time the identification of 75 HIV-1-derived peptides bound to HLA class I complexes that were purified directly from HIV-1-infected human primary CD4+ T cells and the C8166 human T-cell line. Importantly, one-third of eluted HIV-1 peptides had not been previously known to be presented by HLA class I. Over 82% of the identified sequences originated from viral protein regions for which T-cell responses have previously been reported but for which the precise HLA class I-binding sequences have not yet been defined. These results validate and expand the current knowledge of virus-specific antigenic peptide presentation during HIV-1 infection and provide novel targets for T-cell vaccine development.
Engineering a CTL-Tailored Replicon RNA Vaccine against PRRSV

The development of vaccines against porcine reproductive and respiratory syndrome virus (PRRSV) has been hampered by the high mutation rate and the multiple immunoevasive strategies of the virus. With the overall aim of designing a broad coverage vaccine that induces an effective CTL response against PRRSV, we have used a bioinformatics approach to identify common PRRSV type 2 epitopes predicted to react broadly with predominant swine MHC (SLA) alleles.

All possible 9- and 10-mer peptides derived from 104 wild-type strains were analyzed in silico for their predicted binding affinity to 3 common SLA class I alleles and ranked according to genomic conservation and SLA binding coverage. Of the 53 top-ranked peptides, 33 were verified in vitro as high affinity binders. Polyepitope gene cassettes of these peptides, flanked by an upstream ubiquitin sequence and a downstream FLAG tag, were cloned into a classical swine fever virus (CSFV)-derived replicon vector. Virus replicon particles (VRP) were rescued by transfection of a complementing cell line with replicon RNA. Polyepitope expression and subsequent proteasomal degradation was confirmed indirectly by increased FLAG-tagged protein detection in the presence of a proteasome inhibitor.

Finally, a vaccination-challenge experiment using 18 SLA-matched pigs is currently being conducted until July 2016 in which a test group and a control group are being vaccinated twice with VRPs expressing PRRSV epitopes and non-sense control epitopes, respectively, before challenged with live wild type PRRSV. The induced epitope specific cell-mediated immune responses are being monitored by ELISPOT, flow cytometry and cytotoxicity assays, and the degree of protection against infection will be characterized by qPCR and antibody analysis. The results will be available for IVIS.

This study exemplifies how bioinformatics epitope prediction, recombinant SLA molecules and RNA virus replicon design can be used to engineer a replicating non-propagating vaccine tailored to deliver conserved and immunogenic CTL epitopes.
Gapped sequence alignment using artificial neural networks: application to the MHC class I system

Motivation: Many biological processes are guided by receptor interactions with linear ligands of variable length. One such receptor is the MHC class I molecule. The length preferences vary depending on the MHC allele, but are generally limited to peptides of length 8–11 amino acids. On this relatively simple system, we developed a sequence alignment method based on artificial neural networks that allows insertions and deletions in the alignment. Results: We show that prediction methods based on alignments that include insertions and deletions have significantly higher performance than methods trained on peptides of single lengths. Also, we illustrate how the location of deletions can aid the interpretation of the modes of binding of the peptide-MHC, as in the case of long peptides bulging out of the MHC groove or protruding at either terminus. Finally, we demonstrate that the method can learn the length profile of different MHC molecules, and quantified the reduction of the experimental effort required to identify potential epitopes using our prediction algorithm. Availability and implementation: The NetMHC-4.0 method for the prediction of peptide-MHC class I binding affinity using gapped sequence alignment is publicly available at: http://www.cbs.dtu.dk/services/NetMHC-4.0.

General information
State: Published
Organisations: Immunological Bioinformatics, Department of Systems Biology, Center for Biological Sequence Analysis, Universidad Nacional de San Martin
Authors: Andreatta, M. (Ekstern), Nielsen, M. (Intern)
Number of pages: 7
Pages: 511-517
Publication date: 2016
Main Research Area: Technical/natural sciences

Publication information
Journal: Bioinformatics
Volume: 32
Issue number: 4
Article number: btv639
ISSN (Print): 1367-4803
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 6.42
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 6.06
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 5.5
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
HIV infection is associated with preservation of MAIT cells in the lungs but alteration of their phenotype and T cell receptor repertoire

Tuberculosis remains the leading cause of death in HIV-positive people. A better understanding of the impact of HIV on lung immunity may lead to novel immunotherapeutic interventions. MAIT cells are tissue-homing donor-unrestricted T cells with broad anti-microbial activity. HIV infection causes early and irreversible depletion of MAIT cells in the peripheral circulation, but the effect of HIV on MAIT cells in the lungs is unknown. These researchers report, for the first time, that MAIT cells in the lungs are numerically preserved but phenotypically and clonotypically altered by HIV infection. They confirm previous reports that circulating MAIT cells are depleted in HIV. Their results suggest that peripheral MAIT cell depletions observed in HIV infection may be due to compartment-specific microbial alterations and/or tissue redistribution. The presenters emphasized that further study is needed to determine the mechanisms underlying the altered phenotypes of lung-resident MAITs and whether these can be targeted to improve anti-microbial lung immunity in people living with HIV.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immuneinformatics and Machine Learning, KwaZulu Natal Research Institute for Tuberculosis and HIV, Massachusetts General Hospital, Massachusetts Institute of Technology, Durban University of Technology, University of KwaZulu-Natal, Inkosi Albert Luthuli Hospital, McGill University, Oregon Health and Science University
Number of pages: 1
Publication date: 2016
HostPhinder: A Phage Host Prediction Tool

The current dramatic increase of antibiotic resistant bacteria has revitalised the interest in bacteriophages as alternative antibacterial treatment. Meanwhile, the development of bioinformatics methods for analysing genomic data places high-throughput approaches for phage characterization within reach. Here, we present HostPhinder, a tool aimed at predicting the bacterial host of phages by examining the phage genome sequence. Using a reference database of 2196 phages with known hosts, HostPhinder predicts the host species of a query phage as the host of the most genomically similar reference phages. As a measure of genomic similarity the number of co-occurring k-mers (DNA sequences of length k) is used. Using an independent evaluation set, HostPhinder was able to correctly predict host genus and species for 81% and 74% of the phages respectively, giving predictions for more phages than BLAST and significantly outperforming BLAST on phages for which both had predictions. HostPhinder predictions on phage draft genomes from the INTESTI phage cocktail corresponded well with the advertised targets of the cocktail. Our study indicates that for most phages genomic similarity correlates well with related bacterial hosts. HostPhinder is available as an interactive web service [1] and as a stand alone download from the Docker registry [2].

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Technical University of Denmark
Authors: Villarroel, J. (Intern), Kleinheinz, K. A. (Ekstern), Jurtz, V. I. (Intern), Zschach, H. (Intern), Lund, O. (Intern), Nielsen, M. (Intern), Larsen, M. V. (Intern)
Number of pages: 22
Publication date: 2016
Main Research Area: Technical/natural sciences

Publication information
Journal: Viruses
Volume: 8
Issue number: 5
Article number: 116
ISSN (Print): 1999-4915
Ratings:
Web of Science (2018): Indexed yes
Scopus rating (2017): SNIP 1.13 SJR 1.805
Web of Science (2017): Indexed Yes
Scopus rating (2016): CiteScore 3.6 SJR 1.747 SNIP 1.02
Web of Science (2016): Indexed yes
Scopus rating (2015): SJR 1.832 SNIP 1.034 CiteScore 3.74
The aim of the present study was to identify influenza A-derived peptides which bind to both HLA class I and -II molecules and by immunization lead to both HLA class I and class II restricted immune responses. Eight influenza A-derived 9-11mer peptides with simultaneous binding to both HLA-A*02:01 and HLA-DRB1*01:01 molecules were identified by bioinformatics and biochemical technology. Immunization of transgenic HLA-A*02:01/HLA-DRB1*01:01 mice with four of these double binding peptides gave rise to both HLA class I and class II restricted responses by CD8 and CD4 T cells, respectively, whereas four of the double binding peptides did result in HLA-A*02:01 restricted responses only. According to their cytokine profile, the CD4 T cell responses were of the Th2 type. In influenza infected mice, we were unable to detect natural processing in vivo of the double restricted peptides and in line with this, peptide vaccination did not decrease virus titres in the lungs of intranasally influenza challenged mice. Our data show that HLA class I and class II double binding peptides can be identified by bioinformatics and biochemical technology. By immunization, double binding peptides can give rise to both HLA class I and class I restricted responses, a quality which might be of potential interest for peptide-based vaccine development.
Improved pan-specific prediction of MHC class I peptide binding using a novel receptor clustering data partitioning strategy

Pan-specific prediction of receptor-ligand interaction is conventionally done using machine-learning methods that integrates information about both receptor and ligand primary sequences. To achieve optimal performance using machine learning, dealing with overfitting and data redundancy is critical. Most often so-called ligand clustering methods have been used to deal with these issues in the context of pan-specific receptor-ligand predictions, and the MHC system the approach has proven highly effective for extrapolating information from a limited set of receptors with well characterized binding motifs, to others with no or very limited experimental characterization. The success of this approach has however proven to depend strongly on the similarity of the query molecule to the molecules with characterized specificity using in the machine-learning process. Here, we outline an alternative strategy with the aim of altering this and construct data sets optimal for training of pan-specific receptor-ligand predictions focusing on receptor similarity rather than ligand similarity.

We show that this receptor clustering method consistently in benchmarks covering affinity predictions, MHC ligand and MHC epitope identification perform better than the conventional ligand clustering method on the alleles with remote similarity to the training set.

Bibliographical note

© 2016 Pedersen et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Source: FindIt
Source-ID: 2290144236
Publication: Research - peer-review › Journal article – Annual report year: 2016

Improved pan-specific prediction of MHC class I peptide binding using a novel receptor clustering data partitioning strategy

Pan-specific prediction of receptor-ligand interaction is conventionally done using machine-learning methods that integrates information about both receptor and ligand primary sequences. To achieve optimal performance using machine learning, dealing with overfitting and data redundancy is critical. Most often so-called ligand clustering methods have been used to deal with these issues in the context of pan-specific receptor-ligand predictions, and the MHC system the approach has proven highly effective for extrapolating information from a limited set of receptors with well characterized binding motifs, to others with no or very limited experimental characterization. The success of this approach has however proven to depend strongly on the similarity of the query molecule to the molecules with characterized specificity using in the machine-learning process. Here, we outline an alternative strategy with the aim of altering this and construct data sets optimal for training of pan-specific receptor-ligand predictions focusing on receptor similarity rather than ligand similarity.

We show that this receptor clustering method consistently in benchmarks covering affinity predictions, MHC ligand and MHC epitope identification perform better than the conventional ligand clustering method on the alleles with remote similarity to the training set.

General information

State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Evaxion Biotech, University of Copenhagen
Toxoplasma gondii peptide ligands open the gate of the HLA class I binding groove

HLA class I presentation of pathogen-derived peptide ligands is essential for CD8+ T cell recognition of Toxoplasma gondii infected cells. Currently, little data exist pertaining to peptides that are presented after T. gondii infection. Herein we purify HLA-A*02:01 complexes from T. gondii infected cells and characterize the peptide ligands using LCMS. We identify 195 T. gondii encoded ligands originating from both secreted and cytoplasmic proteins. Surprisingly, T. gondii ligands are significantly longer than uninfected host ligands, and these longer pathogen derived peptides maintain a canonical N-terminal binding core yet exhibit a C-terminal extension of 1-30 amino acids. Structural analysis demonstrates that binding of extended peptides opens the HLA class I F’ pocket, allowing the C-terminal extension to protrude through one end of the binding groove. In summary, we demonstrate that unrealized structural flexibility makes MHC class I receptive to parasite-derived ligands that exhibit unique C-terminal peptide extensions.
Bacteriophages are the most abundant biological entity on the planet, but at the same time do not account for much of the genetic material isolated from most environments due to their small genome sizes. They also show great genetic diversity and mosaic genomes making it challenging to analyze and understand them. Here we present MetaPhinder, a method to identify assembled genomic fragments (i.e. contigs) of phage origin in metagenomic data sets. The method is based on a comparison to a database of whole genome bacteriophage sequences, integrating hits to multiple genomes to accommodate for the mosaic genome structure of many bacteriophages. The method is demonstrated to outperform both BLAST methods based on single hits and methods based on k-mer comparisons. MetaPhinder is available as a web service at the Center for Genomic Epidemiology https://cge.cbs.dtu.dk/services/MetaPhinder/, while the source code can be downloaded from https://bitbucket.org/genomicepidemiology/metaphinder or https://github.com/vanessajurtz/MetaPhinder.
NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets

Background: Binding of peptides to MHC class I molecules (MHC-I) is essential for antigen presentation to cytotoxic T-cells. Results: Here, we demonstrate how a simple alignment step allowing insertions and deletions in a pan-specific MHC-I binding machine-learning model enables combining information across both multiple MHC molecules and peptide lengths. This pan-allele/pan-length algorithm significantly outperforms state-of-the-art methods, and captures differences in the length profile of binders to different MHC molecules leading to increased accuracy for ligand identification. Using this model, we demonstrate that percentile ranks in contrast to affinity-based thresholds are optimal for ligand identification due to uniform sampling of the MHC space. Conclusions: We have developed a neural network-based machine-learning algorithm leveraging information across multiple receptor specificities and ligand length scales, and demonstrated how this approach significantly improves the accuracy for prediction of peptide binding and identification of MHC ligands. The
Pan-specific prediction of peptide-MHC Class I complex stability, a correlate of T cell immunogenicity

Binding of peptides to MHC class I (MHC-I) molecules is the most selective event in the processing and presentation of Ags to CTL, and insights into the mechanisms that govern peptide-MHC-I binding should facilitate our understanding of CTL biology. Peptide-MHC-I interactions have traditionally been quantified by the strength of the interaction, that is, the
binding affinity, yet it has been shown that the stability of the peptide-MHC-I complex is a better correlate of immunogenicity compared with binding affinity. In this study, we have experimentally analyzed peptide-MHC-I complex stability of a large panel of human MHC-I allotypes and generated a body of data sufficient to develop a neural network-based pan-specific predictor of peptide-MHC-I complex stability. Integrating the neural network predictors of peptide-MHC-I complex stability with state-of-the-art predictors of peptide-MHC-I binding is shown to significantly improve the prediction of CTL epitopes. The method is publicly available at http://www.cbs.dtu.dk/services/NetMHCstabpan.
Sequence diversity between class I MHC loci of African native and introduced Bos taurus cattle in Theileria parva endemic regions: in silico peptide binding prediction identifies distinct functional clusters

There is strong evidence that the immunity induced by live vaccination for control of the protozoan parasite Theileria parva is mediated by class I MHC-restricted CD8⁺ T cells directed against the schizont stage of the parasite that infects bovine lymphocytes. The functional competency of class I MHC genes is dependent on the presence of codons specifying certain critical amino acid residues that line the peptide binding groove. Compared with European Bos taurus in which class I MHC allelic polymorphisms have been examined extensively, published data on class I MHC transcripts in African taurines in T. parva endemic areas is very limited. We utilized the multiplexing capabilities of 454 pyrosequencing to make an initial assessment of class I MHC allelic diversity in a population of Ankole cattle. We also typed a population of exotic Holstein cattle from an African ranch for class I MHC and investigated the extent, if any, that their peptide-binding motifs overlapped with those of Ankole cattle. We report the identification of 18 novel allelic sequences in Ankole cattle and provide evidence of positive selection for sequence diversity, including in residues that predominantly interact with peptides. In silico functional analysis resulted in peptide binding specificities that were largely distinct between the two breeds. We also demonstrate that CD8⁺ T cells derived from Ankole cattle that are seropositive for T. parva do not recognize vaccine candidate antigens originally identified in Holstein and Boran (Bos indicus) cattle breeds.
T-cell recognition is shaped by epitope sequence conservation in the host proteome and microbiome

Several mechanisms exist to avoid or suppress inflammatory T-cell immune responses that could prove harmful to the host due to targeting self-antigens or commensal microbes. We hypothesized that these mechanisms could become evident when comparing the immunogenicity of a peptide from a pathogen or allergen with the conservation of its sequence in the human proteome or the healthy human microbiome. Indeed, performing such comparisons on large sets of validated T-cell epitopes, we found that epitopes that are similar with self-antigens above a certain threshold showed lower immunogenicity, presumably as a result of negative selection of T cells capable of recognizing such peptides. Moreover, we also found a reduced level of immune recognition for epitopes conserved in the commensal microbiome, presumably as a result of peripheral tolerance. These findings indicate that the existence (and potentially the polarization) of T-cell responses to a given epitope is influenced and to some extent predictable based on its similarity to self-antigens and commensal antigens.
The Length Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference

HLA class I-binding predictions are widely used to identify candidate peptide targets of human CD8+ T cell responses. Many such approaches focus exclusively on a limited range of peptide lengths, typically 9 aa and sometimes 9-10 aa, despite multiple examples of dominant epitopes of other lengths. In this study, we examined whether epitope predictions can be improved by incorporating the natural length distribution of HLA class I ligands. We found that, although different HLA alleles have diverse length-binding preferences, the length profiles of ligands that are naturally presented by these alleles are much more homogeneous. We hypothesized that this is due to a defined length profile of peptides available for HLA binding in the endoplasmic reticulum. Based on this, we created a model of HLA allele-specific ligand length profiles and demonstrate how this model, in combination with HLA-binding predictions, greatly improves comprehensive identification of CD8+ T cell epitopes.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, University of Oklahoma, La Jolla Institute for Allergy & Immunology
Number of pages: 8
Pages: 1480-1487
Publication date: 2016
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Immunology
Volume: 196
Issue number: 4
ISSN (Print): 0022-1767
Ratings:
BFI (2018): BFI-level 2
Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification

A key event in the generation of a cellular response against malicious organisms through the endocytic pathway is binding of peptidic antigens by major histocompatibility complex class II (MHC class II) molecules. The bound peptide is then presented on the cell surface where it can be recognized by T helper lymphocytes. NetMHCIIpan is a state-of-the-art method for the quantitative prediction of peptide binding to any human or mouse MHC class II molecule of known sequence. In this paper, we describe an updated version of the method with improved peptide binding register identification. Binding register prediction is concerned with determining the minimal core region of nine residues directly in contact with the MHC binding cleft, a crucial piece of information both for the identification and design of CD4⁺ T cell antigens. When applied to a set of 51 crystal structures of peptide-MHC complexes with known binding registers, the new method NetMHCIIpan-3.1 significantly outperformed the earlier 3.0 version. We illustrate the impact of accurate binding core identification for the interpretation of T cell cross-reactivity using tetramer double staining with a CMV epitope and its variants mapped to the epitope binding core. NetMHCIIpan is publicly available at http://www.cbs.dtu.dk/services/NetMHCIIpan-3.1.
A modern approach for epitope prediction: identification of foot-and-mouth disease virus peptides binding bovine leukocyte antigen (BoLA) class I molecules

Major histocompatibility complex (MHC) class I molecules regulate adaptive immune responses through the presentation of antigenic peptides to CD8+ T cells. Polymorphisms in the peptide binding region of class I molecules determine peptide binding affinity and stability during antigen presentation, and different antigen peptide motifs are associated with specific genetic sequences of class I molecules. Understanding bovine leukocyte antigen (BoLA), peptide-MHC class I binding specificities may facilitate development of vaccines or reagents for quantifying the adaptive immune response to intracellular pathogens, such as foot-and-mouth disease virus (FMDV). Six synthetic BoLA class I (BoLA-I) molecules were produced, and the peptide binding motif was generated for five of the six molecules using a combined approach of positional scanning combinatorial peptide libraries (PSCPLs) and neural network-based predictions (NetMHCpan). The updated NetMHCpan server was used to predict BoLA-I binding peptides within the P1 structural polyprotein sequence of FMDV (strain A24 Cruzeiro) for BoLA-I*01901, BoLA-I*00801, BoLA-I*01201, and BoLA-I*02401. Peptide binding affinity and stability were determined for these BoLA-I molecules using the luminescent oxygen channeling immunoassay (LOCI) and scintillation proximity assay (SPA). The functional diversity of known BoLA alleles was predicted using the MHCcluster tool, and functional predictions for peptide motifs were compared to observed data from this and prior studies. The results of these analyses showed that BoLA alleles cluster into three distinct groups with the potential to define "BoLA supertypes." This streamlined approach identifies potential T cell epitopes from pathogens, such as FMDV, and provides insight into T cell immunity following infection or vaccination.

General information
State: Published
Organisations: Department of Chemical and Biochemical Engineering, Department of Systems Biology, Center for Biological Sequence Analysis, University of Vermont, University of Copenhagen, Agricultural Research Service
Authors: Pandya, M. (Ekstern), Rasmussen, M. (Ekstern), Hansen, A. (Intern), Nielsen, M. (Intern), Buus, S. (Ekstern), Golde, W. (Ekstern), Barlow, J. (Ekstern)
Number of pages: 13
Automated benchmarking of peptide-MHC class I binding predictions

Motivation: Numerous in silico methods predicting peptide binding to major histocompatibility complex (MHC) class I molecules have been developed over the last decades. However, the multitude of available prediction tools makes it non-trivial for the end-user to select which tool to use for a given task. To provide a solid basis on which to compare different prediction tools, we here describe a framework for the automated benchmarking of peptide-MHC class I binding prediction tools. The framework runs weekly benchmarks on data that are newly entered into the Immune Epitope Database (IEDB), giving the public access to frequent, up-to-date performance evaluations of all participating tools. To overcome potential selection bias in the data included in the IEDB, a strategy was implemented that suggests a set of peptides for which different prediction methods give divergent predictions as to their binding capability. Upon experimental binding validation, these peptides entered the benchmark study.

Results: The benchmark has run for 15 weeks and includes evaluation of 44 datasets covering 17 MHC alleles and more than 4000 peptide-MHC binding measurements. Inspection of the results allows the end-user to make educated selections between participating tools. Of the four participating servers, NetMHCpan performed the best, followed by ANN, SMM and finally ARB.

Availability and implementation: Up-to-date performance evaluations of each server can be found online at http://tools.iedb.org/auto_bench/mhci/weekly. All prediction tool developers are invited to participate in the benchmark. Sign-up instructions are available at http://tools.iedb.org/auto_bench/mhci/join.

General information

State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, La Jolla Institute for Allergy & Immunology
Authors: Trolle, T. (Intern), Metushi, I. G. (Ekstern), Greenbaum, J. (Ekstern), Kim, Y. (Ekstern), Sidney, J. (Ekstern), Lund, O. (Intern), Sette, A. (Ekstern), Peters, B. (Ekstern), Nielsen, M. (Intern)
Number of pages: 8
Pages: 2174-2181
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information

Journal: Bioinformatics
Volume: 31
Issue number: 13
ISSN (Print): 1367-4803
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 6.42
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 6.06
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 5.5
Web of Science (2014): Indexed yes
Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) causes one of the most important diseases in all swine producing countries. The infection has a high impact on animal welfare, food safety and production economics. PRRSV possesses multiple immunoevasive strategies, from suppression of the host cell antiviral machinery, to the deceptive induction of a non-neutralizing antibody response through decoy antigen presentation. This, combined with a very high mutation rate, has hampered the development of safe and effective vaccines. With the overall aim to design a vaccine that induces an effective CTL response against PRRSV, we have taken a bioinformatics approach to identify common PRRSV epitopes predicted to react broadly with predominant swine MHC class I alleles. First, the genomic integrity and sequencing method was examined for 334 available complete PRRSV type 2 genomes leaving 104 strains of high quality. For each strain, a library of all possible 9- and 10-mer peptides was generated considering the known ribosomal frame shift sites and sites for post translational cleavage. All peptides were in silico analyzed for binding affinity to either of five common SLA class I alleles. A quantitative rank score was generated for each peptide by combining two algorithms based on the NetMHCpan neural network and lab determined SLA binding affinity of each amino acid at any position in the peptide, respectively. Peptides with a rank score above a predefined threshold were further analyzed by the PopCover algorithm, providing a final list of 54 epitopes prioritized according to maximum coverage of PRRSV strains and SLA alleles. This bioinformatics approach provides a rational strategy for selecting peptides for a CTL-activating vaccine with broad coverage of both virus and swine diversity. The immunogenicity of the selected peptides is in the process of being verified in vivo.
Ebola virus comparative genomics
The 2014 Ebola outbreak in West Africa is the largest documented for this virus. To examine the dynamics of this genome, we compare more than 100 currently available ebolavirus genomes to each other and to other viral genomes. Based on oligomer frequency analysis, the family Filoviridae forms a distinct group from all other sequenced viral genomes. All filovirus genomes sequenced to date encode proteins with similar functions and gene order, although there is considerable divergence in sequences between the three genera Ebola virus, Cuevavirus, and Marburgvirus within the family Filoviridae. Whereas all ebolavirus genomes are quite similar (multiple sequences of the same strain are often identical), variation is most common in the intergenic regions and within specific areas of the genes encoding the glycoprotein (GP), nucleoprotein (NP) and polymerase (L). We predict regions that could contain epitope-binding sites, which might be good vaccine targets. This information, combined with glycosylation sites and experimentally determined epitopes, can identify the most promising regions for the development of therapeutic strategies.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Oak Ridge National Laboratory, Booz Allen Hamilton, Molecular Microbiology and Genomics Consultants
Authors: Jun S. (Ekstern), Leuze M. R. (Ekstern), Nookaew I. (Ekstern), Uberbacher E. C. (Ekstern), Land M. (Ekstern), Zhang Q. (Ekstern), Wanchai V. (Ekstern), Chai J. (Ekstern), Nielsen M. (Intern), Trolle T. (Intern), Lund O. (Intern), Buzard G. S. (Ekstern), Pedersen T. D. (Intern), Wassenaar T. M. (Ekstern), Ussery D. W. (Intern)
Number of pages: 15
Pages: 764-778
Publication date: 2015
Main Research Area: Technical/natural sciences
Publication information
Journal: FEMS Microbiology Reviews
Volume: 479
Issue number: 1
Article number: fuv031
ISSN (Print): 0168-6445
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 13.54
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 13.38
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 12.9
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
High-throughput sequencing enhanced phage display enables the identification of patient-specific epitope motifs in serum

Phage display is a prominent screening technique with a multitude of applications including therapeutic antibody development and mapping of antigen epitopes. In this study, phages were selected based on their interaction with patient serum and exhaustively characterised by high-throughput sequencing. A bioinformatics approach was developed in order to identify peptide motifs of interest based on clustering and contrasting to control samples. Comparison of patient and control samples confirmed a major issue in phage display, namely the selection of unspecific peptides. The potential of the bioinformatic approach was demonstrated by identifying epitopes of a prominent peanut allergen, Ara h 1, in sera from patients with severe peanut allergy. The identified epitopes were confirmed by high-density peptide micro-arrays. The present study demonstrates that high-throughput sequencing can empower phage display by (i) enabling the analysis of complex biological samples, (ii) circumventing the traditional laborious picking and functional testing of individual phage clones and (iii) reducing the number of selection rounds.

General information
State: Published
Organisations: Department of Micro- and Nanotechnology, Fluidic Array Systems and Technology, Department of Systems Biology, Center for Biological Sequence Analysis, National Food Institute, Regulatory Genomics, Roche NimbleGen, Quadram Institute, Medical University of Vienna
LYRA, a webserver for lymphocyte receptor structural modeling.
The accurate structural modeling of B- and T-cell receptors is fundamental to gain a detailed insight in the mechanisms underlying immunity and in developing new drugs and therapies. The LYRA (LYmphocyte Receptor Automated modeling) web server (http://www.cbs.dtu.dk/services/LYRA/) implements a complete and automated method for building of B- and T-cell receptor structural models starting from their amino acid sequence alone. The webserver is freely available and easy to use for non-specialists. Upon submission, LYRA automatically generates alignments using ad hoc profiles, predicts the structural class of each hypervariable loop, selects the best templates in an automatic fashion, and provides within minutes a complete 3D model that can be downloaded or inspected online. Experienced users can manually select or exclude template structures according to case specific information. LYRA is based on the canonical structure method, that in the last 30 years has been successfully used to generate antibody models of high accuracy, and in our benchmarks this approach proves to achieve similarly good results on TCR modeling, with a benchmarked average RMSD accuracy of 1.29 and 1.48 Å for B- and T-cell receptors, respectively. To the best of our knowledge, LYRA is the first automated server
for the prediction of TCR structure.

**General information**
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Klausen, M. S. (Intern), Anderson, M. V. (Intern), Jespersen, M. C. (Intern), Nielsen, M. (Intern), Marcatili, P. (Intern)
Number of pages: 7
Pages: W349-W355
Publication date: 2015
Main Research Area: Technical/natural sciences

**Publication information**
Journal: Nucleic Acids Research
Volume: 43
Issue number: W1
ISSN (Print): 0305-1048
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 9.28 SJR 7.883 SNIP 2.744
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 7.358 SNIP 2.631 CiteScore 9.48
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 6.64 SNIP 2.552 CiteScore 8.74
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 6.801 SNIP 2.284 CiteScore 8.46
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 6.329 SNIP 2.407 CiteScore 8.62
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 5.976 SNIP 2.19 CiteScore 7.86
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 5.381 SNIP 2.034
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 5.669 SNIP 1.874
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 4.912 SNIP 1.578
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 5.1 SNIP 1.807
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 4.776 SNIP 2.051
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 5.092 SNIP 2.147
MR1-restricted MAIT cells display ligand discrimination and pathogen selectivity through distinct T cell receptor usage

Mucosal-associated invariant T (MAIT) cells express a semi-invariant T cell receptor (TCR) that detects microbial metabolites presented by the nonpolymorphic major histocompatibility complex (MHC)-like molecule MR1. The highly conserved nature of MR1 in conjunction with biased MAIT TCRα chain usage is widely thought to indicate limited ligand presentation and discrimination within a pattern-like recognition system. Here, we evaluated the TCR repertoire of MAIT cells responsive to three classes of microbes. Substantial diversity and heterogeneity were apparent across the functional MAIT cell repertoire as a whole, especially for TCRβ chain sequences. Moreover, different pathogen-specific responses were characterized by distinct TCR usage, both between and within individuals, suggesting that MAIT cell adaptation was a direct consequence of exposure to various exogenous MR1-restricted epitopes. In line with this interpretation, MAIT cell clones with distinct TCRs responded differentially to a riboflavin metabolite. These results suggest that MAIT cells can discriminate between pathogen-derived ligands in a clonotype-dependent manner, providing a basis for adaptive memory via recruitment of specific repertoires shaped by microbial exposure.
Prediction of Antibody Epitopes

Antibodies recognize their cognate antigens in a precise and effective way. In order to do so, they target regions of the antigenic molecules that have specific features such as large exposed areas, presence of charged or polar atoms, specific secondary structure elements, and lack of similarity to self-proteins. Given the sequence or the structure of a protein of interest, several methods exploit such features to predict the residues that are more likely to be recognized by an immunoglobulin. Here, we present two methods (BepiPred and DiscoTope) to predict linear and discontinuous antibody epitopes from the sequence and/or the three-dimensional structure of a target protein.
Towards High-throughput Immunomics for Infectious Diseases: Use of Next-generation Peptide Microarrays for Rapid Discovery and Mapping of Antigenic Determinants

Complete characterization of antibody specificities associated to natural infections is expected to provide a rich source of serologic biomarkers with potential applications in molecular diagnosis, follow-up of chemotherapeutic treatments, and prioritization of targets for vaccine development. Here, we developed a highly-multiplexed platform based on next-generation high-density peptide microarrays to map these specificities in Chagas Disease, an exemplar of a human infectious disease caused by the protozoan Trypanosoma cruzi. We designed a high-density peptide microarray containing more than 175,000 overlapping 15mer peptides derived from T. cruzi proteins. Peptides were synthesized in situ on microarray slides, spanning the complete length of 457 parasite proteins with fully overlapped 15mers (1 residue shift). Screening of these slides with antibodies purified from infected patients and healthy donors demonstrated both a high technical reproducibility as well as epitope mapping consistency when compared with earlier low-throughput technologies. Using a conservative signal threshold to classify positive (reactive) peptides we identified 2,031 disease-specific peptides and 97 novel parasite antigens, effectively doubling the number of known antigens and providing a tenfold increase in the number of fine mapped antigenic determinants for this disease. Finally, further analysis of the chip data showed that optimizing the amount of sequence overlap of displayed peptides can increase the protein space covered in a single chip by at least ~3 fold without sacrificing sensitivity. In conclusion, we show the power of high-density peptide chips for the discovery of pathogen-specific linear B-cell epitopes from clinical samples, thus setting the stage for high-throughput biomarker discovery screenings and proteome-wide studies of immune responses against pathogens.
Protein kinases control cellular responses to environmental cues by swift and accurate signal processing. Breakdowns in this high-fidelity capability are a driving force in cancer and other diseases. Thus, our limited understanding of which amino acids in the kinase domain encode substrate specificity, the so-called determinants of specificity (DoS), constitutes a major obstacle in cancer signaling. Here, we systematically discover several DoS and experimentally validate three of them, named the αC1, αC3, and APE-7 residues. We demonstrate that DoS form sparse networks of non-conserved residues spanning distant regions. Our results reveal a likely role for inter-residue allostery in specificity and an evolutionary decoupling of kinase activity and specificity, which appear loaded on independent groups of residues. Finally, we uncover similar properties driving SH2 domain specificity and demonstrate how the identification of DoS can be utilized to elucidate a greater understanding of the role of signaling networks in cancer (Creixell et al., 2015 [this issue of Cell]).
Characterization of binding specificities of bovine leucocyte class I molecules: impacts for rational epitope discovery.

The binding of peptides to classical major histocompatibility complex (MHC) class I proteins is the single most selective step in antigen presentation. However, the peptide-binding specificity of cattle MHC (bovine leucocyte antigen, BoLA) class I (BoLA-I) molecules remains poorly characterized. Here, we demonstrate how a combination of high-throughput assays using positional scanning combinatorial peptide libraries, peptide dissociation, and peptide-binding affinity binding
measurements can be combined with bioinformatics to effectively characterize the functionality of BoLA-I molecules. Using this strategy, we characterized eight BoLA-I molecules, and found the peptide specificity to resemble that of human MHC-I molecules with primary anchors most often at P2 and P9, and occasional auxiliary P1/P3/P5/P6 anchors. We analyzed nine reported CTL epitopes from *Theileria parva*, and in eight cases, stable and high affinity binding was confirmed. A set of peptides were tested for binding affinity to the eight BoLA proteins and used to refine the predictors of peptide-MHC binding NetMHC and NetMHCpan. The inclusion of BoLA-specific peptide-binding data led to a significant improvement in prediction accuracy for reported *T. parva* CTL epitopes. For reported CTL epitopes with weak or no predicted binding, these refined prediction methods suggested presence of nested minimal epitopes with high-predicted binding affinity. The enhanced affinity of the alternative peptides was in all cases confirmed experimentally. This study demonstrates how biochemical high-throughput assays combined with immunoinformatics can be used to characterize the peptide-binding motifs of BoLA-I molecules, boosting performance of MHC peptide-binding prediction methods, and empowering rational epitope discovery in cattle.

**General information**

State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Immunological Bioinformatics, International Livestock Research Institute, University of Copenhagen, United States Department of Agriculture, University of Vermont
Authors: Hansen, A. M. (Ekstern), Rasmussen, M. (Ekstern), Svitek, N. (Ekstern), Harndahl, M. (Ekstern), Golde, W. T. (Ekstern), Barlow, J. (Ekstern), Nene, V. (Ekstern), Buus, S. (Intern), Nielsen, M. (Intern)
Number of pages: 16
Pages: 705-718
Publication date: 2014
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Immunogenetics
Volume: 66
Issue number: 12
ISSN (Print): 0093-7711
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.638 SJR 0.916
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.2 SJR 1.249 SNIP 0.716
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.573 SNIP 0.813 CiteScore 2.47
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.228 SNIP 0.823 CiteScore 2.28
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.303 SNIP 0.771 CiteScore 2.48
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.382 SNIP 0.943 CiteScore 2.91
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.32 SNIP 0.885 CiteScore 2.83
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.217 SNIP 0.848
Background: It is important to accurately determine the performance of peptide: MHC binding predictions, as this enables users to compare and choose between different prediction methods and provides estimates of the expected error rate. Two common approaches to determine prediction performance are cross-validation, in which all available data are iteratively split into training and testing data, and the use of blind sets generated separately from the data used to construct the predictive method. In the present study, we have compared cross-validated prediction performances generated on our last benchmark dataset from 2009 with prediction performances generated on data subsequently added to the Immune Epitope Database (IEDB) which served as a blind set. Results: We found that cross-validated performances systematically overestimated performance on the blind set. This was found not to be due to the presence of similar peptides in the cross-validation dataset. Rather, we found that small size and low sequence/affinity diversity of either training or blind datasets were associated with large differences in cross-validated vs. blind prediction performances. We use these findings to derive quantitative rules of how large and diverse datasets need to be to provide generalizable performance estimates. Conclusion: It has long been known that cross-validated prediction performance estimates often overestimate performance on independently generated blind set data. We here identify and quantify the specific factors contributing to this effect for MHC-I binding predictions. An increasing number of peptides for which MHC binding affinities are measured experimentally have been selected based on binding predictions and thus are less diverse than historic datasets sampling the entire sequence and affinity space, making them more difficult benchmark data sets. This has to be taken into account when comparing performance metrics between different benchmarks, and when deriving error estimates for predictions based on benchmark performance.

Dataset size and composition impact the reliability of performance benchmarks for peptide-MHC binding predictions

General information

State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Immunological Bioinformatics, La Jolla Institute for Allergy & Immunology, University of Copenhagen
Authors: Kim, Y. (Ekstern), Sidney, J. (Ekstern), Buus, S. (Ekstern), Sette, A. (Ekstern), Nielsen, M. (Intern), Peters, B. (Ekstern)
Number of pages: 9
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Different binding motifs of the celiac disease-associated HLA molecules DQ2.5, DQ2.2, and DQ7.5 revealed by relative quantitative proteomics of endogenous peptide repertoires

Celiac disease is caused by intolerance to cereal gluten proteins, and HLA-DQ molecules are involved in the disease pathogenesis by presentation of gluten peptides to CD4+ T cells. The α- or β-chain sharing HLA molecules DQ2.5, DQ2.2, and DQ7.5 display different risks for the disease. It was recently demonstrated that T cells of DQ2.5 and DQ2.2 patients recognize distinct sets of gluten epitopes, suggesting that these two DQ2 variants select different peptides for display. To explore whether this is the case, we performed a comprehensive comparison of the endogenous self-peptides bound to HLA-DQ molecules of B-lymphoblastoid cell lines. Peptides were eluted from affinity-purified HLA molecules of nine cell lines and subjected to quadrupole orbitrap mass spectrometry and MaxQuant software analysis. Altogether, 12,712 endogenous peptides were identified at very different relative abundances. Hierarchical clustering of normalized quantitative data demonstrated significant differences in repertoires of peptides between the three DQ variant molecules. The neural network-based method, NNAlign, was used to identify peptide-binding motifs. The binding motifs of DQ2.5 and DQ7.5 concurred with previously established binding motifs. The binding motif of DQ2.2 was strikingly different from that of DQ2.5 with position P3 being a major anchor having a preference for threonine and serine. This is notable as three recently identified epitopes of gluten recognized by T cells of DQ2.2 celiac patients harbor serine at position P3. This study demonstrates that relative quantitative comparison of endogenous peptides sampled from our protein metabolism by HLA molecules provides clues to understand HLA association with disease.
Identification and HLA-Tetramer-Validation of Human CD4(+) and CD8(+) T Cell Responses against HCMV Proteins IE1 and IE2

Human cytomegalovirus (HCMV) is an important human pathogen. It is a leading cause of congenital infection and a leading infectious threat to recipients of solid organ transplants as well as of allogeneic hematopoietic cell transplants. Moreover, it has recently been suggested that HCMV may promote tumor development. Both CD4(+) and CD8(+) T cell responses are important for long-term control of the virus, and adoptive transfer of HCMV-specific T cells has led to protection from reactivation and HCMV disease. Identification of HCMV-specific T cell epitopes has primarily focused on CD8(+) T cell responses against the pp65 phosphoprotein. In this study, we have focused on CD4(+) and CD8(+) T cell responses against the immediate early 1 and 2 proteins (IE1 and IE2). Using overlapping peptides spanning the entire IE1 and IE2 sequences, peripheral blood mononuclear cells from 16 healthy, HLA-typed, donors were screened by ex vivo IFN-gamma ELISpot and in vitro intracellular cytokine secretion assays. The specificities of CD4(+) and CD8(+) T cell responses were identified and validated by HLA class II and I tetramers, respectively. Eighty-one CD4(+) and 44 CD8(+) T cell responses were identified representing at least seven different CD4 epitopes and 14 CD8 epitopes restricted by seven and 11 different HLA class II and I molecules, respectively, in total covering 91 and 98% of the Caucasian population, respectively. Presented in the context of several different HLA class II molecules, two epitope areas in IE1 and IE2 were recognized in about half of the analyzed donors. These data may be used to design a versatile anti-HCMV vaccine and/or immunotherapy strategy.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Copenhagen University Hospital , University of Copenhagen
Authors: Braendstrup, P. (Ekstern), Mortensen, B. K. (Ekstern), Justesen, S. F. L. (Ekstern), Osterby, T. (Ekstern), Rasmussen, M. (Ekstern), Hansen, A. M. (Ekstern), Christiansen, C. B. (Ekstern), Hansen, M. B. (Ekstern), Nielsen, M. (Intern), Vindelov, L. (Ekstern), Buus, S. (Ekstern), Stryhn, A. (Ekstern)
Number of pages: 16
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Journal: PLOS ONE
Volume: 9
Issue number: 4
Article number: e94892
ISSN (Print): 1932-6203
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.111 SJR 1.164
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.11 SJR 1.236 SNIP 1.101
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.427 SNIP 1.136 CiteScore 3.32
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.559 SNIP 1.148 CiteScore 3.54
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.772 SNIP 1.153 CiteScore 3.94
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.982 SNIP 1.156 CiteScore 4.15
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Immunoinformatics of Placental Malaria Vaccine Development

Malaria is an infectious disease caused by a protozoan parasite of the genus Plasmodium, which is transferred by female Anopheles mosquitos. WHO estimates that in 2012 there were 207 million cases of malaria, of which 627,000 were fatal. People living in malaria-endemic areas, gradually acquire immunity with multiple infections. Placental malaria (PM) is caused by P. falciparum sequestering in the placenta of pregnant women due to the presence of novel receptors in the placenta. An estimated 200,000 infants die a year as a result of PM. In 2004 the specific protein responsible for the pathogenesis of PM was identified as the P. falciparum Erythrocyte Membrane Protein 1 (PFEMP1) variant VAR2CSA. VAR2CSA is the leading candidate for a vaccine against PM.

The thesis is divided into 4 parts, where part I provide the reader with an introduction and background for the subjects covered in the thesis. Part II presents the first paper: "SigniSite: Identification of residue-level genotype-phenotype correlations in protein multiple sequence alignments". SigniSite is based on a non-parametric statistical evaluation of the positional distribution of amino acid residues in a multiple sequence alignment (MSA), thereby quantifying residue association to MSA phenotype. SigniSite was found to outperform comparable state-of-the-art methods. Furthermore part II addresses the issue of controlling type I and type II error probabilities in multiple testing scenarios and lastly the Dissertation advisor: Prof. Ole Lund Leon Eyrich Jessen analysis of the MHCI:peptide binding interaction by application of the SigniSite method. Part III presents the second paper: "Insight into Antigenic Diversity of VAR2CSA-DBL5ε Domain from Multiple Plasmodium falciparum Placental Isolates". The data consisted of 70 VAR2CSA-DBL5ε sequences each with associated phenotypes. Immunity towards PM is gradually acquired, therefore if a given sequence motif can be phenotype-correlated then the motif may be involved in VAR2CSA immunogenecity. Motifs defining VAR2CSA immunogenecity are naturally interesting in vaccine development context. The motif 'TFKNI' was found to be correlated with the birth weight of the child. Part IV presents the development of two methods for analysis of high-throughput data from a novel High Density Peptide Microarray (HDPMa) chip technology. Subsequently the HDPMa chip is applied for the discovery of linear B-cell VAR2CSA epitopes. Peptides 'GMDEFKNTFKNIKE' and 'SCGSARTMKRGYKNDNYELCKYC' were identified as linear B-cell epitopes. The latter subsequently experimentally found to be highly immunogenic, but not capable of blocking VAR2CSA:receptor interaction.

In summary, the work described in this thesis centres around the development and application of bioinformatics tools for in silico analysis of VAR2CSA, with an emphasis on statistical methodology. It is the hope of the author that the tools, developed, presented and applied in this thesis, may serve as an offset for further research and development in the field of placental malaria vaccine development.

General information
State: Published
Organisations: Department of Systems Biology, Immunological Bioinformatics, Agricultural and Environmental Proteomics, Center for Biological Sequence Analysis, University of Copenhagen
Authors: Jessen, L. E. (Intern), Lund, O. (Intern), Hviid, L. (Ekstern), Nielsen, M. (Intern)
Number of pages: 239
Publication date: 2014
Improved pan-specific MHC class I peptide-binding predictions using a novel representation of the MHC-binding cleft environment

Major histocompatibility complex (MHC) molecules play a key role in cell-mediated immune responses presenting bounded peptides for recognition by the immune system cells. Several in silico methods have been developed to predict the binding affinity of a given peptide to a specific MHC molecule. One of the current state-of-the-art methods for MHC class I is NetMHCpan, which has a core ingredient for the representation of the MHC class I molecule using a pseudo-sequence representation of the binding cleft amino acid environment. New and large MHC-peptide-binding data sets are constantly being made available, and also new structures of MHC class I molecules with a bound peptide have been published. In order to test if the NetMHCpan method can be improved by integrating this novel information, we created new pseudo-sequence definitions for the MHC-binding cleft environment from sequence and structural analyses of different MHC data sets including human leukocyte antigen (HLA), non-human primates (chimpanzee, macaque and gorilla) and other animal alleles (cattle, mouse and swine). From these constructs, we showed that by focusing on MHC sequence positions found to be polymorphic across the MHC molecules used to train the method, the NetMHCpan method achieved a significant increase in the predictive performance, in particular, of non-human MHCs. This study hence showed that an improved performance of MHC-binding methods can be achieved not only by the accumulation of more MHC-peptide-binding data but also by a refined definition of the MHC-binding environment including information from non-human species.
MR1-restricted MAIT cells display ligand discrimination and pathogen selectivity through distinct T cell receptor usage. Mucosal-associated invariant T (MAIT) cells express a semi-invariant T cell receptor (TCR) that detects microbial metabolites presented by the nonpolymorphic major histocompatibility complex (MHC)-like molecule MR1. The highly conserved nature of MR1 in conjunction with biased MAIT TCRα chain usage is widely thought to indicate limited ligand presentation and discrimination within a pattern-like recognition system. Here, we evaluated the TCR repertoire of MAIT cells responsive to three classes of microbes. Substantial diversity and heterogeneity were apparent across the functional MAIT cell repertoire as a whole, especially for TCRβ chain sequences. Moreover, different pathogen-specific responses were characterized by distinct TCR usage, both between and within individuals, suggesting that MAIT cell adaptation was a direct consequence of exposure to various exogenous MR1-restricted epitopes. In line with this interpretation, MAIT cell clones with distinct TCRs responded differentially to a riboflavin metabolite. These results suggest that MAIT cells can discriminate between pathogen-derived ligands in a clonotype-dependent manner, providing a basis for adaptive memory via recruitment of specific repertoires shaped by microbial exposure.
NetTepi: an integrated method for the prediction of T cell epitopes

Multiple factors determine the ability of a peptide to elicit a cytotoxic T cell lymphocyte response. Binding to a major histocompatibility complex class I (MHC-I) molecule is one of the most essential factors, as no peptide can become a T cell epitope unless presented on the cell surface in complex with an MHC-I molecule. As such, peptide-MHC (pMHC) binding affinity predictors are currently the premier methods for T cell epitope prediction, and these prediction methods have been shown to have high predictive performances in multiple studies. However, not all MHC-I binders are T cell epitopes, and multiple studies have investigated what additional factors are important for determining the immunogenicity of a peptide. A recent study suggested that pMHC stability plays an important role in determining if a peptide can become a T cell epitope. Likewise, a T cell propensity model has been proposed for identifying MHC binding peptides with amino acid compositions favoring T cell receptor interactions. In this study, we investigate if improved accuracy for T cell epitope discovery can be achieved by integrating predictions for pMHC binding affinity, pMHC stability, and T cell propensity. We show that a weighted sum approach allows pMHC stability and T cell propensity predictions to enrich pMHC binding affinity predictions. The integrated model leads to a consistent and significant increase in predictive performance and we demonstrate how this can be utilized to decrease the experimental workload of epitope screens.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis
Authors: Trolle, T. (Intern), Nielsen, M. (Intern)
Number of pages: 8
Pages: 449-456
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication Information
Journal: Immunogenetics
Volume: 66
Issue number: 7-8
ISSN (Print): 0093-7711
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.638 SJR 0.916
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.2 SJR 1.249 SNIP 0.716
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.573 SNIP 0.813 CiteScore 2.47
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.228 SNIP 0.823 CiteScore 2.28
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
The interplay of sequence conservation and T cell immune recognition

Predicting which peptides can elicit a T cell response (i.e. are immunogenic) is of great importance for many immunological studies. While it is clear that MHC binding is a necessary requirement for peptide immunogenicity, other variables exist that are incompletely understood. In this study we examined the hypothesis that conservation of a peptide in bacteria that are part of the healthy human microbiome leads to a reduced level of immunogenicity due to tolerization of T cells to the commensal bacteria. This was done by comparing experimentally characterized T cell epitope recognition data from the Immune Epitope Database with their conservation in the human microbiome. Indeed, we did see a lower immunogenicity for conserved peptides conserved. While many aspects how this conservation comparison is done require further optimization, this is a first step towards a better understanding T cell recognition of peptides in bacterial pathogens is influenced by their conservation in commensal bacteria. If the further work proves that this approach is successful, the degree of overlap of a peptide with the human proteome or microbiome could be added to the arsenal of tools available to assess peptide immunogenicity.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, La Jolla Institute for Allergy & Immunology
Uncovering the Peptide-Binding Specificities of HLA-C: A General Strategy To Determine the Specificity of Any MHC Class I Molecule

MHC class I molecules (HLA-I in humans) present peptides derived from endogenous proteins to CTLs. Whereas the peptide-binding specificities of HLA-A and -B molecules have been studied extensively, little is known about HLA-C specificities. Combining a positional scanning combinatorial peptide library approach with a peptide-HLA-I dissociation assay, in this study we present a general strategy to determine the peptide-binding specificity of any MHC class I molecule. We applied this novel strategy to 17 of the most common HLA-C molecules, and for 16 of these we successfully generated matrices representing their peptide-binding motifs. The motifs prominently shared a conserved C-terminal primary anchor with hydrophobic amino acid residues, as well as one or more diverse primary and auxiliary anchors at P1, P2, P3, and/or P7. Matrices were used to generate a large panel of HLA-C-specific peptide-binding data and update our pan-specific NetMHCpan predictor, whose predictive performance was considerably improved with respect to peptide binding to HLA-C. The updated predictor was used to assess the specificities of HLA-C molecules, which were found to cover a more limited sequence space than HLA-A and -B molecules. Assessing the functional significance of these new tools, HLA-C*07:01 transgenic mice were immunized with stable HLA-C*07:01 binders; six of six tested stable peptide binders were immunogenic. Finally, we generated HLA-C tetramers and labeled human CD8(+) T cells and NK cells. These new resources should support future research on the biology of HLA-C molecules. The data are deposited at the Immune Epitope Database, and the updated NetMHCpan predictor is available at the Center for Biological Sequence Analysis and the Immune Epitope Database.
Use of "one-pot, mix-and-read" peptide-MHC class I tetramers and predictive algorithms to improve detection of cytotoxic T lymphocyte responses in cattle

Peptide-major histocompatibility complex (p-MHC) class I tetramer complexes have facilitated the early detection and functional characterisation of epitope specific CD8(+) cytotoxic T lymphocytes (CTL). Here, we report on the generation of seven recombinant bovine leukocyte antigens (BoLA) and recombinant bovine beta 2-microglobulin from which p-MHC class I tetramers can be derived in similar to 48 h. We validated a set of p-MHC class I tetramers against a panel of CTL lines specific to seven epitopes on five different antigens of Theileria parva, a protozoan pathogen causing the lethal bovine disease East Coast fever. One of the p-MHC class I tetramers was tested in ex vivo assays and we detected T. parva specific CTL in peripheral blood of cattle at day 15-17 post-immunization with a live parasite vaccine. The algorithm NetMHCpan predicted alternative epitope sequences for some of the T. parva CTL epitopes. Using an ELISA assay to measure peptide-BoLA monomer formation and p-MHC class I tetramers of new specificity, we demonstrate that a predicted alternative epitope Tp2(29-37) rather than the previously reported Tp2(27-37) epitope is the correct Tp2 epitope presented by BoLA-6*04101. We also verified the prediction by NetMHCpan that the Tp5(87-95) epitope reported as BoLA-T5 restricted can also be presented by BoLA-1*02301, a molecule similar in sequence to BoLA-T5. In addition, Tp5(87-95) specific bovine CTL were simultaneously stained by Tp5-BoLA-1*02301 and Tp5-BoLA-T5 tetramers suggesting that one T cell receptor can bind to two different BoLA MHC class I molecules presenting the Tp5(87-95) epitope and that these BoLA molecules fall into a single functional supertype.
Liabilities Identification of Antigenic Peptide: Predicting the Specificity of Major MHC Class I and II Pathway Players.

Bioinformatics methods for immunology have become increasingly used over the last decade and now form an integrated part of most epitope discovery projects. This wide usage has led to the confusion of defining which of the many methods to use for what problems. In this chapter, an overview is given focusing on the suite of tools developed at the Technical University of Denmark.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis
Authors: Lund, O. (Intern), Karosiene, E. (Intern), Lundegaard, C. (Intern), Larsen, M. V. (Intern), Nielsen, M. (Intern)
Pages: 247-260
Publication date: 2013

Host publication information
Title of host publication: Antigen Processing : Methods and Protocols
ISBN (Print): 978-1-62703-217-9
ISBN (Electronic): 978-1-62703-218-6
Chapter: 19

Series: Methods in Molecular Biology
Volume: 960
ISSN: 1064-3745
Epitope prediction methods

Major histocompatibility complex (MHC) molecules play a crucial role in adaptive immunity by sampling peptides from self and non-self proteins to be recognised by the immune system. MHC molecules present peptides on cell surfaces for recognition by CD8⁺ and CD4⁺ T lymphocytes that can initiate immune responses. Therefore, it is of great importance to be able to identify peptides that bind to MHC molecules, in order to understand the nature of immune responses and discover T cell epitopes useful for designing new vaccines and immunotherapies. MHC molecules in humans, referred to as human leucocyte antigen (HLA) molecules, are encoded by extremely polymorphic genes on chromosome 6. Due to this polymorphism, thousands of different MHC molecules exist, making the experimental identification of peptide-MHC interactions a very costly procedure. This has primed the need for in silico peptide-MHC prediction methods, and over the last decade several such methods have been successfully developed and used for epitope discovery purposes. My PhD project has been dedicated to improve methods for predicting peptide-MHC interactions by developing new strategies for training prediction algorithms based on machine learning techniques. Several MHC class I binding prediction algorithms have been developed and due to their high accuracy they are used by many immunologists to facilitate the conventional experimental process of epitope discovery. However, the accuracy of these methods depends on data defining the MHC molecule in question, making it difficult for the non-expert end-user to choose the most suitable predictor. The first paper in this thesis presents a new, publicly available, consensus method for MHC class I predictions. The NetMHCcons predictor combines three state-of-the-art prediction tools and provides the most accurate predictions for any given MHC molecule. While the methods for MHC class I binding have reached a very high accuracy and are widely used for immunological research, the case of MHC class II is less clear. The open binding groove of MHC class II molecules and differences in polymorphism among MHC encoding genes makes predictions of peptide binding to MHC class II molecules a complicated problem. We addressed these issues in order to develop the first pan-specific predictor common for all three human class II isotypes, HLA-DR, HLA-DP and HLA-DQ. The second paper introduces the NetMHCIIpan-3.0 predictor based on artificial neural networks, which is capable of giving binding affinities to any human MHC class II molecule. Chapter 4 of this thesis gives an overview of bioinformatics tools developed by the Immunological Bioinformatics group at Center for Biological Sequence Analysis. The chapter provides detailed explanations on how to use different methods for T cell epitope discovery research, explaining how input should be given as well as how to interpret the output. In the last chapter, I present the results of a bioinformatics analysis of epitopes from the yellow fever virus. The analysis demonstrated the absence of distinct regions of higher epitope density within the virus polyprotein. Also, the density of epitopes among different proteins was demonstrated to mostly depend on protein length and amino acid composition, underlining the importance of identifying peptide-MHC interactions. Furthermore, using yellow fever virus epitopes, we demonstrated the power of the %Rank score when compared with the binding affinity score of MHC prediction methods, suggesting that this score should be considered to be used for selecting potential T cell epitopes. In summary, this thesis presents methods for prediction of peptides that bind to both MHC class I and class II molecules, which is important for driving immunological research within the field of T cell epitope discovery and for general understanding of the cellular responses.

Evaluation of peptide selection approaches for epitope-based vaccine design

A major challenge in epitope-based vaccine (EV) design stems from the vast genomic variation of pathogens and the diversity of the host cellular immune system. Several computational approaches have been published to assist the selection of potential T cell epitopes for EV design. So far, no thorough comparison between the current methods has
been realized. Using human immunodeficiency virus as test case, different EV selection algorithms were evaluated with respect to their ability to select small peptides sets with broad coverage of allelic and pathogenic diversity. The methods were compared in terms of in silico measurements simulating important vaccine properties like the ability of inducing protection against a multivariant pathogen in a population; the predicted immunogenicity; pathogen, allele, and population coverage; as well as the conservation of selected epitopes. Additionally, we evaluate the use of human leukocyte antigen (HLA) supertypes with regards to their applicability for population-spanning vaccine design. The results showed that in terms of induced protection methods that simultaneously aim to optimize pathogen and HLA coverage significantly outperform methods focusing on pathogen coverage alone. Moreover, supertype-based approaches for coverage of HLA diversity were showed to yield only satisfying results in populations in which the supertype representatives are prevalent.
From viral genome to specific peptide epitopes: methods for identifying porcine T cell epitopes based on in silico predictions, in vitro identification and ex vivo verification

The affinity with which major histocompatibility complex (MHC) class I molecules bind peptides is instrumental to presentation of viral epitopes to cytotoxic T lymphocytes (CTLs). We analyzed three swine leukocyte antigen (SLA) molecules for complete nonamer peptide-based binding matrices in order to predict likely candidates for peptide-SLA binding. These results were combined with binding predictions generated by the algorithm, NetMHCpan (http://www.cbs.dtu.dk/services/NetMHCpan/) in order to select peptide candidates for in vitro analysis. The correlation between high affinity and high stability was determined using luminescence oxygen channeling (LOCI) and scintillation proximity assay (SPA) for peptides bound by the commonly occurring SLA molecules, SLA-1*0401, SLA-2*0401, and SLA-3*0401. Using these tools, peptides bound by SLA with high affinity and stability were identified from a library of 10,000 peptides. T cell epitopes were identified using peptide-SLA complexes assembled into fluorescent tetramers to stain swine influenza specific CTLs derived from immunized animals and MHC-defined pigs vaccinated against foot-and-mouth disease virus. These results demonstrate the broad applicability of methods originally developed for analysis of human leukocyte antigen (HLA) presentation of peptides. The methods presented provide a timely and cost-effective approach to CTL epitope discovery that can be applied to diseases of swine and of other mammalian species of interest.

General information

State: Published
Organisations: National Veterinary Institute, Section for Immunology and Vaccinology, Department of Systems Biology, Center for Biological Sequence Analysis, Immunological Bioinformatics, Agricultural Research Service, University of Copenhagen
Authors: Pedersen, L. E. (Intern), Rasmussen, M. (Ekstern), Harndah, M. (Ekstern), Patch, J. (Ekstern), Nielsen, M. (Intern), Buus, S. (Ekstern), Golde, W. (Ekstern), Jungersen, G. (Intern)
Publication date: 2013
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Immunology
Volume: 190
ISSN (Print): 0022-1767
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): SNIP 1.112 SJR 2.837
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 4.79 SJR 3.474 SNIP 1.176
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 3.571 SNIP 1.26 CiteScore 5.05
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 3.744 SNIP 1.271 CiteScore 5.03
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 3.909 SNIP 1.35 CiteScore 5.61
ISI indexed (2013): ISI indexed yes
Identification of peptides from foot-and-mouth disease virus structural proteins bound by class I swine leukocyte antigen (SLA) alleles, SLA-1*0401 and SLA-2*0401

Characterization of the peptide-binding specificity of swine leukocyte antigen (SLA) class I and II molecules is critical to the understanding of adaptive immune responses of swine toward infectious pathogens. Here, we describe the complete binding motif of the SLA-2*0401 molecule based on a positional scanning combinatorial peptide library approach. By combining this binding motif with data achieved by applying the NetMHCpan peptide prediction algorithm to both SLA-1*0401 and SLA-2*0401, we identified high-affinity binding peptides. A total of 727 different 9mer and 726 different 10mer peptides within the structural proteins of foot-and-mouth disease virus (FMDV), strain A24 were analyzed as candidate T-cell epitopes. Peptides predicted by the NetMHCpan were tested in ELISA for binding to the SLA-1*0401 and SLA-2*0401 major histocompatibility complex class I proteins. Four of the 10 predicted FMDV peptides bound to SLA-2*0401, whereas five of the nine predicted FMDV peptides bound to SLA-1*0401. These methods provide the characterization of T-cell epitopes in response to pathogens in more detail. The development of such approaches to analyze vaccine performance will contribute to a more accelerated improvement of livestock vaccines by virtue of identifying and focusing analysis on bona fide T-cell epitopes.

General information
State: Published
Organisations: National Veterinary Institute, Section for Immunology and Vaccinology, Department of Systems Biology, Center for Biological Sequence Analysis, United States Department of Agriculture, University of Copenhagen
Authors: Pedersen, L. E. (Intern), Hamdahl, M. (Forskerdatabase), Nielsen, M. (Intern), Patch, J. R. (Ekstern), Jungersen, G. (Intern), Buus, S. (Ekstern), Golde, W. T. (Ekstern)
In silico peptide-binding predictions of passerine MHC class I reveal similarities across distantly related species, suggesting convergence on the level of protein function.
The major histocompatibility complex (MHC) genes are the most polymorphic genes found in the vertebrate genome, and they encode proteins that play an essential role in the adaptive immune response. Many songbirds (passerines) have been shown to have a large number of transcribed MHC class I genes compared to most mammals. To elucidate the reason for this large number of genes, we compared 14 MHC class I alleles (α1–α3 domains), from great reed warbler, house sparrow and tree sparrow, via phylogenetic analysis, homology modelling and in silico peptide-binding predictions to investigate their functional and genetic relationships. We found more pronounced clustering of the MHC class I allomorphs (allele specific proteins) in regards to their function (peptide-binding specificities) compared to their genetic relationships (amino acid sequences), indicating that the high number of alleles is of functional significance. The MHC class I allomorphs from house sparrow and tree sparrow, species that diverged 10 million years ago (MYA), had overlapping peptide-binding specificities, and these similarities across species were also confirmed in phylogenetic analyses based on amino acid sequences. Notably, there were also overlapping peptide-binding specificities in the allomorphs from house sparrow and great reed warbler, although these species diverged 30 MYA. This overlap was not found in a tree based on amino acid sequences. Our interpretation is that convergent evolution on the level of the protein function, possibly driven by selection from shared pathogens, has resulted in allomorphs with similar peptide-binding repertoires, although trans-species evolution in combination with gene conversion cannot be ruled out.
MHCcluster, a method for functional clustering of MHC molecules

The identification of peptides binding to major histocompatibility complexes (MHC) is a critical step in the understanding of T cell immune responses. The human MHC genomic region (HLA) is extremely polymorphic comprising several thousand alleles, many encoding a distinct molecule. The potentially unique specificities remain experimentally uncharacterized for the vast majority of HLA molecules. Likewise, for nonhuman species, only a minor fraction of the known MHC molecules have been characterized. Here, we describe a tool, MHCcluster, to functionally cluster MHC molecules based on their predicted binding specificity. The method has a flexible web interface that allows the user to include any MHC of interest in the analysis. The output consists of a static heat map and graphical tree-based visualizations of the functional relationship between MHC variants and a dynamic TreeViewer interface where both the functional relationship and the individual binding specificities of MHC molecules are visualized. We demonstrate that conventional sequence-based clustering will fail to identify the functional relationship between molecules, when applied to MHC system, and only through the use of the predicted binding specificity can a correct clustering be found. Clustering of prevalent HLA-A and HLA-B alleles using MHCcluster confirms the presence of 12 major specificity groups (supertypes) some however with highly divergent specificities. Importantly, some HLA molecules are shown not to fit any supertype classification. Also, we use MHCcluster to show that chimpanzee MHC class I molecules have a reduced functional diversity compared to that of HLA class I molecules.
MISTIC: mutual information server to infer coevolution

MISTIC (mutual information server to infer coevolution) is a web server for graphical representation of the information contained within a MSA (multiple sequence alignment) and a complete analysis tool for Mutual Information networks in protein families. The server outputs a graphical visualization of several information-related quantities using a circos representation. This provides an integrated view of the MSA in terms of (i) the mutual information (MI) between residue pairs, (ii) sequence conservation and (iii) the residue cumulative and proximity MI scores. Further, an interactive interface to explore and characterize the MI network is provided. Several tools are offered for selecting subsets of nodes from the network for visualization. Node coloring can be set to match different attributes, such as conservation, cumulative MI, proximity MI and secondary structure. Finally, a zip file containing all results can be downloaded. The server is available at http://mistic.leloir.org.ar. In summary, MISTIC allows for a comprehensive, compact, visually rich view of the information contained within an MSA in a manner unique to any other publicly available web server. In particular, the use of circos representation of MI networks and the visualization of the cumulative MI and proximity MI concepts is novel.
Monitoring Expansion of T Cell Specificities Against Foot-and-Mouth-Disease Virus (FMDV) in Swine With MHC Class I Tetramers Following a Prime/Boost Vaccination

**General information**

State: Published
Organisations: National Veterinary Institute, Section for Immunology and Vaccinology, Department of Systems Biology, Center for Biological Sequence Analysis, Immunological Bioinformatics
Authors: Pedersen, L. E. (Intern), Patch, J. (Ekstern), Glabman, R. (Ekstern), Kenney, M. (Ekstern), Nielsen, M. (Intern), Buus, S. (Ekstern), Jungersen, G. (Intern), Golde, W. T. (Ekstern)
Number of pages: 1
Pages: 72
Publication date: 2013

**Host publication information**

Title of host publication: 10th International Veterinary Immunology Symposium - IVIS 2013
Article number: CS05.5
Main Research Area: Technical/natural sciences
Conference: 10th International Veterinary Immunology Symposium, Milano, Italy, 28/08/2013 - 28/08/2013
Electronic versions: Pedersen_et_al_Abstract_10IVIS_Milan_2013.pdf
Publication: Research - peer-review › Conference abstract in proceedings – Annual report year: 2013

**NETMHCSTAB - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery**

Major histocompatibility complex class I (MHC-I) molecules play an essential role in the cellular immune response, presenting peptides to cytotoxic T lymphocytes (CTLs) allowing the immune system to scrutinize ongoing intracellular production of proteins. In the early 1990s, immunogenicity and stability of the peptide-MHC-I (pMHC-I) complex were
shown to be correlated. At that time, measuring stability was cumbersome and time consuming and only small data sets were analysed. Here, we investigate this fairly unexplored area on a large scale compared with earlier studies. A recent small-scale study demonstrated that pMHC-I complex stability was a better correlate of CTL immunogenicity than peptide-MHC-I affinity. We here extended this study and analysed a total of 5509 distinct peptide stability measurements covering 10 different HLA class I molecules. Artificial neural networks were used to construct stability predictors capable of predicting the half-life of the pMHC-I complex. These predictors were shown to predict T-cell epitopes and MHC ligands from SYFPEITHI and IEDB to form significantly more stable MHC-I complexes compared with affinity-matched non-epitopes. Combining the stability predictions with a state-of-the-art affinity predictions NetMHCcons significantly improved the performance for identification of T-cell epitopes and ligands. For the HLA alleles included in the study, we could identify distinct sub-motifs that differentiate between stable and unstable peptide binders and demonstrate that anchor positions in the N-terminal of the binding motif (primarily P2 and P3) play a critical role for the formation of stable pMHC-I complexes. A webserver implementing the method is available at www.cbs.dtu.dk/services/NetMHCstab.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Technical University of Denmark , University of Copenhagen
Authors: Jørgensen, K. W. (Ekstern), Rasmussen, M. (Ekstern), Buus, S. (Ekstern), Nielsen, M. (Intern)
Number of pages: 9
Pages: 18-26
Publication date: 2013
Main Research Area: Technical/natural sciences

Publication information
Journal: Immunology
Volume: 141
Issue number: 1
ISSN (Print): 0019-2805
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.938 SJR 1.69
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.74 SJR 1.964 SNIP 0.965
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 2.075 SNIP 0.965 CiteScore 3.83
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 2.048 SNIP 1.043 CiteScore 3.61
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 2.086 SNIP 1.084 CiteScore 3.97
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.941 SNIP 1.04 CiteScore 3.94
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.884 SNIP 0.992 CiteScore 3.75
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 0.121 SNIP 0.912
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
SigniSite: Identification of residue-level genotype-phenotype correlations in protein multiple sequence alignments

Identifying which mutation(s) within a given genotype is responsible for an observable phenotype is important in many aspects of molecular biology. Here, we present SigniSite, an online application for subgroup-free residue-level genotype–phenotype correlation. In contrast to similar methods, SigniSite does not require any pre-definition of subgroups or binary classification. Input is a set of protein sequences where each sequence has an associated real number, quantifying a given phenotype. SigniSite will then identify which amino acid residues are significantly associated with the data set phenotype. As output, SigniSite displays a sequence logo, depicting the strength of the phenotype association of each residue and a heat-map identifying ‘hot’ or ‘cold’ regions. SigniSite was benchmarked against SPEER, a state-of-the-art method for the prediction of specificity determining positions (SDP) using a set of human immunodeficiency virus protease-inhibitor genotype–phenotype data and corresponding resistance mutation scores from the Stanford University HIV Drug Resistance Database, and a data set of protein families with experimentally annotated SDPs. For both data sets, SigniSite was found to outperform SPEER. SigniSite is available at: http://www.cbs.dtu.dk/services/SigniSite/.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, University of Copenhagen
Authors: Jessen, L. I. (Intern), Hoof, I. (Ekstern), Lund, O. (Intern), Nielsen, M. (Intern)
Pages: 20W286-W291
Publication date: 2013
Main Research Area: Technical/natural sciences

Publication information
Journal: Nucleic Acids Research
Volume: 41
Issue number: W1
ISSN (Print): 0305-1048
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach

Motivation: Proteins recognizing short peptide fragments play a central role in cellular signaling. As a result of high-throughput technologies, peptide-binding protein specificities can be studied using large peptide libraries at dramatically
lower cost and time. Interpretation of such large peptide datasets, however, is a complex task, especially when the data contain multiple receptor binding motifs, and/or the motifs are found at different locations within distinct peptides. Results: The algorithm presented in this article, based on Gibbs sampling, identifies multiple specificities in peptide data by performing two essential tasks simultaneously: alignment and clustering of peptide data. We apply the method to de-convolute binding motifs in a panel of peptide datasets with different degrees of complexity spanning from the simplest case of pre-aligned fixed-length peptides to cases of unaligned peptide datasets of variable length. Example applications described in this article include mixtures of binders to different MHC class I and class II alleles, distinct classes of ligands for SH3 domains and sub-specificities of the HLA-A*02:01 molecule. Availability: The Gibbs clustering method is available online as a web server at http://www.cbs.dtu.dk/services/GibbsCluster. Contact: massimo@cbs.dtu.dk Supplementary information: Supplementary data are available at Bioinformatics online.
Structural analysis of B-cell epitopes in antibody:protein complexes

The binding of antigens to antibodies is one of the key events in an immune response against foreign molecules and is a critical element of several biomedical applications including vaccines and immunotherapeutics. For development of such applications, the identification of antibody binding sites (B-cell epitopes) is essential. However experimental epitope mapping is highly cost-intensive and computer-aided methods do in general have moderate performance. One major reason for this moderate performance is an incomplete understanding of what characterizes an epitope. To fill this gap, we here developed a novel framework for comparing and superimposing B-cell epitopes and applied it on a dataset of 107 non-similar antigen:antibody structures extracted from the PDB database. With the presented framework, we were able to describe the general B-cell epitope as a flat, oblong, oval shaped volume consisting of predominantly hydrophobic amino acids in the center flanked by charged residues. The average epitope was found to be made up of ~15 residues with one linear stretch of 5 or more residues constituting more than half of the epitope size. Furthermore, the epitope area is predominantly constrained to a plane above the antibody tip, in which the epitope is orientated in a −30° to 60° angle relative to the light to heavy chain antibody direction. Contrary to previously findings, we did not find a significant deviation between the amino acid composition in epitopes and the composition of equally exposed parts of the antigen surface. Our results, in combination with previously findings, give a detailed picture of the B-cell epitope that may be used in development of improved B-cell prediction methods.
CD4+ T cells orchestrate immunity against viral infections, but their importance in HIV infection remains controversial. Nevertheless, comprehensive studies have associated increase in breadth and functional characteristics of HIV-specific CD4+ T cells with decreased viral load. A major challenge for the identification of HIV-specific CD4+ T cells targeting broadly reactive epitopes in populations with diverse ethnic background stems from the vast genomic variation of HIV and the diversity of the host cellular immune system. Here, we describe a novel epitope selection strategy, PopCover, that aims to resolve this challenge, and identify a set of potential HLA class II-restricted HIV epitopes that in concert will provide optimal viral and host coverage. Using this selection strategy, we identified 64 putative epitopes (peptides) located in the Gag, Nef, Env, Pol and Tat protein regions of HIV. In total, 73% of the predicted peptides were found to induce HIV-specific CD4+ T cell responses. The Gag and Nef peptides induced most responses. The vast majority of the peptides (93%) had predicted restriction to the patient's HLA alleles. Interestingly, the viral load in viremic patients was inversely correlated to the number of targeted Gag peptides. In addition, the predicted Gag peptides were found to induce broader polyfunctional CD4+ T cell responses compared to the commonly used Gag-p55 peptide pool. These results demonstrate the power of the PopCover method for the identification of broadly recognized HLA class II-restricted epitopes. All together, selection strategies, such as PopCover, might with success be used for the evaluation of antigen-specific CD4+ T cell responses and design of future vaccines.
Characterizing the binding motifs of 11 common human HLA-DP and HLA-DQ molecules using NNAlign

Compared with HLA-DR molecules, the specificities of HLA-DP and HLA-DQ molecules have only been studied to a limited extent. The description of the binding motifs has been mostly anecdotal and does not provide a quantitative measure of the importance of each position in the binding core and the relative weight of different amino acids at a given position. The recent publication of larger data sets of peptide-binding to DP and DQ molecules opens the possibility of using data-driven bioinformatics methods to accurately define the binding motifs of these molecules. Using the neural network-based method NNAlign, we characterized the binding specificities of five HLA-DP and six HLA-DQ among the most frequent in the human population. The identified binding motifs showed an overall concurrence with earlier studies but revealed subtle differences. The DP molecules revealed a large overlap in the pattern of amino acid preferences at core positions, with conserved hydrophobic/aromatic anchors at P1 and P6, and an additional hydrophobic anchor at P9 in some variants. These results confirm the existence of a previously hypothesized supertype encompassing the most common DP alleles. Conversely, the binding motifs for DQ molecules appear more divergent, displaying unconventional anchor positions and in some cases rather unspecific amino acid preferences.
Describing the Peptide Binding Specificity of HLA-C

Human leukocyte antigen (HLA) presents peptides to T-cells for immune scrutiny. Whereas HLA-A and -B have been described in great detail, HLA-C has received much less attention. Here, to increase the coverage of HLA-C and the accuracy of the corresponding tools, we have generated HLA-C molecules; peptide-binding assays, data and predictors; and tetramers; representing the most prevalent HLA-C molecules. We have combined positional scanning combinatorial peptide library (PSCPL) with a homogenous high-throughput dissociation assay and generated specificity matrices for 11 different HLA-C molecules. We find preference for hydrophobic residues at the peptide C-terminus for all HLA-C molecules. Most molecules were found to have an additional strong anchor at P2 or P3, with auxiliary anchor observed at P1, P2, P3, and P7. The binding affinity is measured for peptides fitting the specificity matrix for 5 HLA-C molecules and for all, but one, molecule we find a high frequency of binders, >70%, among these peptides. To extend the examined peptide space, we use bioinformatic prediction tools to search for additional binders. Finally, we update our prediction tool, NetMHCpan, with the HLA-C affinity data and show that the predictive performance for HLA-C molecules now is increased to a level comparable with that of HLA-A and -B. These novel HLA-C molecules and predictors are successfully used to generate HLA-C tetramers and validate HLA-C-restricted T cell responses.

General information

State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, University of Copenhagen
Authors: Rasmussen, M. (Ekstern), Harndahl, M. N. (Forskerdatabase), Nielsen, M. (Intern), Buus, S. (Ekstern), Stryhn, A. (Ekstern)
Publication date: 2012
Main Research Area: Technical/natural sciences
Source: dtu
Source-ID: n:oai:DTIC-ART:isi/365995138::17115
Publication: Research - peer-review › Conference abstract for conference – Annual report year: 2012

Designing bovine T cell vaccines via reverse immunology

T cell responses contribute to immunity against many intracellular infections. There is, for example, strong evidence that major histocompatibility complex (MHC) class I-restricted cytotoxic T lymphocytes (CTLs) play an essential role in mediating immunity to East Coast fever (ECF), a fatal lymphoproliferative disease of cattle prevalent in sub-Saharan Africa and caused by Theileria parva. To complement the more traditional approaches to CTL antigen identification and vaccine development that we have previously undertaken we propose a use of immunoinformatics to predict CTL peptide epitopes followed by experimental verification of T cell specificity to candidate epitopes using peptide–MHC (pMHC) tetramers. This system, adapted from human and rodent studies, is in the process of being developed for cattle. Briefly, we have used an artificial neural network called NetMHCpan, which has been trained mainly on existing human, mouse, and non-human
primate MHC–peptide binding data in an attempt to predict the peptide-binding specificity of bovine MHC class I molecules. Our data indicate that this algorithm needs to be further optimized by incorporation of bovine MHC–peptide binding data. When retrained, NetMHCpan may be used to predict parasite peptide epitopes by scanning the predicted T. parva proteome and known parasite CTL antigens. A range of pMHC tetramers, made “on-demand”, will then be used to assay cattle that are immune to ECF or in vaccine trials to determine if CTLs of the predicted epitope specificity are present or not. Thus, pMHC tetramers can be used in one step to identify candidate CTL antigens and to map CTL epitopes. Our current research focuses on 9 different BoLA class I molecules. By expanding this repertoire to include the most common bovine MHCs, these methods could be used as generic assays to predict and measure bovine T cell immune responses to any pathogen.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, International Livestock Research Institute, Agricultural Research Service, University of Vermont, University of Copenhagen
Authors: Nene, V. (Ekstern), Svitek, N. (Ekstern), Toye, P. (Ekstern), Golde, W. T. (Ekstern), Barlow, J. (Ekstern), Harndahl, M. N. (Forskerdatabase), Buus, S. (Ekstern), Nielsen, M. (Intern)
Pages: 188-192
Publication date: 2012
Main Research Area: Technical/natural sciences

Publication information
Journal: Ticks and Tick-borne Diseases
Volume: 3
Issue number: 3
ISSN (Print): 1877-959X
Ratings:
Web of Science (2018): Indexed yes
Scopus rating (2017): SNIP 1.057 SJR 1.421
Web of Science (2017): Indexed Yes
Scopus rating (2016): SJR 1.308 SNIP 1.251 CiteScore 3.15
Scopus rating (2015): SJR 1.248 SNIP 1.059 CiteScore 2.75
Scopus rating (2014): SJR 1.011 SNIP 1.095 CiteScore 2.35
Scopus rating (2013): SJR 1.04 SNIP 0.942 CiteScore 2.63
ISI indexed (2013): ISI indexed yes
Scopus rating (2012): SJR 0.788 SNIP 1.177 CiteScore 2.09
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
Scopus rating (2011): SJR 0.578 SNIP 1.154 CiteScore 1.91
ISI indexed (2011): ISI indexed no
Original language: English
Theileria parva, Cytotoxic T lymphocytes, Epitope prediction, Peptide–MHC tetramers, Vaccine development
DOIs:
10.1016/j.ttbdis.2011.12.001
Source: dtu
Source-ID: n::oai:DTIC-ART:elsevier/367543058::18371
Publication: Research - peer-review › Journal article – Annual report year: 2012

Discovering sequence motifs in quantitative and qualitative peptide data
Proteins are central to virtually all processes within the cell. The vast amount of functions performed by proteins in biological processes is conferred by their ability to bind in a selective and specific manner to other molecules. The nature of these interactions is, in general terms, three-dimensional, as binding sites normally consist of a pocket or a groove on the protein surface. However, in many cases such interactions contain a linear component and can be more conveniently represented, or approximated, by a protein-peptide interaction. Whereas time-consuming structural studies are necessary in systems where the three-dimensional aspect of the interaction is prevalent, protein-peptide interactions can normally be represented simply by a linear binding motif. Phage display and peptide microarray technologies allow generating large libraries of peptide sequences and the parallel detection of thousands of interactions in a single experiment, with virtually unlimited choice of potential targets and variants of these targets. However, the amount and complexity of data produced by high-throughput techniques poses serious challenges to researchers of limited bioinformatics expertise who need to analyze and interpret such data. The first paper in this thesis presents a new, publicly available method based on artificial neural networks that allows custom analysis of quantitative peptide data. The online NNAAlign web-server provides a simple yet powerful tool for the discovery of sequence motifs in large-scale peptide data sets. It was successfully applied to characterize the binding motifs of MHC class I and class II molecules, and for the prediction of protease cleavage on data generated by a large-scale peptide microarray technology.
In the second paper, NNAlign was applied to binding data for HLA-DP and DQ molecules, two classes of HLA molecules with recognized importance in immune response but poorly characterized sequence motifs. The sequence logos of 5 HLADP and 6 HLA-DQ molecules provide a characterization of their binding motifs at an unprecedented level of detail. The third paper in this thesis deals with the presence of multiple motifs, due to the experimental setup or the actual poly-specificity of the receptor, in peptide data. A new algorithm, based on Gibbs sampling, identifies multiple specificities by performing two tasks simultaneously: alignment and clustering of peptide data. The method, available online as a web-server, was applied to various data sets including mixtures of MHC binding data and distinct classes of ligands to SH3 domains.

Next, we investigated how string kernels could be used to identify pattern in peptide data, with particular focus on the MHC class I system. We suggest a strategy that, unlike most available methods, allows to learn from peptides of multiple lengths to achieve improved predictive performance. This appeared particularly important in alleles and peptide lengths where experimental data was limited.

The last chapter presents a method to rationally guide the discovery of T-cell epitopes from ELISPOT and ICS assays based on peptide pool matrices. By prediction of binding affinity, analysis of peptide pools intersections, and combination of information from different donors, we show that the method can effectively rank potential epitope candidates and reduce the number of experimental tests needed to identify new epitopes.

Taken as a whole, this thesis provides a valuable series of algorithms and tools for the analysis of peptide data, both from the point of view of characterization of sequence motifs and the prediction of protein-peptide interactions.

General information
State: Published
Organisations: Department of Systems Biology
Authors: Andreatta, M. (Intern), Nielsen, M. (Intern), Lund, O. (Intern)
Number of pages: 105
Publication date: 2012

Disentangling evolutionary signals: conservation, specificity determining positions and coevolution. Implication for catalytic residue prediction
Background: A large panel of methods exists that aim to identify residues with critical impact on protein function based on evolutionary signals, sequence and structure information. However, it is not clear to what extent these different methods overlap, and if any of the methods have higher predictive potential compared to others when it comes to, in particular, the identification of catalytic residues (CR) in proteins. Using a large set of enzymatic protein families and measures based on different evolutionary signals, we sought to break up the different components of the information content within a multiple sequence alignment to investigate their predictive potential and degree of overlap. Results: Our results demonstrate that the different methods included in the benchmark in general can be divided into three groups with a limited mutual overlap. One group containing real-value Evolutionary Trace (rvET) methods and conservation, another containing mutual information (MI) methods, and the last containing methods designed explicitly for the identification of specificity determining positions (SDPs): integer-value Evolutionary Trace (ivET), SDPfox, and XDET. In terms of prediction of CR, we find using a proximity score integrating structural information (as the sum of the scores of residues located within a given distance of the residue in question) that only the methods from the first two groups displayed a reliable performance. Next, we investigated to what degree proximity scores for conservation, rvET and cumulative MI (cMI) provide complementary information capable of improving the performance for CR identification. We found that integrating conservation with proximity scores for rvET and cMI achieved the highest performance. The proximity conservation score contained no complementary information when integrated with proximity rvET. Moreover, the signal from rvET provided only a limited gain in predictive performance when integrated with mutual information and conservation proximity scores. Combined, these observations demonstrate that the rvET and cMI scores add complementary information to the prediction system. Conclusions: This work contributes to the understanding of the different signals of evolution and also shows that it is possible to improve the detection of catalytic residues by integrating structural and higher order sequence evolutionary information with sequence conservation.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Fundación Instituto Leloir, Baylor College of Medicine
Authors: Teppa, E. (Ekstern), Wilkins, A. D. (Ekstern), Nielsen, M. (Intern), Marino Buslje, C. (Ekstern)
Pages: -
Publication date: 2012
From Viral genome to specific peptide epitopes - Methods for identifying porcine T cell epitopes based on in silico predictions, in vitro identification and ex vivo verification

The affinity for and stability of peptides bound by major histocompatibility complex (MHC) class I molecules are instrumental factors in presentation of viral epitopes to cytotoxic T lymphocytes (CTLs). In swine, such peptide presentations by swine leukocyte antigens (SLA) are crucial for swine immunity during viral infections and disease. Here we combine the ability of complete nonamer peptide based binding matrices for three different SLA proteins to predict good candidates for peptide-SLA (pSLA) binding with that of an online available algorithm, NetMHCpan. Further we analyze the correlation between high affinity and high stability peptides bound by the highly expressed SLA molecules, SLA-1*0401, SLA-2*0401, and SLA-3*0401, using a luminescence oxygen channeling (LOCI) and a scintillation proximity assay, respectively. With this procedure, high affinity and highly stable SLA peptide epitopes can be identified within a given viral genome, along with the elimination of hundreds, or even thousands, of peptide sequences, which are not likely to be bound. Applying these methods can save enormous amounts of time and costs of epitope discovery studies and MHC binding analysis not only in swine but in almost any species of interest. Finally, peptide candidates of interest were verified as actual T cell epitopes using peptide-SLA complexes assembled into fluorescent tetramers to stain influenza-specific CTLs derived from vaccinated animals. From 20 such animals 16 had the correct SLA allele match and 7 of these qualified as potential candidates for tetramer staining. From the 7 animals 3 responded with a positive tetramer staining of 1% or higher.

General information
State: Published
Organisations: National Veterinary Institute, Division of Veterinary Diagnostics and Research, Adaptive Immunology & Parasitology, Department of Systems Biology, Center for Biological Sequence Analysis, University of Copenhagen
Authors: Pedersen, L. E. (Intern), Rasmussen, M. (Ekstern), Harndahl, M. (Ekstern), Nielsen, M. (Intern), Buus, S. (Forskerdatabase), Jungersen, G. (Intern)
Number of pages: 1
Publication date: 2012
Event: Abstract from 4th European Veterinary Immunology Workshop (EVIW), Edinburgh, United Kingdom.
Main Research Area: Technical/natural sciences
Electronic versions:
Abstract_Edinburg.pdf

Immune epitope database analysis resource
The immune epitope database analysis resource (IEDB-AR: http://tools.iedb.org) is a collection of tools for prediction and analysis of molecular targets of T- and B-cell immune responses (i.e. epitopes). Since its last publication in the NAR webserver issue in 2008, a new generation of peptide:MHC binding and T-cell epitope predictive tools have been added. As validated by different labs and in the first international competition for predicting peptide:MHC-I binding, their predictive performances have improved considerably. In addition, a new B-cell epitope prediction tool was added, and the homology mapping tool was updated to enable mapping of discontinuous epitopes onto 3D structures. Furthermore, to serve a wider range of users, the number of ways in which IEDB-AR can be accessed has been expanded. Specifically, the predictive tools can be programatically accessed using a web interface and can also be downloaded as software packages.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, La Jolla Institute for Allergy & Immunology, San Diego Supercomputer Center, Shanghai Advanced Research Institute
Authors: Kim, Y. (Ekstern), Ponomarenko, J. (Ekstern), Zhu, Z. (Ekstern), Tamang, D. (Ekstern), Wang, P. (Forskerdatabase), Greenbaum, J. (Ekstern), Lundegaard, C. (Intern), Sette, A. (Ekstern), Lund, O. (Intern), Bourne, P. E. (Ekstern), Nielsen, M. (Intern), Peters, B. (Ekstern)
Pages: W525-W530
Publication date: 2012
Main Research Area: Technical/natural sciences

Publication information
Journal: Nucleic Acids Research
Volume: 40
Issue number: W1
ISSN (Print): 0305-1048

Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 9.28 SJR 7.883 SNIP 2.744
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 7.358 SNIP 2.631 CiteScore 9.48
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 6.64 SNIP 2.552 CiteScore 8.74
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 6.801 SNIP 2.284 CiteScore 8.46
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 6.329 SNIP 2.407 CiteScore 8.62
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 5.976 SNIP 2.19 CiteScore 7.86
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 5.381 SNIP 2.034
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 5.669 SNIP 1.874
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 4.912 SNIP 1.578
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 5.1 SNIP 1.807
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 4.776 SNIP 2.051
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 5.092 SNIP 2.147
Web of Science (2005): Indexed yes
Scopus rating (2004): SJR 4.912 SNIP 1.971
Web of Science (2004): Indexed yes
Scopus rating (2003): SJR 0.111 SNIP 1.849
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 0.131 SNIP 1.529
Scopus rating (2001): SJR 0.161 SNIP 1.393
Scopus rating (2000): SJR 0.136 SNIP 1.661
NetMHCcons: a consensus method for the major histocompatibility complex class I predictions

A key role in cell-mediated immunity is dedicated to the major histocompatibility complex (MHC) molecules that bind peptides for presentation on the cell surface. Several in silico methods capable of predicting peptide binding to MHC class I have been developed. The accuracy of these methods depends on the data available characterizing the binding specificity of the MHC molecules. It has, moreover, been demonstrated that consensus methods defined as combinations of two or more different methods led to improved prediction accuracy. This plethora of methods makes it very difficult for the non-expert user to choose the most suitable method for predicting binding to a given MHC molecule. In this study, we have therefore made an in-depth analysis of combinations of three state-of-the-art MHC–peptide binding prediction methods (NetMHC, NetMHCpan and PickPocket). We demonstrate that a simple combination of NetMHC and NetMHCpan gives the highest performance when the allele in question is included in the training and is characterized by at least 50 data points with at least ten binders. Otherwise, NetMHCpan is the best predictor. When an allele has not been characterized, the performance depends on the distance to the training data. NetMHCpan has the highest performance when close neighbours are present in the training set, while the combination of NetMHCpan and PickPocket outperforms either of the two methods for alleles with more remote neighbours. The final method, NetMHCcons, is publicly available at www.cbs.dtu.dk/services/NetMHCcons, and allows the user in an automatic manner to obtain the most accurate predictions for any given MHC molecule.
Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity

Efficient presentation of peptide-MHC class I (pMHC-I) complexes to immune T cells should benefit from a stable peptide-MHC-I interaction. However, it has been difficult to distinguish stability from other requirements for MHC-I binding, for example, affinity. We have recently established a high-throughput assay for pMHC-I stability. Here, we have generated a large database containing stability measurements of pMHC-I complexes, and re-examined a previously reported unbiased analysis of the relative contributions of antigen processing and presentation in defining cytotoxic T lymphocyte (CTL) immunogenicity [Assarsson et al., J. Immunol. 2007. 178: 7890–7901]. Using an affinity-balanced approach, we demonstrated that immunogenic peptides tend to be more stably bound to MHC-I molecules compared with nonimmunogenic peptides. We also developed a bioinformatics method to predict pMHC-I stability, which suggested that 30% of the nonimmunogenic binders hitherto classified as "holes in the T-cell repertoire" can be explained as being unstably bound to MHC-I. Finally, we suggest that nonoptimal anchor residues in position 2 of the peptide are particularly prone to cause unstable interactions with MHC-I. We conclude that the availability of accurate predictors of pMHC-I
stability might be helpful in the elucidation of MHC-I restricted antigen presentation, and might be instrumental in future search strategies for MHC-I epitopes.
Peptide-MHC class I stability is a stronger predictor of CTL immunogenicity than peptide affinity

Peptide-MHC class I stability is a stronger predictor of CTL immunogenicity than peptide affinity

Mikkel Harndahl, Michael Rasmusse, Morten Nielsen, Soren Buus,

a Laboratory of Experimental Immunology, Faculty of Health Sciences, University of Copenhagen, Denmark
b Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Denmark

Efficient presentation of peptide-MHC class I (pMHC-I) complexes to immune T cells should benefit from a stable peptide-MHC-I interaction. However, it has been difficult to distinguish stability from other requirements for MHC-I binding e.g. affinity.

We have recently established a high-throughput assay for pMHCI stability. Here, we have generated a large database containing stability measurements of pMHCI complexes, and re-examined a previously reported unbiased analysis of the relative contributions of antigen processing and presentation in defining cytotoxic T lymphocyte (CTL) immunogenicity Assarsson et al., 2007. Using an affinity-balanced approach, we demonstrated that immunogenic peptides tend to be more stably bound to MHC-I molecules compared with non-immunogenic peptides. We also developed a bioinformatics method to predict pMHC-I stability, which suggested that 30% of the non-immunogenic binders hitherto classified as “holes in the T cell repertoire” can be explained as being unstably bound to MHC-I. Finally, we suggest that non-optimal anchor residues in position 2 of the peptide are particularly prone to cause unstable interactions with MHC-I. We conclude that the availability of accurate predictors of pMHC-I stability might be helpful in the elucidation of MHC-I restricted antigen presentation, and might be instrumental in future search strategies for MHC-I epitopes.

Reference

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, University of Copenhagen
Authors: Harndahl, M. N. (Forskerdatabase), Rasmusse, M. (Ekstern), Nielsen, M. (Intern), Buus, S. (Ekstern)
Pages: 11
Publication date: 2012

Host publication information
Title of host publication: Molecular Immunology
Volume: 51
Publisher: Elsevier
Main Research Area: Technical/natural sciences
Conference: 7th International EMBO Workshop on Antigen Presentation and Processing, Amsterdam, Netherlands, 24/04/2012 - 24/04/2012
Predictions versus high-throughput experiments in T-cell epitope discovery: competition or synergy?

Prediction methods as well as experimental methods for T-cell epitope discovery have developed significantly in recent years. High-throughput experimental methods have made it possible to perform full-length protein scans for epitopes restricted to a limited number of MHC alleles. The high costs and limitations regarding the number of proteins and MHC alleles that are feasibly handled by such experimental methods have made in silico prediction models of high interest. MHC binding prediction methods are today of a very high quality and can predict MHC binding peptides with high accuracy. This is possible for a large range of MHC alleles and relevant length of binding peptides. The predictions can easily be performed for complete proteomes of any size. Prediction methods are still, however, dependent on good experimental methods for validation, and should merely be used as a guide for rational epitope discovery. We expect prediction methods as well as experimental validation methods to continue to develop and that we will soon see clinical trials of products whose development has been guided by prediction methods.

General information

State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Lundegaard, C. (Intern), Lund, O. (Intern), Nielsen, M. (Intern)
Pages: 43
Publication date: 2012
Main Research Area: Technical/natural sciences

Publication information

Journal: Expert Review of Vaccines
Volume: 11
Issue number: 1
ISSN (Print): 1476-0584
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.946 SJR 1.551
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): SJR 1.482 SNIP 0.965 CiteScore 3.08
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.772 SNIP 0.947 CiteScore 3.21
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.533 SNIP 1.133 CiteScore 3.28
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.381 SNIP 1.085 CiteScore 3.27
ISI indexed (2013): ISI indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.331 SNIP 1.111 CiteScore 3.3
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.405 SNIP 1.275 CiteScore 3.55
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.482 SNIP 1.05
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 1.278 SNIP 1.025
Reliable B cell epitope predictions: impacts of method development and improved benchmarking.

The interaction between antibodies and antigens is one of the most important immune system mechanisms for clearing infectious organisms from the host. Antibodies bind to antigens at sites referred to as B-cell epitopes. Identification of the exact location of B-cell epitopes is essential in several biomedical applications such as; rational vaccine design, development of disease diagnostics and immunotherapeutics. However, experimental mapping of epitopes is resource intensive making in silico methods an appealing complementary approach. To date, the reported performance of methods for in silico mapping of B-cell epitopes has been moderate. Several issues regarding the evaluation data sets may however have led to the performance values being underestimated: Rarely, all potential epitopes have been mapped on an antigen, and antibodies are generally raised against the antigen in a given biological context not against the antigen monomer. Improper dealing with these aspects leads to many artificial false positive predictions and hence to incorrect low performance values. To demonstrate the impact of proper benchmark definitions, we here present an updated version of the DiscoTope method incorporating a novel spatial neighborhood definition and half-sphere exposure as surface measure. Compared to other state-of-the-art prediction methods, Discotope-2.0 displayed improved performance both in cross-validation and in independent evaluations. Using DiscoTope-2.0, we assessed the impact on performance when using proper benchmark definitions. For 13 proteins in the training data set where sufficient biological information was available to make a proper benchmark redefinition, the average AUC performance was improved from 0.791 to 0.824. Similarly, the average AUC performance on an independent evaluation data set improved from 0.712 to 0.727. Our results thus demonstrate that given proper benchmark definitions, B-cell epitope prediction methods achieve highly significant predictive performances suggesting these tools to be a powerful asset in rational epitope discovery. The updated version of DiscoTope is available at www.cbs.dtu.dk/services/DiscoTope-2.0.
Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion

Seq2Logo is a web-based sequence logo generator. Sequence logos are a graphical representation of the information content stored in a multiple sequence alignment (MSA) and provide a compact and highly intuitive representation of the position-specific amino acid composition of binding motifs, active sites, etc. in biological sequences. Accurate generation of sequence logos is often compromised by sequence redundancy and low number of observations. Moreover, most methods available for sequence logo generation focus on displaying the position-specific enrichment of amino acids, discarding the equally valuable information related to amino acid depletion. Seq2Logo aims at resolving these issues allowing the user to include sequence weighting to correct for data redundancy, pseudo counts to correct for low number of observations and different logotype representations each capturing different aspects related to amino acid enrichment and depletion. Besides allowing input in the format of peptides and MSA, Seq2Logo accepts input as Blast sequence profiles, providing easy access for non-expert end-users to characterize and identify functionally conserved/variable amino acids in any given protein of interest. The output from the server is a sequence logo and a PSSM. Seq2Logo is available at http://www.cbs.dtu.dk/biotools/Seq2Logo (14 May 2012, date last accessed).

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis
Authors: Thomsen, M. C. F. (Intern), Nielsen, M. (Intern)
Pages: W281-W287
Publication date: 2012
Main Research Area: Technical/natural sciences

Publication information
Journal: Nucleic Acids Research
Volume: 40
Issue number: W1
ISSN (Print): 0305-1048
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 9.28 SJR 7.883 SNIP 2.744
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 7.358 SNIP 2.631 CiteScore 9.48
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 6.64 SNIP 2.552 CiteScore 8.74
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 6.801 SNIP 2.284 CiteScore 8.46
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 6.329 SNIP 2.407 CiteScore 8.62
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 5.976 SNIP 2.19 CiteScore 7.86
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 5.381 SNIP 2.034
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 5.669 SNIP 1.874
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 4.912 SNIP 1.578
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 5.1 SNIP 1.807
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 4.776 SNIP 2.051
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 5.092 SNIP 2.147
Web of Science (2005): Indexed yes
Scopus rating (2004): SJR 4.912 SNIP 1.971
Web of Science (2004): Indexed yes
Scopus rating (2003): SJR 0.111 SNIP 1.849
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 0.131 SNIP 1.529
Scopus rating (2001): SJR 0.161 SNIP 1.393
Scopus rating (2000): SJR 0.136 SNIP 1.661
Web of Science (2000): Indexed yes
Scopus rating (1999): SJR 0.241 SNIP 1.548
Original language: English
Electronic versions:
PDF7E.pdf
The Cancer Exome Generated by Alternative mRNA Splicing Dilutes Predicted HLA Class I Epitope Density

Several studies have shown that cancers actively regulate alternative splicing. Altered splicing mechanisms in cancer lead to cancer-specific transcripts different from the pool of transcripts occurring only in healthy tissue. At the same time, altered presentation of HLA class I epitopes is frequently observed in various types of cancer. Down-regulation of genes related to HLA class I antigen processing has been observed in several cancer types, leading to fewer HLA class I antigens on the cell surface. Here, we use a peptidome wide analysis of predicted alternative splice forms, based on a publicly available database, to show that peptides over-represented in cancer splice variants comprise significantly fewer predicted HLA class I epitopes compared to peptides from normal transcripts. Peptides over-represented in cancer transcripts are in the case of the three most common HLA class I supertype representatives consistently found to contain fewer predicted epitopes compared to normal tissue. We observed a significant difference in amino acid composition between protein sequences associated with normal versus cancer tissue, as transcripts found in cancer are enriched with hydrophilic amino acids. This variation contributes to the observed significant lower likelihood of cancer-specific peptides to be predicted epitopes compared to peptides found in normal tissue.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis
Authors: Stranzl, T. (Intern), Larsen, M. V. (Intern), Lund, O. (Intern), Nielsen, M. (Intern), Brunak, S. (Intern)
Number of pages: 8
Pages: e38670
Publication date: 2012
Main Research Area: Technical/natural sciences

Publication information
Journal: P L o S One
Volume: 7
Issue number: 9
ISSN (Print): 1932-6203
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.111 SJR 1.164
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.11 SJR 1.236 SNIP 1.101
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.427 SNIP 1.136 CiteScore 3.32
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.559 SNIP 1.148 CiteScore 3.54
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.772 SNIP 1.153 CiteScore 3.94
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.982 SNIP 1.156 CiteScore 4.15
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 2.425 SNIP 1.233 CiteScore 4.58
Traditionally, T cell epitope discovery requires considerable amounts of tedious, slow, and costly experimental work. During the last decade, prediction tools have emerged as essential tools allowing researchers to select a manageable list of epitope candidates to test from a larger peptide, protein, or even proteome. However, no current tools address the complexity caused by the highly polymorphic nature of the restricting HLA molecules, which effectively individualizes T cell responses. To fill this gap, we here present an easy-to-use prediction tool named HLArestrictor (http://www.cbs.dtu.dk/services/HLArestrictor), which is based on the highly versatile and accurate NetMHCpan predictor, which here has been optimized for the identification of both the MHC restriction element and the corresponding minimal epitope of a T cell response in a given individual. As input, it requires high-resolution (i.e., 4-digit) HLA typing of the individual. HLArestrictor then predicts all 8-11mer peptide binders within one or more larger peptides and provides an overview of the predicted HLA restrictions and minimal epitopes. The method was tested on a large dataset of HIV IFNγ ELIspot peptide responses and was shown to identify HLA restrictions and minimal epitopes for about 90% of the positive peptide/patient pairs while rejecting more than 95% of the negative peptide-HLA pairs. Furthermore, for 18 peptide/HLA tetramer validated responses, HLArestrictor in all cases predicted both the HLA restriction element and minimal epitope. Thus, HLArestrictor should be a valuable tool in any T cell epitope discovery process aimed at identifying new epitopes from infectious diseases and other disease models.
Human Leukocyte Antigen (HLA) Class I Restricted Epitope Discovery in Yellow Fever and Dengue Viruses: Importance of HLA Binding Strength.

Epitopes from all available full-length sequences of yellow fever virus (YFV) and dengue fever virus (DENV) restricted by Human Leukocyte Antigen class I (HLA-I) alleles covering 12 HLA-I supertypes were predicted using the NetCTL algorithm. A subset of 179 predicted YFV and 158 predicted DENV epitopes were selected using the EpiSelect algorithm to allow for optimal coverage of viral strains. The selected predicted epitopes were synthesized and approximately 75% were found to bind the predicted restricting HLA molecule with an affinity, K(D), stronger than 500 nM. The immunogenicity of 25 HLA-A*02:01, 28 HLA-A*2402 and 28 HLA-B*07:02 binding peptides was tested in three HLA-transgenic mice models and led to the identification of 17 HLA-A*02:01, 4 HLA-A*2402 and 4 HLA-B*07:02 immunogenic peptides. The immunogenic peptides bound HLA significantly stronger than the non-immunogenic peptides. All except one of the immunogenic peptides had K(D) below 100 nM and the peptides with K(D) below 5 nM were more likely to be immunogenic. In addition, all the immunogenic peptides that were identified as having a high functional avidity had K(D) below 20 nM. A*02:01 transgenic mice were also inoculated twice with the 17DD YFV vaccine strain. Three of the YFV A*02:01 restricted peptides activated T-cells from the infected mice in vitro. All three peptides that elicited responses had an HLA binding affinity of 2 nM or less. The results indicate the importance of the strength of HLA binding in shaping the immune response.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, Johns Hopkins University, University of Copenhagen, University of Pittsburgh
Authors: Lund, O. (Intern), Nascimento, E. J. M. (Ekstern), Maciel, M. J. (Ekstern), Nielsen, M. (Intern), Larsen, M. V. (Intern), Lundegaard, C. (Intern), Harndahl, M. (Ekstern), Lamberth, K. (Ekstern), Buus, S. (Ekstern), Salmon, J. (Ekstern), August, T. J. (Ekstern), Marques, E. T. A. J. (Ekstern)
Pages: e26494
Publication date: 2011
Main Research Area: Technical/natural sciences

Publication information
Journal: P L o S One
Volume: 6
Issue number: 10
ISSN (Print): 1932-6203
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.111 SJR 1.164
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.11 SJR 1.236 SNIP 1.101
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.427 SNIP 1.136 CiteScore 3.32
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.559 SNIP 1.148 CiteScore 3.54
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.772 SNIP 1.153 CiteScore 3.94
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.982 SNIP 1.156 CiteScore 4.15
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
**Bibliographical note**
This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Source:** orbit
**Source-ID:** 286756
**Publication:** Research - peer-review › Journal article – Annual report year: 2011

---

**Immunogenic CTL Epitopes Tend to be Stably Bound to MHC Class I Molecules: Implications for 'Holes in the Stably Bound MHC-I Repertoire'**

---

**General information**

**State:** Published
**Organisations:** Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark
**Authors:** Harndahl, M. (Ekstern), Rasmussen, M. (Ekstern), Roder, G. (Ekstern), Pedersen, I. D. (Ekstern), Sørensen, M. (Ekstern), Nielsen, M. (Intern), Buus, S. (Ekstern)
**Pages:** 353-353
**Publication date:** 2011
**Main Research Area:** Technical/natural sciences

**Publication information**

**Journal:** Scandinavian Journal of Immunology
**Volume:** 73
**Issue number:** 4
**ISSN (Print):** 0300-9475
**Ratings:**
**BFI (2018):** BFI-level 1
**Web of Science (2018):** Indexed yes
**BFI (2017):** BFI-level 1
**Scopus rating (2017):** SNIP 0.621 SJR 0.891
**Web of Science (2017):** Indexed yes
**BFI (2016):** BFI-level 1
**Scopus rating (2016):** CiteScore 2.03 SJR 0.979 SNIP 0.644
**BFI (2015):** BFI-level 1
**Scopus rating (2015):** SJR 0.933 SNIP 0.679 CiteScore 1.97
**Web of Science (2015):** Indexed yes
**BFI (2014):** BFI-level 1
**Scopus rating (2014):** SJR 0.901 SNIP 0.665 CiteScore 1.91
Induction of Foot-and-Mouth Disease Virus-Specific Cytotoxic T Cell Killing by Vaccination

Foot-and-mouth disease (FMD) continues to be a significant threat to the health and economic value of livestock species. This acute infection is caused by the highly contagious FMD virus (FMDV), which infects cloven-hoofed animals including large and small ruminants and swine. Current vaccine strategies are all directed toward the induction of neutralizing antibody responses. However, the role of cytotoxic T lymphocytes (CTLs) has not received a great deal of attention, in part because of the technical difficulties associated with establishing a reliable assay of cell killing for this highly cytopathic virus. Here, we have used recombinant human adenovirus vectors as a means of delivering FMDV antigens in a T cell-directed vaccine in pigs. We tested the hypothesis that impaired processing of the FMDV capsid would enhance cytolytic activity, presumably by targeting all proteins for degradation and effectively increasing the class I MHC/FMDV peptide concentration for stimulation of a CTL response. We compared such a T cell targeting vaccine with the parental vaccine, previously shown to effectively induce a neutralizing antibody response. Our results show induction of FMDV-specific CD8+(+) CTL killing of MHC matched target cells in an antigen specific manner. Further, we confirm these results by MHC tetramer staining. This work presents the first demonstration of FMDV specific, CTL killing and confirmation by MHC tetramer staining in response to vaccination against FMDV.
Interdisciplinary Evaluation of Broadly-Reactive HLA Class II Restricted Epitopes Eliciting HIV-Specific CD4+T Cell Responses: Abstract of poster presentation

Background: CD4+ T cells orchestrate immune protection by “helping” other cells of our immune system to clear viral infections. It is well known that the preferential infection and depletion of CD4+ T cells contributes to hampered systemic T cell help following HIV infection. However, the functional and immunodominant discrepancies of CD4+ T cell responses targeting promiscuous MHC II restricted HIV epitopes remains poorly defined. Thus, utilization of interdisciplinary approaches might aid revealing broadly-reactive peptides eliciting CD4+ T cell responses. Methods: We utilized the novel bioinformatic prediction program NetMHCIIpan to select 64 optimized MHC II restricted HIV epitopes located in the HIV Gag, Pol, Env, Nef and Tat regions. The epitopes were selected to cover the global diversity of the virus (multiple subtypes) and the human immune system (diverse MHC II types). Optimized polychromatic flow cytometry analysis, including the functional markers IFNc, IL-2, IL-21, MIP-Ib and TNFa, revealed immunogenicity of the individual epitopes. The study subjects (n = 38) were of diverse ethnic background infected by different HIV subtypes. High resolution HLA typing and sequences of the HIV-Gag and Nef regions were obtained. Results: The FACS analysis revealed immunogenicity against 73% of the epitopes. All subjects, except one, recognized at least one epitope. Interestingly, almost all epitopes located in Gag (15/15) and Nef (14/15) elicited responses, while epitopes in Pol (10/15) and Env (5/15) revealed restricted CD4+ T cell immunogenicity. This difference in immunogenicity between the regions was significant (One-way ANOVA: p <0.001). Additionally, Gag and Nef epitopes generated greater polyfunctionality than Poland Env-specific CD4+ T cells. Importantly, we found that the use of optimized epitopes improved the polyfunctionality compared with overlapping HIV Gag (p55) peptides. Conclusion: Using an unbiased approach where we have predicted peptides with same prerequisites, we demonstrate that HIV-specific CD4+ T cell immunodominance is heavily skewed, targeting particularly Gag and Nef.
Main Research Area: Technical/natural sciences

Publication information
Journal: AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume: 27
Issue number: 10
ISSN (Print): 0889-2229
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.655 SJR 1.066
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): SJR 1.07 SNIP 0.588 CiteScore 1.66
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.091 SNIP 0.603 CiteScore 1.67
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.253 SNIP 0.673 CiteScore 2.01
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.194 SNIP 0.738 CiteScore 2.15
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.17 SNIP 0.783 CiteScore 2.34
ISI indexed (2012): ISI indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.18 SNIP 0.666 CiteScore 2.22
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.207 SNIP 0.679
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 1.051 SNIP 0.655
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 1.142 SNIP 0.714
Scopus rating (2007): SJR 1.146 SNIP 0.707
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 1.215 SNIP 0.649
Scopus rating (2005): SJR 1.307 SNIP 0.77
Scopus rating (2004): SJR 1.127 SNIP 0.742
Scopus rating (2003): SJR 1.066 SNIP 0.619
Scopus rating (2002): SJR 1.053 SNIP 0.622
Web of Science (2002): Indexed yes
Scopus rating (2001): SJR 1.363 SNIP 0.706
Scopus rating (2000): SJR 1.329 SNIP 0.665
Scopus rating (1999): SJR 1.252 SNIP 0.749
Original language: English
DOIs:
10.1089/aid.2011.1502
Links:
http://www.hivvaccineenterprise.org/conference/2011/
Source: orbit
Source-ID: 286565
Publication: Research - peer-review › Conference article – Annual report year: 2011
Machine learning competition in immunology – Prediction of HLA class I binding peptides

Experimental studies of immune system and related applications such as characterization of immune responses against pathogens, vaccine design, or optimization of therapies are combinatorially complex, time-consuming and expensive. The main methods for large-scale identification of T-cell epitopes from pathogens or cancer proteomes involve either reverse immunology or high-throughput mass spectrometry (HTMS). Reverse immunology approaches involve pre-screening of proteomes by computational algorithms, followed by experimental validation of selected targets ([Mora et al., 2006], [De Groot et al., 2008] and [Larsen et al., 2010]). HTMS involves HLA typing, immunoaffinity chromatography of HLA molecules, HLA extraction, and chromatography combined with tandem mass spectrometry, followed by the application of computational algorithms for peptide characterization (Bassani-Sternberg et al., 2010). Hundreds of naturally processed HLA class I associated peptides have been identified in individual studies using HTMS in normal (Escobar et al., 2008), cancer ([Antwi et al., 2009] and [Bassani-Sternberg et al., 2010]), autoimmunity-related (Ben Dror et al., 2010), and infected samples (Wahl et al, 2010). Computational algorithms are essential steps in high-throughput identification of T-cell epitope candidates using both reverse immunology and HTMS approaches. Peptide binding to MHC molecules is the single most selective step in defining T cell epitope and the accuracy of computational algorithms for prediction of peptide binding, therefore, determines the accuracy of the overall method. Computational predictions of peptide binding to HLA, both class I and class II, use a variety of algorithms ranging from binding motifs to advanced machine learning techniques ([Brusic et al., 2004] and [Lafuente and Reche, 2009]) and standards for their assessments have been developed. The assessments of computational servers that predict peptide binding to several common HLA class I alleles have been performed by different groups (see [Peters et al., 2006], [Lin et al., 2008] and [Gowthaman et al., 2010]). Some of these models were reported to be highly accurate while others need improvement.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, Dana-Farber Cancer Institute, Institute of Microbial Technology, Fred Hutchinson Cancer Research Center, University of East Anglia, Fudan University, Microsoft Research Redmond, La Jolla Institute for Allergy & Immunology, University of Tubingen, Frederik University Cyprus, Bar-Ilan University, Iowa State University, Vanderbilt University, Nicolaus Copernicus University in Torun, Nanyang Technological University, University of Cyprus
Pages: 1-4
Publication date: 2011
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Immunological Methods
Volume: 374
Issue number: 1-2
ISSN (Print): 0022-1759
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.715 SJR 1.289
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 1.92 SJR 1.089 SNIP 0.65
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.064 SNIP 0.739 CiteScore 2.07
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.018 SNIP 0.824 CiteScore 1.99
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.087 SNIP 0.834 CiteScore 2.31
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Computational models, Peptide binding, Human leukocyte antigen, HLA, Predictions, Machine learning

NNAlign: A Web-Based Prediction Method Allowing Non-Expert End-User Discovery of Sequence Motifs in Quantitative Peptide Data

Recent advances in high-throughput technologies have made it possible to generate both gene and protein sequence data at an unprecedented rate and scale thereby enabling entirely new "omics"-based approaches towards the analysis of complex biological processes. However, the amount and complexity of data that even a single experiment can produce seriously challenges researchers with limited bioinformatics expertise, who need to handle, analyze and interpret the data before it can be understood in a biological context. Thus, there is an unmet need for tools allowing non-bioinformatics users to interpret large data sets. We have recently developed a method, NNAlign, which is generally applicable to any biological problem where quantitative peptide data is available. This method efficiently identifies underlying sequence patterns by simultaneously aligning peptide sequences and identifying motifs associated with quantitative readouts. Here, we provide a web-based implementation of NNAlign allowing non-expert end-users to submit their data (optionally adjusting method parameters), and in return receive a trained method (including a visual representation of the identified motif) that subsequently can be used as prediction method and applied to unknown proteins/peptides. We have successfully applied this method to several different data sets including peptide microarray-derived sets containing more than 100,000 data points. NNAlign is available online at http://www.cbs.dtu.dk/services/NNAlign.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, Schafer-N, University of Copenhagen
Authors: Andreatta, M. (Intern), Schafer-Nielsen, C. (Ekstern), Lund, O. (Intern), Buus, S. (Ekstern), Nielsen, M. (Intern)
Pages: e26781
Publication date: 2011
Main Research Area: Technical/natural sciences
Porcine major histocompatibility complex (MHC) class I molecules and analysis of their peptide-binding specificities

In all vertebrate animals, CD8+ cytotoxic T lymphocytes (CTLs) are controlled by major histocompatibility complex class I (MHC-I) molecules. These are highly polymorphic peptide receptors selecting and presenting endogenously derived epitopes to circulating CTLs. The polymorphism of the MHC effectively individualizes the immune response of each member of the species. We have recently developed efficient methods to generate recombinant human MHC-I (also known as human leukocyte antigen class I, HLA-I) molecules, accompanying peptide-binding assays and predictors, and HLA tetramers for specific CTL staining and manipulation. This has enabled a complete mapping of all HLA-I specificities ("the Human MHC Project"). Here, we demonstrate that these approaches can be applied to other species. We systematically transferred domains of the frequently expressed swine MHC-I molecule, SLA-1*0401, onto a HLA-I molecule (HLA-A*11:01), thereby generating recombinant human/swine chimeric MHC-I molecules as well as the intact SLA-1*0401 molecule. Biochemical peptide-binding assays and positional scanning combinatorial peptide libraries were used to analyze the peptide-binding motifs of these molecules. A pan-specific predictor of peptide–MHC-I binding, NetMHCpan, which was originally developed to cover the binding specificities of all known HLA-I molecules, was successfully used to predict the specificities of the SLA-1*0401 molecule as well as the porcine/human chimeric MHC-I molecules. These data indicate that it is possible to extend the biochemical and bioinformatics tools of the Human MHC Project to other vertebrate species.
Prediction of epitopes using neural network based methods

In this paper, we describe the methodologies behind three different aspects of the NetMHC family for prediction of MHC class I binding, mainly to HLAs. We have updated the prediction servers, NetMHC-3.2, NetMHCpan-2.2, and a new consensus method, NetMHCcons, which, in their previous versions, have been evaluated to be among the very best performing MHC:peptide binding predictors available. Here we describe the background for these methods, and the rationale behind the different optimization steps implemented in the methods. We go through the practical use of the methods, which are publicly available in the form of relatively fast and simple web interfaces. Furthermore, we will review results obtained in actual epitope discovery projects where previous implementations of the described methods have been used in the initial selection of potential epitopes. Selected potential epitopes were all evaluated experimentally using ex vivo assays.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Lundegaard, C. (Intern), Lund, O. (Intern), Nielsen, M. (Intern)
Pages: 26-34
Publication date: 2011
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Immunological Methods
Volume: 374
Issue number: 1-2
ISSN (Print): 0022-1759
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles

CPHmodels-3.0 is a web server predicting protein 3D structure by use of single template homology modeling. The server employs a hybrid of the scoring functions of CPHmodels-2.0 and a novel remote homology-modeling algorithm. A query sequence is first attempted modeled using the fast CPHmodels-2.0 profile-profile scoring function suitable for close homology modeling. The new computational costly remote homology-modeling algorithm is only engaged provided that no suitable PDB template is identified in the initial search. CPHmodels-3.0 was benchmarked in the CASP8 competition and
produced models for 94% of the targets (117 out of 128), 74% were predicted as high reliability models (87 out of 117). These achieved an average RMSD of 4.6 Å when superimposed to the 3D structure. The remaining 26% low reliability models (30 out of 117) could superimpose to the true 3D structure with an average RMSD of 9.3 Å. These performance values place the CPHmodels-3.0 method in the group of high performing 3D prediction tools. Beside its accuracy, one of the important features of the method is its speed. For most queries, the response time of the server is

**General information**

State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Nielsen, M. (Intern), Lundegaard, C. (Intern), Lund, O. (Intern), Petersen, T. N. (Intern)
Pages: W576-W581
Publication date: 2010
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Nucleic Acids Research
Volume: 38
ISSN (Print): 0305-1048
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 9.28 SJR 7.883 SNIP 2.744
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 7.358 SNIP 2.631 CiteScore 9.48
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 6.64 SNIP 2.552 CiteScore 8.74
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 6.801 SNIP 2.284 CiteScore 8.46
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 6.329 SNIP 2.407 CiteScore 8.62
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 5.976 SNIP 2.19 CiteScore 7.86
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 5.381 SNIP 2.034
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 5.669 SNIP 1.874
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 4.912 SNIP 1.578
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 5.1 SNIP 1.807
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 4.776 SNIP 2.051
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 5.092 SNIP 2.147
Web of Science (2005): Indexed yes
Scopus rating (2004): SJR 4.912 SNIP 1.971
Web of Science (2004): Indexed yes
Scopus rating (2003): SJR 0.111 SNIP 1.849
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 0.131 SNIP 1.529
Scopus rating (2001): SJR 0.161 SNIP 1.393
Scopus rating (2000): SJR 0.136 SNIP 1.661
Web of Science (2000): Indexed yes
Scopus rating (1999): SJR 0.241 SNIP 1.548

Original language: English
Electronic versions:
E805O0d01.pdf
DOIs:
10.1093/nar/gkq535

Bibliographical note
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Source: orbit
Source-ID: 265767
Publication: Research - peer-review › Journal article – Annual report year: 2010

CTL epitopes of FMDV determined by NetMHCpan-driven predictions of SLA/peptide binding, confirmed by tetramer complex formation and staining

General information
State: Published
Organisations: Adaptive Immunology & Parasitology, Division of Veterinary Diagnostics and Research, National Veterinary Institute, Center for Biological Sequence Analysis, Department of Systems Biology, Department of Acoustic Technology, Agricultural Research Service
Authors: Pedersen, L. E. (Intern), Nielsen, M. (Intern), Patch, J. R. (Ekstern), Jungersen, G. (Intern), Buus, S. (Ekstern), Golde, W. T. (Ekstern)
Publication date: 2010
Event: Poster session presented at 9th International Veterinary Immunology Symposium, Tokyo, Japan.
Main Research Area: Technical/natural sciences
Electronic versions:
FINAL, Lasse Japan 9IVIS 2010 Poster 90x90 July 23 [Kompatibilitetstilstand][1].pdf
Source: orbit
Source-ID: 271673
Publication: Research › Poster – Annual report year: 2010

Degree of Predicted Minor Histocompatibility Antigen Mismatch Correlates with Poorer Clinical Outcomes of Nonmyeloablative Allogeneic Hematopoietic Cell Transplantation
In fully HLA-matched allogeneic hematopoietic cell transplantsations (HCT), the main mechanism of the beneficial graft-versus-tumor (GVT) effect and of the detrimental graft-versus-host disease (GVHD) is believed to be caused by donor cytotoxic T cells directed against disparate recipient minor histocompatibility antigens (miHAs). The most common origin of disparate miHAs is non-synonymous single nucleotide polymorphism (nsSNP) differences between donors and patients. At this time, only some 30 miHAs have been identified and registered, but considering the numerous different HLA-types in the human population as well as all the possible nsSNP differences between any two individuals, it is likely that many miHAs have yet to be discovered. The objective of the current study was to predict novel HLA-A and HLA-B restricted miHAs in a cohort of patients treated with non-myeloablative conditioning allogeneic HCT (matched related donor, n=70; matched unrelated donor, n=56) for hematologic malignancies. Initially, the cohort was genotyped for 53 nsSNPs in 11 known miHA source proteins. Twenty-three nsSNPs within six miHA source proteins showed variation in the graft-versus-host (GVH) direction. No correlation between the number of disparate nsSNPs and clinical outcome could be observed. Next, miHAs in the GVH direction were predicted for each patient-donor pair. Using the NetMHCpan predictor, we identified peptides encompassing a nsSNP variant uniquely expressed by the patient and with predicted binding to any of the HLA-A or -B molecules expressed by the patient and donor. Patients with more than the median of three predicted miHAs had a significantly lower five-year overall survival (42% vs 70%, P=0.0060, adjusted hazard ratio (HR) 2.6, P=0.0047) and significantly higher treatment related mortality (39% vs 10%, P=0.0094, adjusted HR 4.6, P=0.0038). No
association between number of predicted miHAs and any other clinical outcome parameters was observed. Collectively, our data suggest that the clinical outcome of HCT is not affected by disparate nsSNPs per se, but rather by the HLA-restricted presentation and recognition of peptides encompassing these. Our data also suggest that 6 of the 11 proteins included in the current study could contain more miHAs yet to be identified, and that the presence of multiple miHAs confers a higher risk of mortality after non-myeloablative conditioning HCT. Furthermore, our data suggest a possible role for in silico based miHA predictions, in donor selection as well as in selecting candidate miHAs for further evaluation in vitro and in vivo experiments. Copyright © 2010. Published by Elsevier Inc.

**General information**
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis
Authors: Larsen, M. E. (Intern), Kornblit, B. (Ekstern), Larsen, M. V. (Intern), Masmas, T. (Ekstern), Nielsen, M. (Intern), Thiim, M. H. (Intern), Garred, P. (Ekstern), Stryhn, A. (Ekstern), Lund, O. (Intern), Buus, S. (Ekstern), Vindelov, L. (Ekstern)
Publication date: 2010

**Publication information**
Journal: Blood & Marrow Transplant Newsletter
Volume: 16
Issue number: 10
ISSN (Print): 1525-6855
Ratings:
ISI indexed (2013): ISI indexed no
ISI indexed (2012): ISI indexed no
ISI indexed (2011): ISI indexed no
Original language: English
GVHD, mHags, Nonsynonymous SNPs, Survival, Allo-HCT
DOIs:
10.1016/j.bbmt.2010.03.022
Source: orbit
Source-ID: 265694
Publication: Research - peer-review › Journal article – Annual report year: 2010

**HLA Class I Binding 9mer Peptides from Influenza A Virus Induce CD4(+) T Cell Responses**
Background: Identification of human leukocyte antigen class I (HLA-I) restricted cytotoxic T cell (CTL) epitopes from influenza virus is of importance for the development of new effective peptide-based vaccines. Methodology/Principal Findings: In the present work, bioinformatics was used to predict 9mer peptides derived from available influenza A viral proteins with binding affinity for at least one of the 12 HLA-I supertypes. The predicted peptides were then selected in a way that ensured maximal coverage of the available influenza A strains. One hundred and thirty one peptides were synthesized and their binding affinities for the HLA-I supertypes were measured in a biochemical assay. Influenza-specific T cell responses towards the peptides were quantified using IFN gamma ELISPOT assays with peripheral blood mononuclear cells (PBMC) from adult healthy HLA-I typed donors as responder cells. Of the 131 peptides, 21 were found to induce T cell responses in 19 donors. In the ELISPOT assay, five peptides induced responses that could be blocked by the pan-specific anti-HLA-I antibody W6/32, whereas 15 peptides induced responses that could be completely blocked in the presence of the pan-specific anti-HLA class II (HLA-II) antibody IVA12. Blocking of HLA-II subtype reactivity revealed that 8 and 6 peptide responses were blocked by anti-HLA-DR and -DP antibodies, respectively. Peptide reactivity of PBMC depleted of CD4(+) or CD8(+) T cells prior to the ELISPOT culture revealed that effectors are either CD4(+) (the majority of reactivities) or CD8(+) T cells, never a mixture of these subsets. Three of the peptides, recognized by CD4(+) T cells showed binding to recombinant DRA1*0101/DRB1*0401 or DRA1*0101/DRB5*0101 molecules in a recently developed biochemical assay. Conclusions/Significance: HLA-I binding 9mer influenza virus-derived peptides induce in many cases CD4(+) T cell responses restricted by HLA-II molecules.

**General information**
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Pages: e10533
Publication date: 2010
Main Research Area: Technical/natural sciences

**Publication information**
Journal: PLoS One
Volume: 5
Identification of CD8(+) T Cell Epitopes in the West Nile Virus Polyprotein by Reverse-Immunology Using NetCTL

Background: West Nile virus (WNV) is a growing threat to public health and a greater understanding of the immune response raised against WNV is important for the development of prophylactic and therapeutic strategies.

Methodology/Principal Findings: In a reverse-immunology approach, we used bioinformatics methods to predict WNV-
specific CD8(+) T cell epitopes and selected a set of peptides that constitutes maximum coverage of 20 fully-sequenced WNV strains. We then tested these putative epitopes for cellular reactivity in a cohort of WNV-infected patients. We identified 26 new CD8(+) T cell epitopes, which we propose are restricted by 11 different HLA class I alleles. Aiming for optimal coverage of human populations, we suggest that 11 of these new WNV epitopes would be sufficient to cover from 48% to 93% of ethnic populations in various areas of the World. Conclusions/Significance: The 26 identified CD8(+) T cell epitopes contribute to our knowledge of the immune response against WNV infection and greatly extend the list of known WNV CD8(+) T cell epitopes. A polytope incorporating these and other epitopes could possibly serve as the basis for a WNV vaccine.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Larsen, M. V. (Intern), Lelic, A. (Ekstern), Parsons, R. (Ekstern), Nielsen, M. (Intern), Hoof, I. (Ekstern), Lamberth, K. (Ekstern), Loeb, M. (Ekstern), Buus, S. (Ekstern), Bramson, J. (Ekstern), Lund, O. (Intern)
Publication date: 2010
Main Research Area: Technical/natural sciences

Publication information
Journal: P L o S One
Volume: 5
Issue number: 9
ISSN (Print): 1932-6203
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.111 SJR 1.164
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.11 SJR 1.236 SNIP 1.101
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.427 SNIP 1.136 CiteScore 3.32
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.559 SNIP 1.148 CiteScore 3.54
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.772 SNIP 1.153 CiteScore 3.94
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.982 SNIP 1.156 CiteScore 4.15
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 2.425 SNIP 1.233 CiteScore 4.58
ISI indexed (2011): ISI indexed no
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 2.705 SNIP 1.178
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 2.614 SNIP 1.046
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 2.506 SNIP 1.006
Web of Science (2008): Indexed yes
In Silico Prediction of Human Pathogenicity in the gamma-Proteobacteria

In Background: Although the majority of bacteria are innocuous or even beneficial for their host, others are highly infectious pathogens that can cause widespread and deadly diseases. When investigating the relationships between bacteria and other living organisms, it is therefore essential to be able to separate pathogenic organisms from non-pathogenic ones. Using traditional experimental methods for this purpose can be very costly and time-consuming, and also uncertain since animal models are not always good predictors for pathogenicity in humans. Bioinformatics-based methods are therefore strongly needed to mine the fast growing number of genome sequences and assess in a rapid and reliable way the pathogenicity of novel bacteria. Methodology/Principal Findings: We describe a new in silico method for the prediction of bacterial pathogenicity, based on the identification in microbial genomes of features that appear to correlate with virulence. The method does not rely on identifying genes known to be involved in pathogenicity (for instance virulence factors), but rather it inherently builds families of proteins that, irrespective of their function, are consistently present in only one of the two kinds of organisms, pathogens or non-pathogens. Whether a new bacterium carries proteins contained in these families determines its prediction as pathogenic or non-pathogenic. The application of the method on a set of known genomes correctly classified the virulence potential of 86% of the organisms tested. An additional validation on an independent test-set assigned correctly 22 out of 24 bacteria. Conclusions: The proposed approach was demonstrated to go beyond the species bias imposed by evolutionary relatedness, and performs better than predictors based solely on taxonomy or sequence similarity. A set of protein families that differentiate pathogenic and non-pathogenic strains were identified, including families of yet uncharacterized proteins that are suggested to be involved in bacterial pathogenicity.
Interdisciplinary Analysis of HIV-Specific CD8(+) T Cell Responses against Variant Epitopes Reveals Restricted TCR Promiscuity

HIV-1 specific CTL responses play a key role in limiting viral replication. CTL responses are sensitive to viral escape mutations, which influence recognition of the virus. Although CTLs have been shown to recognize epitope variants, the extent of this cross-reactivity has not been quantitatively investigated in a genetically diverse cohort of HIV-1 infected patients. Using a novel bioinformatic binding prediction method, we aimed to explain the pattern of epitope-specific CTL responses based on the patients' HLA genotype and autologous virus sequence quantitatively. Sequences covering predicted and tested HLA class I-restricted epitopes (peptides) within the HIV-Gag, Pol, and Nef regions were obtained from 26 study subjects resulting in 1492 patient-specific peptide pairs. Epitopes that were recognized in ELISPOT assays were found to be significantly more similar to the autologous virus than those that did not elicit a response. A single substitution in the presented epitope decreased the chance of a CTL response by 40%. The impact of sequence similarity on cross-recognition was confirmed by testing immune responses against multiple variants of six selected epitopes. Substitutions at central positions in the epitope were particularly likely to result in abrogation of recognition. In summary, the presented data demonstrate a highly restricted promiscuity of HIV-1 specific CTL in the recognition of variant epitopes. In addition, our results illustrate that bioinformatic prediction methods are useful to study the complex pattern of CTL responses exhibited by an HIV-1 infected patient cohort and for identification of optimal targets for novel therapeutic or vaccine approaches. The Journal of Immunology, 2010, 184: 5383-5391.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Hoof, I. (Intern), Perez, C. (Ekstern), Buggert, M. (Ekstern), Gustafsson, R. (Ekstern), Nielsen, M. (Intern), Lund, O. (Intern), Karlsson, A. (Ekstern)
Limitations of Ab Initio Predictions of Peptide Binding to MHC Class II Molecules

Successful predictions of peptide MHC binding typically require a large set of binding data for the specific MHC molecule that is examined. Structure based prediction methods promise to circumvent this requirement by evaluating the physical contacts a peptide can make with an MHC molecule based on the highly conserved 3D structure of peptide:MHC complexes. While several such methods have been described before, most are not publicly available and have not been independently tested for their performance. We here implemented and evaluated three prediction methods for MHC class II molecules: statistical potentials derived from the analysis of known protein structures; energetic evaluation of different peptide snapshots in a molecular dynamics simulation; and direct analysis of contacts made in known 3D structures of peptide:MHC complexes. These methods are ab initio in that they require structural data of the MHC molecule examined, but no specific peptide:MHC binding data. Moreover, these methods retain the ability to make predictions in a sufficiently short time scale to be useful in a real world application, such as screening a whole proteome for candidate binding peptides. A rigorous evaluation of each methods prediction performance showed that these are significantly better than random, but still substantially lower than the best performing sequence based class II prediction methods available. While the approaches presented here were developed independently, we have chosen to present our results together in order to support the notion that generating structure based predictions of peptide:MHC binding without using binding data is unlikely to give satisfactory results.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis
Authors: Zhang, H. (Intern), Lund, O. (Intern), Nielsen, M. (Intern), Wang, P. (Ekstern), Papangelopoulos, N. (Ekstern), Xu, Y. (Ekstern), Sette, A. (Ekstern), Bourne, P. (Ekstern), Ponomarenko, J. (Ekstern), Peters, B. (Ekstern)
Pages: e9272
Publication date: 2010
Main Research Area: Technical/natural sciences

Publication information
Journal: P L o S One
Volume: 5
Issue number: 2
ISSN (Print): 1932-6203
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.111 SJR 1.164
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.11 SJR 1.236 SNIP 1.101
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.427 SNIP 1.136 CiteScore 3.32
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.559 SNIP 1.148 CiteScore 3.54
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.772 SNIP 1.153 CiteScore 3.94
ISI indexed (2013): ISI indexed yes
Major histocompatibility complex class I binding predictions as a tool in epitope discovery

Over the last decade, in silico models of the major histocompatibility complex (MHC) class I pathway have developed significantly. Before, peptide binding could only be reliably modelled for a few major human or mouse histocompatibility molecules; now, high-accuracy predictions are available for any human leucocyte antigen (HLA) -A or -B molecule with known protein sequence. Furthermore, peptide binding to MHC molecules from several non-human primates, mouse strains and other mammals can now be predicted. In this review, a number of different prediction methods are briefly explained, highlighting the most useful and historically important. Selected case stories, where these ‘reverse immunology’ systems have been used in actual epitope discovery, are briefly reviewed. We conclude that this new generation of epitope discovery systems has become a highly efficient tool for epitope discovery, and recommend that the less accurate prediction systems of the past be abandoned, as these are obsolete.
MHC Class II epitope predictive algorithms

Major histocompatibility complex class II (MHC-II) molecules sample peptides from the extracellular space, allowing the immune system to detect the presence of foreign microbes from this compartment. To be able to predict the immune response to given pathogens, a number of methods have been developed to predict peptide-MHC binding. However, few methods other than the pioneering TEPITOPE/ProPred method have been developed for MHC-II. Despite recent progress in method development, the predictive performance for MHC-II remains significantly lower than what can be obtained for MHC-I. One reason for this is that the MHC-II molecule is open at both ends allowing binding of peptides extending out of the groove. The binding core of MHC-II-bound peptides is therefore not known a priori and the binding motif is hence not readily discernible. Recent progress has been obtained by including the flanking residues in the predictions. All attempts to make ab initio predictions based on protein structure have failed to reach predictive performances similar to those that can be obtained by data-driven methods. Thousands of different MHC-II alleles exist in humans. Recently developed pan-specific methods have been able to make reasonably accurate predictions for alleles that were not included in the training data. These methods can be used to define supertypes (clusters) of MHC-II alleles where alleles within each supertype have similar binding specificities. Furthermore, the pan-specific methods have been used to make a graphical atlas such as the MHCMotifviewer, which allows for visual comparison of specificities of different alleles.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Nielsen, M. (Intern), Lund, O. (Intern), Buus, S. (Ekstern), Lundegaard, C. (Intern)
Pages: 319-328
Publication date: 2010
Main Research Area: Technical/natural sciences

Publication Information
Journal: Immunology
Volume: 130
Issue number: 3
ISSN (Print): 0019-2805
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.938 SJR 1.69
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.74 SJR 1.964 SNIP 0.965
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 2.075 SNIP 0.965 CiteScore 3.83
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 2.048 SNIP 1.043 CiteScore 3.61
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 2.086 SNIP 1.084 CiteScore 3.97
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.941 SNIP 1.04 CiteScore 3.94
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.884 SNIP 0.992 CiteScore 3.75
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 0.121 SNIP 0.912
Web of Science (2010): Indexed yes
NetCTLpan: pan-specific MHC class I pathway epitope predictions
Reliable predictions of immunogenic peptides are essential in rational vaccine design and can minimize the experimental effort needed to identify epitopes. In this work, we describe a pan-specific major histocompatibility complex (MHC) class I epitope predictor, NetCTLpan. The method integrates predictions of proteasomal cleavage, transporter associated with antigen processing (TAP) transport efficiency, and MHC class I binding affinity into a MHC class I pathway likelihood score and is an improved and extended version of NetCTL. The NetCTLpan method performs predictions for all MHC class I molecules with known protein sequence and allows predictions for 8-, 9-, 10-, and 11-mer peptides. In order to meet the need for a low false positive rate, the method is optimized to achieve high specificity. The method was trained and validated on large datasets of experimentally identified MHC class I ligands and cytotoxic T lymphocyte (CTL) epitopes. It has been reported that MHC molecules are differentially dependent on TAP transport and proteasomal cleavage. Here, we did not find any consistent signs of such MHC dependencies, and the NetCTLpan method is implemented with fixed weights for proteasomal cleavage and TAP transport for all MHC molecules. The predictive performance of the NetCTLpan method was shown to outperform other state-of-the-art CTL epitope prediction methods. Our results further confirm the importance of using full-type human leukocyte antigen restriction information when identifying MHC class I epitopes. Using the NetCTLpan method, the experimental effort to identify 90% of new epitopes can be reduced by 15% and 40%, respectively, when compared to the NetMHCpan and NetCTL methods. The method and benchmark datasets are available at http://www.cbs.dtu.dk/services/NetCTLpan/.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Stranzl, T. (Intern), Larsen, M. V. (Intern), Lundegaard, C. (Intern), Nielsen, M. (Intern)
Pages: 357-368
Publication date: 2010
Main Research Area: Technical/natural sciences

Publication information
Journal: Immunogenetics
Volume: 62
Issue number: 6
ISSN (Print): 0093-7711
Ratings:

BFI (2018): BFI-level 1
NetMHCIpan-2.0 - Improved pan-specific HLA-DR predictions using a novel concurrent alignment and weight optimization training procedure

BACKGROUND: Binding of peptides to Major Histocompatibility class II (MHC-II) molecules play a central role in governing responses of the adaptive immune system. MHC-II molecules sample peptides from the extracellular space allowing the immune system to detect the presence of foreign microbes from this compartment. Predicting which peptides bind to an MHC-II molecule is therefore of pivotal importance for understanding the immune response and its effect on host-pathogen interactions. The experimental cost associated with characterizing the binding motif of an MHC-II molecule is significant and large efforts have therefore been placed in developing accurate computer methods capable of predicting this binding event. Prediction of peptide binding to MHC-II is complicated by the open binding cleft of the MHC-II molecule, allowing binding of peptides extending out of the binding groove. Moreover, the genes encoding the MHC molecules are immensely diverse leading to a large set of different MHC molecules each potentially binding a unique set of peptides. Characterizing each MHC-II molecule using peptide-screening binding assays is hence not a viable option. RESULTS: Here, we present an MHC-II binding prediction algorithm aiming at dealing with these challenges. The method is a pan-specific version of the earlier published allele-specific NN-align algorithm and does not require any pre-alignment of the input data. This allows the method to benefit also from information from alleles covered by limited binding data. The method is evaluated on a large and diverse set of benchmark data, and is shown to significantly out-perform state-of-the-art MHC-II prediction methods. In particular, the method is found to boost the performance for alleles characterized by limited binding data where conventional allele-specific methods tend to achieve poor prediction accuracy.

CONCLUSIONS: The method thus shows great potential for efficient boosting the accuracy of MHC-II binding prediction, as accurate predictions can be obtained for novel alleles at highly reduced experimental costs. Pan-specific binding predictions can be obtained for all alleles with known protein sequence and the method can benefit by including data in the training from alleles even where only few binders are known. The method and benchmark data are available at http://www.cbs.dtu.dk/services/NetMHCIpan-2.0.
Identification of catalytic residues (CR) is essential for the characterization of enzyme function. CR are, in general, conserved and located in the functional site of a protein in order to attain their function. However, many non-catalytic residues are highly conserved and not all CR are conserved throughout a given protein family making identification of CR a challenging task. Here, we put forward the hypothesis that CR carry a particular signature defined by networks of close proximity residues with high mutual information (MI), and that this signature can be applied to distinguish functional from other non-functional conserved residues. Using a data set of 434 Pfam families included in the catalytic site atlas (CSA) database, we tested this hypothesis and demonstrated that MI can complement amino acid conservation scores to detect CR. The Kullback-Leibler (KL) conservation measurement was shown to significantly outperform both the Shannon entropy and maximal frequency measurements. Residues in the proximity of catalytic sites were shown to be rich in shared MI. A structural proximity MI average score (termed pMI) was demonstrated to be a strong predictor for CR, thus confirming the proposed hypothesis. A structural proximity conservation average score (termed pC) was also calculated and demonstrated to carry distinct information from pMI. A catalytic likeliness score (Cls), combining the KL, pC and pMI measures, was shown to lead to significantly improved prediction accuracy. At a specificity of 0.90, theCls method was found to have a sensitivity of 0.816. In summary, we demonstrate that networks of residues with high MI provide a distinct signature on CR and propose that such a signature should be present in other classes of functional residues where the requirement to maintain a particular function places limitations on the diversification of the structural environment along the course of evolution.
Peptide binding predictions for HLA DR, DP and DQ molecules

BACKGROUND: MHC class II binding predictions are widely used to identify epitope candidates in infectious agents, allergens, cancer and autoantigens. The vast majority of prediction algorithms for human MHC class II to date have targeted HLA molecules encoded in the DR locus. This reflects a significant gap in knowledge as HLA DP and DQ molecules are presumably equally important, and have only been studied less because they are more difficult to handle experimentally. RESULTS: In this study, we aimed to narrow this gap by providing a large scale dataset of over 17,000 HLA-peptide binding affinities for a set of 11 HLA DP and DQ alleles. We also expanded our dataset for HLA DR alleles resulting in a total of 40,000 MHC class II binding affinities covering 26 allelic variants. Utilizing this dataset, we generated prediction tools utilizing several machine learning algorithms and evaluated their performance. CONCLUSION: We found that 1) prediction methodologies developed for HLA DR molecules perform equally well for DP or DQ molecules. 2) Prediction performances were significantly increased compared to previous reports due to the larger amounts of training data available. 3) The presence of homologous peptides between training and testing datasets should be avoided to give real-world estimates of prediction performance metrics, but the relative ranking of different predictors is largely unaffected by the presence of homologous peptides, and predictors intended for end-user applications should include all training data for maximum performance. 4) The recently developed NN-align prediction method significantly outperformed all other algorithms, including a naïve consensus based on all prediction methods. A new consensus method dropping the comparably weak ARB prediction method could outperform the NN-align method, but further research into how to best combine MHC class II binding predictions is required.
State of the art and challenges in sequence based T-cell epitope prediction

Sequence based T-cell epitope predictions have improved immensely in the last decade. From predictions of peptide binding to major histocompatibility complex molecules with moderate accuracy, limited allele coverage, and no good estimates of the other events in the antigen-processing pathway, the field has evolved significantly. Methods have now been developed that produce highly accurate binding predictions for many alleles and integrate both proteasomal cleavage and transport events. Moreover, so-called pan-specific methods have been developed, which allow for prediction of peptide binding to MHC alleles characterized by limited or no peptide binding data. Most of the developed methods are publicly available, and have proven to be very useful as a shortcut in epitope discovery. Here, we will go through some of the history of sequence-based predictions of helper as well as cytotoxic T cell epitopes. We will focus on some of the most accurate methods and their basic background.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Lundegaard, C. (Intern), Hoof, I. (Intern), Lund, O. (Intern), Nielsen, M. (Intern)
Pages: 21067545
Publication date: 2010
Main Research Area: Technical/natural sciences

Publication Information
Journal: Immunome Research
Volume: 3
Issue number: 6
ISSN (Print): 1745-7580
Ratings:
Scopus rating (2016): SJR 0.175 SNIP 0.656
Scopus rating (2015): SJR 1.412 SNIP 1.064
Scopus rating (2014): SJR 1.809 SNIP 0.969
Scopus rating (2013): SJR 1.478 SNIP 1.067
ISI indexed (2013): ISI indexed no
Scopus rating (2012): SJR 0.734 SNIP 0.521
ISI indexed (2012): ISI indexed no
Scopus rating (2011): SJR 1.379 SNIP 0.461
ISI indexed (2011): ISI indexed no
Scopus rating (2010): SJR 1.776 SNIP 0.754
Scopus rating (2009): SJR 1.306 SNIP 0.672
Scopus rating (2008): SJR 0.741 SNIP 0.829
Original language: English
Electronic versions:
B55BDd01.pdf
DOIs:
10.1186/1745-7580-6-S2-S3

Bibliographical note
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Source: orbit
Source-ID: 272210
Publication: Research - peer-review › Journal article – Annual report year: 2010

Structural Properties of MHC Class II Ligands, Implications for the Prediction of MHC Class II Epitopes

Major Histocompatibility class II (MHC-II) molecules sample peptides from the extracellular space allowing the immune system to detect the presence of foreign microbes from this compartment. Prediction of MHC class II ligands is complicated by the open binding cleft of the MHC class II molecule, allowing binding of peptides extending out of the binding groove. Furthermore, only a few HLA-DR alleles have been characterized with a sufficient number of peptides (100–200 peptides per allele) to derive accurate description of their binding motif. Little work has been performed characterizing structural properties of MHC class II ligands. Here, we perform one such large-scale analysis. A large set of SYFPEITHI MHC class II ligands covering more than 20 different HLA-DR molecules was analyzed in terms of their secondary structure and surface exposure characteristics in the context of the native structure of the corresponding source
protein. We demonstrated that MHC class II ligands are significantly more exposed and have significantly more coil content than other peptides in the same protein with similar predicted binding affinity. We next exploited this observation to derive an improved prediction method for MHC class II ligands by integrating prediction of MHC-peptide binding with prediction of surface exposure and protein secondary structure. This combined prediction method was shown to significantly outperform the state-of-the-art MHC class II peptide binding prediction method when used to identify MHC class II ligands. We also tried to integrate N- and O-glycosylation in our prediction methods but this additional information was found not to improve prediction performance. In summary, these findings strongly suggest that local structural properties influence antigen processing and/or the accessibility of peptides to the MHC class II molecule.
The MHC motif viewer: a visualization tool for MHC binding motifs

In vertebrates, the onset of cellular immune reactions is controlled by presentation of peptides in complex with major histocompatibility complex (MHC) molecules to T cell receptors. In humans, MHCs are called human leukocyte antigens (HLAs). Different MHC molecules present different subsets of peptides, and knowledge of their binding specificities is important for understanding differences in the immune response between individuals. Algorithms predicting which peptides bind a given MHC molecule have recently been developed with high prediction accuracy. The utility of these algorithms is hampered by the lack of tools for browsing and comparing specificity of these molecules. We have developed a Web server, MHC Motif Viewer, which allows the display of the binding motif for MHC class I proteins for human, chimpanzee, rhesus monkey, mouse, and swine, as well as HLA-DR protein sequences. The binding motif for each MHC molecule is predicted using state-of-the-art, pan-specific peptide-MHC binding-prediction methods, and is visualized as a sequence logo, in a format that allows for a comprehensive interpretation of binding motif anchor positions and amino acid preferences.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Rapin, N. (Intern), Hoof, I. (Intern), Lund, O. (Intern), Nielsen, M. (Intern)
Publication date: 2010
Main Research Area: Technical/natural sciences

Publication information
Journal: Current Protocols in Immunology
ISSN (Print): 1934-3671
Ratings:
Web of Science (2018): Indexed yes
Scopus rating (2017): SNIP 1.086 SJR 2.172
Scopus rating (2016): SJR 1.393 SNIP 0.74 CiteScore 1.72
Scopus rating (2015): SJR 1.56 SNIP 0.685 CiteScore 1.59
Scopus rating (2014): SJR 1.074 SNIP 0.569 CiteScore 0.88
Scopus rating (2013): SJR 1.469 SNIP 0.586 CiteScore 1.05
ISI indexed (2013): ISI indexed no
Scopus rating (2012): SJR 1.516 SNIP 0.626 CiteScore 1.26
ISI indexed (2012): ISI indexed no
Scopus rating (2011): SJR 1.222 SNIP 0.75 CiteScore 1.32
ISI indexed (2011): ISI indexed no
Scopus rating (2010): SJR 1.16 SNIP 0.936
Scopus rating (2009): SJR 0.614 SNIP 0.481
Scopus rating (2008): SJR 0.132
Scopus rating (2007): SJR 0.1
Scopus rating (2006): SJR 0.17
Scopus rating (2005): SJR 0.144
Scopus rating (2004): SJR 0.11
Scopus rating (2003): SJR 0.106
Scopus rating (2002): SJR 0.1
A generic method for assignment of reliability scores applied to solvent accessibility predictions

Estimation of the reliability of specific real value predictions is nontrivial and the efficacy of this is often questionable. It is important to know if you can trust a given prediction and therefore the best methods associate a prediction with a reliability score or index. For discrete qualitative predictions, the reliability is conventionally estimated as the difference between output scores of selected classes. Such an approach is not feasible for methods that predict a biological feature as a single real value rather than a classification. As a solution to this challenge, we have implemented a method that predicts the relative surface accessibility of an amino acid and simultaneously predicts the reliability for each prediction, in the form of a Z-score. RESULTS: An ensemble of artificial neural networks has been trained on a set of experimentally solved protein structures to predict the relative exposure of the amino acids. The method assigns a reliability score to each surface accessibility prediction as an inherent part of the training process. This is in contrast to the most commonly used procedures where reliabilities are obtained by post-processing the output. CONCLUSION: The performance of the neural networks was evaluated on a commonly used set of sequences known as the CB513 set. An overall Pearson's correlation coefficient of 0.72 was obtained, which is comparable to the performance of the currently best public available method, Real-SPINE. Both methods associate a reliability score with the individual predictions. However, our implementation of reliability scores in the form of a Z-score is shown to be the more informative measure for discriminating good predictions from bad ones in the entire range from completely buried to fully exposed amino acids. This is evident when comparing the Pearson's correlation coefficient for the upper 20% of predictions sorted according to reliability. For this subset, values of 0.79 and 0.74 are obtained using our and the compared method, respectively. This tendency is true for any selected subset.
Correction for phylogeny, small number of observations and data redundancy improves the identification of coevolving amino acid pairs using mutual information

Motivation: Mutual information (MI) theory is often applied to predict positional correlations in a multiple sequence alignment (MSA) to make possible the analysis of those positions structurally or functionally important in a given fold or protein family. Accurate identification of coevolving positions in protein sequences is difficult due to the high background signal imposed by phylogeny and noise. Several methods have been proposed using MI to identify coevolving amino acids in protein families. Results: After evaluating two current methods, we demonstrate how the use of sequence-weighting techniques to reduce sequence redundancy and low-count corrections to account for small number of observations in limited size sequence families, can significantly improve the predictability of MI. The evaluation is made on large sets of both in silico-generated alignments as well as on biological sequence data. The methods included in the analysis are the APC (average product correction) and RCW (row-column weighting) methods. The best performing method was APC including sequence-weighting and low-count corrections. The use of sequence-permutations to calculate a MI rescaling is shown to significantly improve the prediction accuracy and allows for direct comparison of information values across protein families. Finally, we demonstrate how a lower bound of 400 sequences

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Buslje, C. (Ekstern), Santos, J. (Ekstern), Delfino, J. (Ekstern), Nielsen, M. (Intern)
Pages: 1125-1131
Publication date: 2009
Main Research Area: Technical/natural sciences

Publication information
Full genome sequence of a Danish isolate of Mycobacterium avium subspecies paratuberculosis, strain Ejlskov2007

We have sequenced a Danish isolate of Mycobacterium avium subspecies paratuberculosis, strain Ejlskov2007. The strain was isolated from faecal material of a 48 month old second parity Danish Holstein cow, with clinical symptoms of chronic diarrhoea and emaciation. The cultures were grown on Löwenstein-Jensen media by standard procedures and passed once to new tubes before DNA extraction and being sequenced on a 454 FLX machine. Currently, the genome has been assembled into 70 contiguous pieces, for a total of around 5.0 Mbp, with a 63% GC content. We have predicted
a total of 4687 proteins, consisting of 4317 unique gene families. Comparison with M. avium paratuberculosis strain K10 revealed only 3436 genes in common (~70%). We have used GenomeAtlases to show conserved (and unique) regions along the Ejlskov2007 chromosome, compared to 2 other Mycobacterium avium sequenced genomes. Pan-genome analyses of the sequenced Mycobacterium genomes reveal a surprisingly open and diverse set of genes for this bacterial genera.

**General information**

State: Published
Organisations: Adaptive Immunology & Parasitology, Division of Veterinary Diagnostics and Research, National Veterinary Institute, Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark
Publication date: 2009
Event: Poster session presented at 10th International Colloquium on Paratuberculosis, Minneapolis, MN, United States.
Main Research Area: Technical/natural sciences
Source: orbit
Source-ID: 262316
Publication: Research › Poster – Annual report year: 2009

**NetMHCpan, a method for MHC class I binding prediction beyond humans.**

Binding of peptides to major histocompatibility complex (MHC) molecules is the single most selective step in the recognition of pathogens by the cellular immune system. The human MHC genomic region (called HLA) is extremely polymorphic comprising several thousand alleles, each encoding a distinct MHC molecule. The potentially unique specificity of the majority of HLA alleles that have been identified to date remains uncharacterized. Likewise, only a limited number of chimpanzee and rhesus macaque MHC class I molecules have been characterized experimentally. Here, we present NetMHCpan-2.0, a method that generates quantitative predictions of the affinity of any peptide-MHC class I interaction. NetMHCpan-2.0 has been trained on the hitherto largest set of quantitative MHC binding data available, covering HLA-A and HLA-B, as well as chimpanzee, rhesus macaque, gorilla, and mouse MHC class I molecules. We show that the NetMHCpan-2.0 method can accurately predict binding to uncharacterized HLA molecules, including HLA-C and HLA-G. Moreover, NetMHCpan-2.0 is demonstrated to accurately predict peptide binding to chimpanzee and macaque MHC class I molecules. The power of NetMHCpan-2.0 to guide immunologists in interpreting cellular immune
responses in large out-bred populations is demonstrated. Further, we used NetMHCpan-2.0 to predict potential binding peptides for the pig MHC class I molecule SLA-1*0401. Ninety-three percent of the predicted peptides were demonstrated to bind stronger than 500 nM. The high performance of NetMHCpan-2.0 for non-human primates documents the method's ability to provide broad allelic coverage also beyond human MHC molecules. The method is available at http://www.cbs.dtu.dk/services/NetMHCpan.

**General information**
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, National Veterinary Institute
Authors: Hoof, I. (Intern), Peters, B. (Ekstern), Sidney, J. (Ekstern), Pedersen, L. E. (Intern), Sette, A. (Ekstern), Lund, O. (Intern), Buus, S. (Ekstern), Nielsen, M. (Intern)
Pages: 1-13
Publication date: 2009
Main Research Area: Technical/natural sciences

**Publication information**
Journal: Immunogenetics
Volume: 61
Issue number: 1
ISSN (Print): 0093-7711
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.638 SJR 0.916
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.2 SJR 1.249 SNIP 0.716
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.573 SNIP 0.813 CiteScore 2.47
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.228 SNIP 0.823 CiteScore 2.28
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.303 SNIP 0.771 CiteScore 2.48
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.382 SNIP 0.943 CiteScore 2.91
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.32 SNIP 0.885 CiteScore 2.83
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.217 SNIP 0.848
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 1.502 SNIP 0.843
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 1.408 SNIP 0.774
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 1.266 SNIP 0.742
Scopus rating (2006): SJR 1.232 SNIP 0.767
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 1.565 SNIP 0.82
Web of Science (2005): Indexed yes
Scopus rating (2004): SJR 1.535 SNIP 0.923
Web of Science (2004): Indexed yes
Scopus rating (2003): SJR 1.382 SNIP 0.713
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 1.357 SNIP 0.712
Scopus rating (2001): SJR 1.264 SNIP 0.639
Scopus rating (2000): SJR 1.206 SNIP 0.663
Web of Science (2000): Indexed yes
Scopus rating (1999): SJR 1.336 SNIP 0.902

Original language: English
Binding specificity, MHC class I, Artificial neural networks, CTL epitopes, Non-human primates

DOIs:
10.1007/s00251-008-0341-z

Links:
http://www.springerlink.com/content/n8l42545722u8261/

Source: orbit
Source-ID: 231396
Publication: Research - peer-review › Journal article – Annual report year: 2009

NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction
The major histocompatibility complex (MHC) molecule plays a central role in controlling the adaptive immune response to infections. MHC class I molecules present peptides derived from intracellular proteins to cytotoxic T cells, whereas MHC class II molecules stimulate cellular and humoral immunity through presentation of extracellularly derived peptides to helper T cells. Identification of which peptides will bind a given MHC molecule is thus of great importance for the understanding of host-pathogen interactions, and large efforts have been placed in developing algorithms capable of predicting this binding event. RESULTS: Here, we present a novel artificial neural network-based method, NN-align that allows for simultaneous identification of the MHC class II binding core and binding affinity. NN-align is trained using a novel training algorithm that allows for correction of bias in the training data due to redundant binding core representation. Incorporation of information about the residues flanking the peptide-binding core is shown to significantly improve the prediction accuracy. The method is evaluated on a large-scale benchmark consisting of six independent data sets covering 14 human MHC class II alleles, and is demonstrated to outperform other state-of-the-art MHC class II prediction methods. CONCLUSION: The NN-align method is competitive with the state-of-the-art MHC class II peptide binding prediction algorithms. The method is publicly available at http://www.cbs.dtu.dk/services/NetMHCII-2.0.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Nielsen, M. (Intern), Lund, O. (Intern)
Pages: 296
Publication date: 2009
Main Research Area: Technical/natural sciences

Publication information
Journal: B M C Bioinformatics
Volume: 10
ISSN (Print): 1471-2105
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.878 SJR 1.479
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.54 SJR 1.581 SNIP 0.974
Web of Science (2016): Indexed yes
Pan-specific MHC class I predictors: A benchmark of HLA class I pan-specific prediction methods

Motivation: MHC:peptide binding plays a central role in activating the immune surveillance. Computational approaches to determine T-cell epitopes restricted to any given MHC molecule are of special practical value in the development of for instance vaccines with broad population coverage against emerging pathogens. Methods have recently been published that are able to predict peptide binding to any human MHC class I molecule. In contrast to conventional allele-specific methods, these methods do allow for extrapolation to un-characterized MHC molecules. These pan-specific HLA predictors have not previously been compared using independent evaluation sets. Results: A diverse set of quantitative
peptide binding affinity measurements was collected from IEDB, together with a large set of HLA class I ligands from the SYFPEITHI database. Based on these data sets, three different pan-specific HLA web-accessible predictors NetMHCpan, Adaptive-Double-.Threading (ADT), and KISS were evaluated. The performance of the pan-specific predictors was also compared to a well performing allele-specific MHC class I predictor, NetMHC, as well as a consensus approach integrating the predictions from the NetMHC and NetMHCpan methods. Conclusions: The benchmark demonstrated that pan-specific methods do provide accurate predictions also for previously uncharacterized MHC molecules. The NetMHCpan method trained to predict actual binding affinities was consistently top ranking both on quantitative (affinity) and binary (ligand) data. However, the KISS method trained to predict binary data was one of the best performing when benchmarked on binary data. Finally, a consensus method integrating predictions from the two best-performing methods was shown to improve the prediction accuracy. Associate Editor: Prof. Thomas Lengauer.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis
Authors: Zhang, H. (Intern), Lundegaard, C. (Intern), Nielsen, M. (Intern)
Pages: 83-89
Publication date: 2009
Main Research Area: Technical/natural sciences

Publication information
Journal: Bioinformatics
Volume: 25
Issue number: 1
ISSN (Print): 1367-4803
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 6.42
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 6.06
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 5.5
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 5.78
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): CiteScore 6.73
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): CiteScore 5.61
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 2
Web of Science (2008): Indexed yes
Web of Science (2007): Indexed yes
Web of Science (2006): Indexed yes
Peptide Binding to HLA Class I Molecules: Homogenous, High-Throughput Screening, and Affinity Assays

The Human MHC Project aims at large-scale description of peptide-HLA binding to a wide range of HLA molecules covering all populations of the world and the accompanying generation of bioinformatics tools capable of predicting binding of any given peptide to any given HLA molecule. Here, the authors present a homogenous, proximity-based assay for detection of peptide binding to HLA class I molecules. It uses a conformation-dependent anti-HLA class I antibody, W6/32, as one tag and a biotinylated recombinant HLA class I molecule as the other tag, and a proximity-based signal is generated through the luminescent oxygen channeling immunoassay technology (abbreviated LOCI and commercialized as AlphaScreen (TM)). Compared with an enzyme-linked immunosorbent assay-based peptide-HLA class I binding assay, the LOCI assay yields virtually identical affinity measurements, although having a broader dynamic range, better signal-to-background ratios, and a higher capacity. They also describe an efficient approach to screen peptides for binding to HLA molecules. For the occasional user, this will serve as a robust, simple peptide-HLA binding assay. For the more dedicated user, it can easily be performed in a high-throughput screening mode using standard liquid handling robotics and 384-well plates. We have successfully applied this assay to more than 60 different HLA molecules, leading to more than 2 million measurements. (Journal of Biomolecular Screening 2009: 173-180)
Systematic Characterisation of Cellular Localisation and Expression Profiles of Proteins Containing MHC Ligands

Background: Presentation of peptides on Major Histocompatibility Complex (MHC) molecules is the cornerstone in immune system activation and increased knowledge of the characteristics of MHC ligands and their source proteins is highly desirable. Methodology/Principal Finding: In the present large-scale study, we used a large data set of proteins containing experimentally identified MHC class I or II ligands and examined the proteins according to their expression profiles at the mRNA level and their Gene Ontology (GO) classification within the cellular component ontology. Proteins encoded by highly abundant mRNA were found to be much more likely to be the source of MHC ligands. Of the 2.5% most abundant mRNAs as much as 41% of the proteins encoded by these mRNAs contained MHC class I ligands. For proteins containing MHC class II ligands, the corresponding percentage was 11%. Furthermore, we found that most proteins containing MHC class I ligands were localised to the intracellular parts of the cell including the cytoplasm and nucleus. MHC class II ligand donors were, on the other hand, mostly membrane proteins. Conclusions/Significance: The results contribute to the ongoing debate concerning the nature of MHC ligand-containing proteins and can be used to extend the existing methods for MHC ligand predictions by including the source protein's localisation and expression profile. Improving the current methods is important in the growing quest for epitopes that can be used for vaccine or diagnostic purposes, especially when it comes to large DNA viruses and cancer.
Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers

Several accurate prediction systems have been developed for prediction of class I major histocompatibility complex (MHC):peptide binding. Most of these are trained on binding affinity data of primarily 9mer peptides. Here, we show how prediction methods trained on 9mer data can be used for accurate binding affinity prediction of peptides of length 8, 10, and 11. The method gives the opportunity to predict peptides with a different length than nine for MHC alleles where no such peptides have been measured. As validation, the performance of this approach is compared to predictors trained on peptides of the peptide length in question. In this validation, the approximation method has an accuracy that is comparable to or better than methods trained on a peptide length identical to the predicted peptides.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Lundegaard, C. (Intern), Lund, O. (Intern), Nielsen, M. (Intern)
Pages: 1397-1398
Publication date: 2008
Amino acid similarity accounts for T cell cross-reactivity and for “holes” in the T cell repertoire

Background: Cytotoxic T cell (CTL) cross-reactivity is believed to play a pivotal role in generating immune responses but the extent and mechanisms of CTL cross-reactivity remain largely unknown. Several studies suggest that CTL clones can...
recognize highly diverse peptides, some sharing no obvious sequence identity. The emerging realization in the field is that T cell receptors (TcR) recognize multiple distinct ligands. Principal Findings: First, we analyzed peptide scans of the HIV epitope SLFNTVATL (SFL9) and found that TCR specificity is position dependent and that biochemically similar amino acid substitutions do not drastically affect recognition. Inspired by this, we developed a general model of TCR peptide recognition using amino acid similarity matrices and found that such a model was able to predict the cross-reactivity of a diverse set of CTL epitopes. With this model, we were able to demonstrate that seemingly distinct T cell epitopes, i.e., ones with low sequence identity, are in fact more biochemically similar than expected. Additionally, an analysis of HIV immunogenicity data with our model showed that CTLs have the tendency to respond mostly to peptides that do not resemble self-antigens. Conclusions: T cell cross-reactivity can thus, to an extent greater than earlier appreciated, be explained by amino acid similarity. The results presented in this paper will help resolving some of the long-lasting discussions in the field of T cell cross-reactivity.
Broadly Immunogenic HLA Class I Supertype-Restricted Elite CTL Epitopes Recognized in a Diverse Population Infected with Different HIV-1 Subtypes

The genetic variations of the HIV-1 virus and its human host constitute major obstacles for obtaining potent HIV-1-specific CTL responses in individuals of diverse ethnic backgrounds infected with different HIV-1 variants. In this study, we developed and used a novel algorithm to select 184 predicted epitopes representing seven different HLA class I supertypes that together constitute a broad coverage of the different HIV-1 strains as well as the human HLA alleles. Of the tested 184 HLA class I-restricted epitopes, 114 were recognized by at least one study subject, and 45 were novel epitopes, not previously described in the HIV-1 immunology database. In addition, we identified 21 "elite" epitopes that induced CTL responses in at least 4 of the 31 patients. A majority (27 of 31) of the study population recognized one or more of these highly immunogenic epitopes. We also found a limited set of 9 epitopes that together induced HIV-1-specific CTL responses in all HIV-1-responsive patients in this study. Our results have important implications for the validation of potent CTL responses and show that the goal for a vaccine candidate in inducing broadly reactive CTL immune responses is attainable.
Humans with chimpanzee-like major histocompatibility complex-specificities control HIV-1 infection

Background: Major histocompatibility complex (MHC) class I molecules allow immune surveillance by presenting a snapshot of the intracellular state of a cell to circulating cytotoxic T lymphocytes. The MHC class I alleles of an HIV-1 infected individual strongly influence the level of viremia and the progression rate to AIDS. Chimpanzees control HIV-1 viral replication and develop a chronic infection without progressing to AIDS. A similar course of disease is observed in human long-term non-progressors. Objective: To investigate if long-term non-progressors and chimpanzees have functional similarities in their MHC class I repertoire. Methods: We compared the specificity of groups of human MHC molecules associated with different levels of viremia in HIV-1 infected individuals with those of chimpanzee. Results and conclusion: We demonstrate that human MHC with control of HIV-1 viral load share binding motifs with chimpanzee MHC. Moreover, we find that chimpanzee and human MHC associated with low viral load are predicted to elicit broader Gag-specific immune responses than human MHC associated with high viral load, thus supporting earlier findings that Gag-specific immune responses are essential for HIV-1 control.
Immune epitope database analysis resource (IEDB-AR)

We present a new release of the immune epitope database analysis resource (IEDB-AR, http://tools.immuneepitope.org), a repository of web-based tools for the prediction and analysis of immune epitopes. New functionalities have been added to most of the previously implemented tools, and a total of eight new tools were added, including two B-cell epitope
prediction tools, four T-cell epitope prediction tools and two analysis tools.

**General information**

State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, La Jolla Institute for Allergy & Immunology, University of California
Authors: Zhang, Q. (Ekstern), Wang, P. (Intern), Kim, Y. (Ekstern), Andersen, P. (Intern), Beaver, J. (Ekstern), Bourne, P. (Ekstern), Sui, H. (Ekstern), Buus, S. (Intern), Pletscher-Frankild, S. (Intern), Greenbaum, J. (Ekstern), Lund, O. (Intern), Lundegaard, C. (Intern), Nielsen, M. (Intern), Ponomarenko, J. (Ekstern), Sette, A. (Ekstern), Zhu, Z. (Ekstern), Peters, B. (Ekstern)
Pages: W513-W518
Publication date: 2008
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Nucleic Acids Research
Volume: W36
ISSN (Print): 0305-1048
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 9.28 SJR 7.883 SNIP 2.744
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 7.358 SNIP 2.631 CiteScore 9.48
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 6.64 SNIP 2.552 CiteScore 8.74
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 6.801 SNIP 2.284 CiteScore 8.46
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 6.329 SNIP 2.407 CiteScore 8.62
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 5.976 SNIP 2.19 CiteScore 7.86
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 5.381 SNIP 2.034
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 5.669 SNIP 1.874
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 4.912 SNIP 1.578
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 5.1 SNIP 1.807
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 4.776 SNIP 2.051
Web of Science (2006): Indexed yes
MHC motif viewer

In vertebrates, the major histocompatibility complex (MHC) presents peptides to the immune system. In humans, MHCs are called human leukocyte antigens (HLAs), and some of the loci encoding them are the most polymorphic in the human genome. Different MHC molecules present different subsets of peptides, and knowledge of their binding specificities is important for understanding the differences in the immune response between individuals. Knowledge of motifs may be used to identify epitopes, to understand the MHC restriction of epitopes, and to compare the specificities of different MHC molecules. Algorithms that predict which peptides MHC molecules bind have recently been developed and cover many different alleles, but the utility of these algorithms is hampered by the lack of tools for browsing and comparing the specificity of these molecules. We have, therefore, developed a web server, MHC motif viewer, that allows the display of the likely binding motif for all human class I proteins of the loci HLA A, B, C, and E and for MHC class I molecules from chimpanzee (Pan troglodytes), rhesus monkey (Macaca mulatta), and mouse (Mus musculus). Furthermore, it covers all HLA-DR protein sequences. A special viewing feature, MHC fight, allows for display of the specificity of two different MHC molecules side by side. We show how the web server can be used to discover and display surprising similarities as well as differences between MHC molecules within and between different species. The MHC motif viewer is available at http://www.cbs.dtu.dk/biotools/MHCMotifViewer/.

General information

State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Rapin, N. P. J. (Intern), Hoof, I. (Intern), Lund, O. (Intern), Nielsen, M. (Intern)
Pages: 759-765
Publication date: 2008
Main Research Area: Technical/natural sciences

Publication information

Journal: Immunogenetics
Volume: 60
Issue number: 12
ISSN (Print): 0093-7711
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.638 SJR 0.916
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.2 SJR 1.249 SNIP 0.716
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.573 SNIP 0.813 CiteScore 2.47
NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11

NetMHC-3.0 is trained on a large number of quantitative peptide data using both affinity data from the Immune Epitope Database and Analysis Resource (IEDB) and elution data from SYFPEITHI. The method generates high-accuracy predictions of major histocompatibility complex (MHC): peptide binding. The predictions are based on artificial neural networks trained on data from 55 MHC alleles (43 Human and 12 non-human), and position-specific scoring matrices (PSSMs) for additional 67 HLA alleles. As only the MHC class I prediction server is available, predictions are possible for peptides of length 8–11 for all 122 alleles. Artificial neural network predictions are given as actual IC50 values whereas PSSM predictions are given as a log-odds likelihood scores. The output is optionally available as download for easy post-processing. The training method underlying the server is the best available, and has been used to predict possible MHC-binding peptides in a series of pathogen viral proteomes including SARS, Influenza and HIV, resulting in an average of...
75–80% confirmed MHC binders. Here, the performance is further validated and benchmarked using a large set of newly published affinity data, non-redundant to the training set. The server is free of use and available at:
http://www.cbs.dtu.dk/services/NetMHC.

**General information**

- **State:** Published
- **Organisations:** Center for Biological Sequence Analysis, Department of Systems Biology
- **Authors:** Lundegaard, C. (Intern), Lamberth, K. (Ekstern), Harndahl, M. (Ekstern), Buus, S. (Ekstern), Lund, O. (Intern), Nielsen, M. (Intern)
- **Pages:** W509-W512
- **Publication date:** 2008
- **Main Research Area:** Technical/natural sciences

**Publication information**

- **Journal:** Nucleic Acids Research
- **Volume:** 36
- **ISSN (Print):** 0305-1048
- **Ratings:**
  - BFI (2018): BFI-level 2
  - Web of Science (2018): Indexed yes
  - BFI (2017): BFI-level 2
  - Web of Science (2017): Indexed yes
  - BFI (2016): BFI-level 2
  - Scopus rating (2016): CiteScore 9.28 SJR 7.883 SNIP 2.744
  - Web of Science (2016): Indexed yes
  - BFI (2015): BFI-level 2
  - Scopus rating (2015): SJR 7.358 SNIP 2.631 CiteScore 9.48
  - Web of Science (2015): Indexed yes
  - BFI (2014): BFI-level 2
  - Scopus rating (2014): SJR 6.64 SNIP 2.552 CiteScore 8.74
  - Web of Science (2014): Indexed yes
  - BFI (2013): BFI-level 2
  - Scopus rating (2013): SJR 6.801 SNIP 2.284 CiteScore 8.46
  - ISI indexed (2013): ISI indexed yes
  - Web of Science (2013): Indexed yes
  - BFI (2012): BFI-level 2
  - Scopus rating (2012): SJR 6.329 SNIP 2.407 CiteScore 8.62
  - ISI indexed (2012): ISI indexed yes
  - Web of Science (2012): Indexed yes
  - BFI (2011): BFI-level 2
  - Scopus rating (2011): SJR 5.976 SNIP 2.19 CiteScore 7.86
  - ISI indexed (2011): ISI indexed yes
  - Web of Science (2011): Indexed yes
  - BFI (2010): BFI-level 2
  - Scopus rating (2010): SJR 5.381 SNIP 2.034
  - Web of Science (2010): Indexed yes
  - BFI (2009): BFI-level 2
  - Scopus rating (2009): SJR 5.669 SNIP 1.874
  - BFI (2008): BFI-level 2
  - Scopus rating (2008): SJR 4.912 SNIP 1.578
  - Web of Science (2008): Indexed yes
  - Scopus rating (2007): SJR 5.1 SNIP 1.807
  - Web of Science (2007): Indexed yes
  - Scopus rating (2006): SJR 4.776 SNIP 2.051
  - Web of Science (2006): Indexed yes
  - Scopus rating (2005): SJR 5.092 SNIP 2.147
Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan

CD4 positive T helper cells control many aspects of specific immunity. These cells are specific for peptides derived from protein antigens and presented by molecules of the extremely polymorphic major histocompatibility complex (MHC) class II system. The identification of peptides that bind to MHC class II molecules is therefore of pivotal importance for rational discovery of immune epitopes. HLA-DR is a prominent example of a human MHC class II. Here, we present a method, NetMHCIIpan, that allows for pan-specific predictions of peptide binding to any HLA-DR molecule of known sequence. The method is derived from a large compilation of quantitative HLA-DR binding events covering 14 of the more than 500 known HLA-DR alleles. Taking both peptide and HLA sequence information into account, the method can generalize and predict peptide binding also for HLA-DR molecules where experimental data is absent. Validation of the method includes identification of endogenously derived HLA class II ligands, cross-validation, leave-one-molecule-out, and binding motif identification for hitherto uncharacterized HLA-DR molecules. The validation shows that the method can successfully predict binding for HLA-DR molecules—even in the absence of specific data for the particular molecule in question. Moreover, when compared to TEPITOPE, currently the only other publicly available prediction method aiming at providing broad HLA-DR allelic coverage, NetMHCIIpan performs equivalently for alleles included in the training of TEPITOPE while outperforming TEPITOPE on novel alleles. We propose that the method can be used to identify those hitherto uncharacterized alleles, which should be addressed experimentally in future updates of the method to cover the polymorphism of HLA-DR most efficiently. We thus conclude that the presented method meets the challenge of keeping up with the MHC polymorphism discovery rate and that it can be used to sample the MHC "space," enabling a highly efficient iterative process for improving MHC class II binding predictions.
The peptide-binding specificity of HLA-A*3001 demonstrates membership of the HLA-A3 supertype

Human leukocyte antigen class I (HLA-I) molecules are highly polymorphic peptide receptors, which select and present endogenously derived peptide epitopes to CD8+ cytotoxic T cells (CTL). The specificity of the HLA-I system is an important component of the overall specificity of the CTL immune system. Unfortunately, the large and rapidly increasing number of known HLA-I molecules seriously complicates a comprehensive analysis of the specificities of the entire HLA-I system (as of June 2008, the international HLA registry holds >1,650 unique HLA-I protein entries). In an attempt to reduce this complexity, it has been suggested to cluster the different HLA-I molecules into "supertypes" of largely overlapping peptide-binding specificities. Obviously, the HLA supertype concept is only valuable if membership can be assigned with reasonable accuracy. The supertype assignment of HLA-A*3001, a common HLA haplotype in populations of African descent, has variously been assigned to the A1, A3, or A24 supertypes. Using a biochemical HLA-A*3001 binding assay, and a large panel of nonamer peptides and peptide libraries, we here demonstrate that the specificity of HLA-A*3001 most closely resembles that of the HLA-A3 supertype. We discuss approaches to supertype assignment and underscore the importance of experimental verification.
CTL epitopes for influenza A including the H5N1 bird flu; genome-, pathogen-, and HLA-wide screening

The purpose of the present study is to perform a global screening for new immunogenic HLA class I (HLA-I) restricted cytotoxic T cell (CTL) epitopes of potential utility as candidates of influenza A-virus diagnostics and vaccines. We used predictions of antigen processing and presentation, the latter encompassing 12 different HLA class I supertypes with >99% population coverage, and searched for conserved epitopes from available influenza A viral protein sequences. Peptides corresponding to 167 predicted peptide-HLA-1 interactions were synthesized, tested for peptide-HLA-1 interactions in a biochemical assay and for influenza-specific, HLA-1-restricted CTL responses in an IFN-gamma ELISPOT assay. Eighty-nine peptides could be confirmed as HLA-1 binders, and 13 could be confirmed as CTL targets. The 13 epitopes, are highly conserved among human influenza A pathogens, and all of these epitopes are present in the emerging bird flu isolates. Our study demonstrates that present technology enables a fast global screening for T cell immune epitopes of potential diagnostics and vaccine interest. This technology includes immuno-bioinformatics predictors with the capacity to perform fast genome-, pathogen-, and HLA-wide searches for immune targets. To exploit this new potential, a coordinated international effort to analyze the precious source of information represented by rare patients, such as the current victims of bird flu, would be essential.
Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Larsen, M. V. (Intern), Lundegaard, C. (Intern), Lambert, K. (Ekstern), Buus, S. (Ekstern), Lund, O. (Intern), Nielsen, M. (Intern)
Pages: 424
Publication date: 2007
Main Research Area: Technical/natural sciences

Publication information
Journal: BMC Bioinformatics
Volume: 8
ISSN (Print): 1471-2105
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.878 SJR 1.479
Web of Science (2017): Indexed yes

Original language: English
CTL epitope, peptide, influenza A virus, diagnostics, vaccine

DOIs:
10.1016/j.vaccine.2006.12.038
Source: orbit
Source-ID: 213822
Publication: Research - peer-review › Journal article – Annual report year: 2007
NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence

BACKGROUND: Binding of peptides to Major Histocompatibility Complex (MHC) molecules is the single most selective step in the recognition of pathogens by the cellular immune system. The human MHC class I system (HLA-I) is extremely polymorphic. The number of registered HLA-I molecules has now surpassed 1500. Characterizing the specificity of each separately would be a major undertaking. PRINCIPAL FINDINGS: Here, we have drawn on a large database of known peptide-HLA-I interactions to develop a bioinformatics method, which takes both peptide and HLA sequence information into account, and generates quantitative predictions of the affinity of any peptide-HLA-I interaction. Prospective experimental validation of peptides predicted to bind to previously untested HLA-I molecules, cross-validation, and retrospective prediction of known HIV immune epitopes and endogenous presented peptides, all successfully validate this method. We further demonstrate that the method can be applied to perform a clustering analysis of MHC specificities and suggest using this clustering to select particularly informative novel MHC molecules for future biochemical and functional analysis. CONCLUSIONS: Encompassing all HLA molecules, this high-throughput computational method lends itself to epitope searches that are not only genome- and pathogen-wide, but also HLA-wide. Thus, it offers a truly global analysis of immune responses supporting rational development of vaccines and immunotherapy. It also promises to provide new basic insights into HLA structure-function relationships. The method is available at http://www.cbs.dtu.dk/services/NetMHCpan.
Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method

BACKGROUND: Antigen presenting cells (APCs) sample the extra cellular space and present peptides from here to T helper cells, which can be activated if the peptides are of foreign origin. The peptides are presented on the surface of the cells in complex with major histocompatibility class II (MHC II) molecules. Identification of peptides that bind MHC II molecules is thus a key step in rational vaccine design and developing methods for accurate prediction of the peptide:MHC interactions play a central role in epitope discovery. The MHC class II binding groove is open at both ends making the correct alignment of a peptide in the binding groove a crucial part of identifying the core of an MHC class II binding motif. Here, we present a novel stabilization matrix alignment method, SMM-align, that allows for direct prediction of peptide:MHC binding affinities. The predictive performance of the method is validated on a large MHC class II benchmark data set covering 14 HLA-DR (human MHC) and three mouse H2-IA alleles. RESULTS: The predictive performance of the SMM-align method was demonstrated to be superior to that of the Gibbs sampler, TEPITOPE, SVRMHC, and MHCpred methods. Cross validation between peptide data set obtained from different sources demonstrated that direct incorporation of peptide length potentially results in over-fitting of the binding prediction method. Focusing on amino terminal peptide flanking residues (PFR), we demonstrate a consistent gain in predictive performance by favoring binding registers with a minimum PFR length of two amino acids. Visualizing the binding motif as obtained by the SMM-align and TEPITOPE methods highlights a series of fundamental discrepancies between the two predicted motifs. For the DRB1*1302 allele for instance, the TEPITOPE method favors basic amino acids at most anchor positions, whereas the SMM-align method identifies a preference for hydrophobic or neutral amino acids at the anchors. CONCLUSION: The SMM-align method was shown to outperform other state of the art MHC class II prediction methods. The method predicts quantitative peptide:MHC binding affinity values, making it ideally suited for rational epitope discovery. The method has been trained and evaluated on the, to our knowledge, largest benchmark data set publicly available and covers the nine HLA-DR supertypes suggested as well as three mouse H2-IA allele. Both the peptide benchmark data set, and SMM-align prediction method (NetMHCII) are made publicly available.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Nielsen, M. (Intern), Lundegaard, C. (Intern), Lund, O. (Intern)
Pages: 238
Publication date: 2007
Main Research Area: Technical/natural sciences

Publication information
Journal: BMC Bioinformatics
Volume: 8
ISSN (Print): 1471-2105
Ratings:
BFI (2018): BFI-level 1
A community resource benchmarking predictions of peptide binding to MHC-I molecules

**General information**

State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Pages: 574-584
Publication date: 2006
Main Research Area: Technical/natural sciences

**Publication information**

Journal: *P L o S Computational Biology (Online)*
Volume: 2
Issue number: 6
ISSN (Print): 1553-7358
Ratings:
- BFI (2018): BFI-level 1
- Web of Science (2018): Indexed yes
- BFI (2017): BFI-level 1
- Scopus rating (2017): SNIP 1.348 SJR 3.097
- Web of Science (2017): Indexed yes
- BFI (2016): BFI-level 1
- Scopus rating (2016): CiteScore 4.41 SJR 3.243 SNIP 1.363
- Web of Science (2016): Indexed yes
- BFI (2015): BFI-level 1
- Scopus rating (2015): SJR 3.476 SNIP 1.442 CiteScore 4.69
- Web of Science (2015): Indexed yes
- BFI (2014): BFI-level 1
- Scopus rating (2014): SJR 3.412 SNIP 1.442 CiteScore 4.74
- Web of Science (2014): Indexed yes
- BFI (2013): BFI-level 1
- Scopus rating (2013): SJR 3.467 SNIP 1.483 CiteScore 4.91
- ISI indexed (2013): ISI indexed yes
- Web of Science (2013): Indexed yes
- Scopus rating (2012): SJR 3.523 SNIP 1.645 CiteScore 5.36
- ISI indexed (2012): ISI indexed no
- Web of Science (2012): Indexed yes
- Scopus rating (2011): SJR 3.613 SNIP 1.591 CiteScore 5.25
- ISI indexed (2011): ISI indexed no
- Web of Science (2011): Indexed yes
- Scopus rating (2010): SJR 3.709 SNIP 1.555
- Web of Science (2010): Indexed yes
- Scopus rating (2009): SJR 3.428 SNIP 1.428
- Web of Science (2009): Indexed yes
- Scopus rating (2008): SJR 4.045 SNIP 1.397
- Web of Science (2008): Indexed yes
- Scopus rating (2007): SJR 3.396 SNIP 1.329
- Scopus rating (2006): SJR 2.419 SNIP 1.082
- Web of Science (2006): Indexed yes

Original language: English
Electronic versions:
- journal.pcbi.0020065.pdf
- DOIs: 10.1371/journal.pcbi.0020065
Improved method for predicting linear B-cell epitopes

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis
Authors: Larsen, J. E. P. (Intern), Lund, O. (Intern), Nielsen, M. (Intern)
Pages: 2
Publication date: 2006
Main Research Area: Technical/natural sciences

Publication information
Journal: Immunome Research
Volume: 2
ISSN (Print): 1745-7580
Ratings:
Scopus rating (2016): SJR 0.175 SNIP 0.656
Scopus rating (2015): SJR 1.412 SNIP 1.064
Scopus rating (2014): SJR 1.809 SNIP 0.969
Scopus rating (2013): SJR 1.478 SNIP 1.067
ISI indexed (2013): ISI indexed no
Scopus rating (2012): SJR 0.734 SNIP 0.521
ISI indexed (2012): ISI indexed no
Scopus rating (2011): SJR 1.379 SNIP 0.461
ISI indexed (2011): ISI indexed no
Scopus rating (2010): SJR 1.776 SNIP 0.754
Scopus rating (2009): SJR 1.306 SNIP 0.672
Scopus rating (2008): SJR 0.741 SNIP 0.829
Original language: English
Electronic versions:
1745-7580-2-2.pdf
DOIs:
10.1186/1745-7580-2-2
Source: orbit
Source-ID: 193398
Publication: Research › Journal article – Annual report year: 2006

Modelling the human immune system by combining bioinformatics and systems biology approaches

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Pages: 335-353
Publication date: 2006
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of Biological Physics
Volume: 32
Issue number: 3-4
ISSN (Print): 0092-0606
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
No evidence for the use of DIR, D-D fusions, chromosome 15 open reading frames or V(H)replacement in the peripheral repertoire was found on application of an improved algorithm, JointML, to 6329 human immunoglobulin H rearrangements.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Ohm-Laursen, L. (Ekstern), Nielsen, M. (Intern), Larsen, T. S. (Intern), Barington, T. (Ekstern)
Pages: 265-277
Publication date: 2006
Main Research Area: Technical/natural sciences

Publication information
Journal: Immunology
Prediction of residues in discontinuous B-cell epitopes using protein 3D structures

Discovery of discontinuous B-cell epitopes is a major challenge in vaccine design. Previous epitope prediction methods have mostly been based on protein sequences and are not very effective. Here, we present DiscoTope, a novel method for discontinuous epitope prediction that uses protein three-dimensional structural data. The method is based on amino acid statistics, spatial information, and surface accessibility in a compiled data set of discontinuous epitopes determined by X-ray crystallography of antibody/antigen protein complexes. DiscoTope is the first method to focus explicitly on discontinuous epitopes. We show that the new structure-based method has a better performance for predicting residues of discontinuous epitopes than methods based solely on sequence information, and that it can successfully predict epitope residues that have been identified by different techniques. DiscoTope detects 15.5% of residues located in discontinuous epitopes with a specificity of 95%. At this level of specificity, the conventional Parker hydrophilicity scale for predicting linear B-cell epitopes identifies only 11.0% of residues located in discontinuous epitopes. Predictions by the DiscoTope method can guide experimental epitope mapping in both rational vaccine design and development of diagnostic tools, and may lead to more efficient epitope identification.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Andersen, P. (Ekstern), Nielsen, M. (Intern), Lund, O. (Intern)
Pages: 2558-2567
Publication date: 2006
Main Research Area: Technical/natural sciences

Publication information
Journal: Protein Science
Volume: 15
Issue number: 11
ISSN (Print): 0961-8368
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.809 SJR 1.652
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): SJR 1.73 SNIP 0.784 CiteScore 2.68
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 2.009 SNIP 0.901 CiteScore 2.99
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.825 SNIP 0.846 CiteScore 2.77
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.995 SNIP 0.813 CiteScore 2.96
ISI indexed (2013): ISI indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.964 SNIP 0.816 CiteScore 2.76
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.973 SNIP 0.953 CiteScore 2.94
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.776 SNIP 0.807
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 2.29 SNIP 0.997
TAP-independent MHC class I presentation.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Larsen, M. V. (Intern), Nielsen, M. (Intern), Weinzierl, A. (Ekstern), Lund, O. (Intern)
Pages: 233-245
Publication date: 2006
Main Research Area: Technical/natural sciences

Publication information
Journal: Current Immunology Reviews
Volume: 2
Issue number: 3
ISSN (Print): 1573-3955
Ratings:
Web of Science (2018): Indexed yes
Scopus rating (2017): SNIP 0.065 SJR 0.124
Scopus rating (2016): SJR 0.271 SNIP 0.24 CiteScore 0.53
Scopus rating (2015): SJR 0.347 SNIP 0.195 CiteScore 0.65
Scopus rating (2014): SJR 0.208 SNIP 0.141 CiteScore 0.35
Scopus rating (2013): SJR 0.351 SNIP 0.243 CiteScore 0.81
Scopus rating (2012): SJR 0.332 SNIP 0.208 CiteScore 0.86
Scopus rating (2011): SJR 0.379 SNIP 0.208 CiteScore 0.8
Scopus rating (2010): SJR 0.308 SNIP 0.113
Scopus rating (2009): SJR 0.332 SNIP 0.147
Scopus rating (2008): SJR 0.294 SNIP 0.345
Scopus rating (2007): SJR 0.218 SNIP 0.129
Original language: English
Source: orbit
Source-ID: 193390
Publication: Research - peer-review › Journal article – Annual report year: 2006

The validity of predicted T-cell epitopes

General information
An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Panum Institute
Authors: Larsen, M. V. (Intern), Lundegaard, C. (Intern), Lamberth, K. (Ekstern), Buus, S. (Ekstern), Brunak, S. (Intern), Lund, O. (Intern), Nielsen, M. (Intern)
Pages: 2295-2303
Publication date: 2005
Main Research Area: Technical/natural sciences

Publication information
Journal: European Journal of Immunology
Volume: 35
Issue number: 8
ISSN (Print): 0014-2980
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.92 SJR 2.206
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.61 SJR 2.525 SNIP 0.927
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 2.588 SNIP 0.965 CiteScore 3.85
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 2.672 SNIP 0.972 CiteScore 3.83
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 2.876 SNIP 1.05 CiteScore 4.3
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 2.989 SNIP 1.063 CiteScore 4.62
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 3.255 SNIP 1.025 CiteScore 4.69
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 3.363 SNIP 0.99
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Nielsen, M. (Intern), Lundegaard, C. (Intern), Lund, O. (Intern), Kesmir, C. (Intern)
Pages: 33-41
Publication date: 2005
Main Research Area: Technical/natural sciences

Publication information
Journal: Immunogenetics
Volume: 57
Issue number: 1-2
ISSN (Print): 0093-7711
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.638 SJR 0.916
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.2 SJR 1.249 SNIP 0.716
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.573 SNIP 0.813 CiteScore 2.47
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.228 SNIP 0.823 CiteScore 2.28
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.303 SNIP 0.771 CiteScore 2.48
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
Definition of supertypes for HLA molecules using clustering of specificity matrices

Major histocompatibility complex (MHC) proteins are encoded by extremely polymorphic genes and play a crucial role in immunity. However, not all genetically different MHC molecules are functionally different. Sette and Sidney (1999) have defined nine HLA class I supertypes and showed that with only nine main functional binding specificities it is possible to cover the binding properties of almost all known HLA class I molecules. Here we present a comprehensive study of the functional relationship between all HLA molecules with known specificities in a uniform and automated way. We have developed a novel method for clustering sequence motifs. We construct hidden Markov models for HLA class I molecules using a Gibbs sampling procedure and use the similarities among these to define clusters of specificities. These clusters are extensions of the previously suggested ones. We suggest splitting some of the alleles in the A1 supertype into a new A26 supertype, and some of the alleles in the B27 supertype into a new B39 supertype. Furthermore the B8 alleles may define their own supertype. We also use the published specificities for a number of HLA-DR types to define clusters with similar specificities. We report that the previously observed specificities of these class II molecules can be clustered into nine classes, which only partly correspond to the serological classification. We show that classification of HLA molecules may be done in a uniform and automated way. The definition of clusters allows for selection of representative HLA molecules that can cover the HLA specificity space better. This makes it possible to target most of the known HLA alleles with known specificities using only a few peptides, and may be used in construction of vaccines. Supplementary material is available at http://www.cbs.dtu.dk/researchgroups/immunology/supertypes.html.
Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach

Prediction of which peptides will bind a specific major histocompatibility complex (MHC) constitutes an important step in identifying potential T-cell epitopes suitable as vaccine candidates. MHC class II binding peptides have a broad length distribution complicating such predictions. Thus, identifying the correct alignment is a crucial part of identifying the core of an MHC class II binding motif. In this context, we wish to describe a novel Gibbs motif sampler method ideally suited for recognizing such weak sequence motifs. The method is based on the Gibbs sampling method, and it incorporates novel features optimized for the task of recognizing the binding motif of MHC classes I and II. The method locates the binding motif in a set of sequences and characterizes the motif in terms of a weight-matrix. Subsequently, the weight-matrix can be applied to identifying effectively potential MHC binding peptides and to guiding the process of rational vaccine design.

Results: We apply the motif sampler method to the complex problem of MHC class II binding. The input to the method is amino acid peptide sequences extracted from the public databases of SYFPEITHI and MHCPEP and known to bind to the MHC class II complex HLA-DR4(B1*0401). Prior identification of information-rich (anchor) positions in the binding motif is shown to improve the predictive performance of the Gibbs sampler. Similarly, a consensus solution obtained from an ensemble average over suboptimal solutions is shown to outperform the use of a single optimal solution. In a large-scale benchmark calculation, the performance is quantified using relative operating characteristics curve (ROC) plots and we make a detailed comparison of the performance with that of both the TEPITOPE method and a weight-matrix derived using the conventional alignment algorithm of ClustalW. The calculation demonstrates that the predictive performance of the Gibbs sampler is higher than that of ClustalW and in most cases also higher than that of the TEPITOPE method.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Nielsen, M. (Intern), Lundegaard, C. (Intern), Worning, P. (Intern), Hvid, C. (Ekstern), Lamberth, K. (Ekstern), Buus, S. (Ekstern), Brunak, S. (Intern), Lund, O. (Intern)
Pages: 1388-1397
Publication date: 2004
Main Research Area: Technical/natural sciences

Publication information
Journal: Bioinformatics
Volume: 20
ISSN (Print): 1367-4803
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 6.42
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 6.06
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 5.5
Web of Science (2014): Indexed yes
MHC class I epitope binding prediction trained on small data sets
The identification of potential T-cell epitopes is important for development of new human or veterinary vaccines, both considering single protein/subunit vaccines, and for epitope/peptide vaccines as such. The highly diverse MHC class I alleles bind very different peptides, and accurate binding prediction methods exist only for alleles were the binding pattern have been deduced from peptide motifs. Using empirical knowledge of important anchor positions within the binding peptides dramatically reduces the number of peptides needed for reliable predictions. We here present a general method for predicting peptides binding to specific MHC class I alleles. The method combines advanced automatic scoring matrix generation with empirical position specific differential anchor weighting. The method leads to predictions with a comparable or higher accuracy than other established prediction servers, even in situations where only very limited data are available for training.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Lundegaard, C. (Intern), Nielsen, M. (Intern), Lamberth, K. (Ekstern), Worning, P. (Intern), Sylvester-Hvid, C. (Ekstern), Buus, S. (Ekstern), Brunak, S. (Intern), Lund, O. (Intern)
Pages: 217-225
Publication date: 2004

Host publication information
Volume: 3239
Publisher: Springer

Series: Lecture Notes in Computer Science
Volume: 3239
Modeling lipid-sterol bilayers: Applications to structural evolution, lateral diffusion, and rafts

General information
State: Published
Organisations: Department of Systems Biology
Authors: Zuckermann, M. (Ekstern), Ipsen, J. (Ekstern), Miao, L. (Ekstern), Mouritsen, O. (Ekstern), Nielsen, M. (Intern), Polson, J. (Ekstern), Thewalt, J. (Ekstern), Vattulainen, I. (Ekstern), Zhu, H. (Ekstern)
Pages: 198-229
Publication date: 2004
Main Research Area: Technical/natural sciences

SARS CTL vaccine candidates; HLA supertype-, genome-wide scanning and biochemical validation

An effective Severe Acute Respiratory Syndrome (SARS) vaccine is likely to include components that can induce specific cytotoxic T-lymphocyte (CTL) responses. The specificities of such responses are governed by human leukocyte antigen (HLA)-restricted presentation of SARS-derived peptide epitopes. Exact knowledge of how the immune system handles protein antigens would allow for the identification of such linear sequences directly, from genomic/proteomic sequence information (Lauemoller et al., Rev Immunogenet 2001: 2: 477-91). The latter was recently established when a causative coronavirus (SARS-CoV) was isolated and full-length sequenced (Marra et al., Science 2003: 300: 1399404). Here, we have combined advanced bioinformatics and high-throughput immunology to perform an HLA supertype-, genome-wide scan for SARS-specific CTL epitopes. The scan includes all nine human HLA supertypes in total covering >99% of all individuals of all major human populations (Sette & Sidney, Immunogenetics 1999: 50: 201-12). For each HLA supertype, we have selected the 15 top candidates for test in biochemical binding assays. At this time (approximately 2-6 months after the genome was established), we have tested the majority of the HLA supertypes and identified almost 100 potential vaccine candidates. These should be further validated in SARS survivors and used for vaccine formulation. We suggest that immunobioinformatics may become a fast and valuable tool in rational vaccine design.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Pages: 295-400
Publication date: 2004
Main Research Area: Technical/natural sciences
Reliable prediction of T-cell epitopes using neural networks with novel sequence representations

In this paper we describe an improved neural network method to predict T-cell class I epitopes. A novel input representation has been developed consisting of a combination of sparse encoding, Blosum encoding, and input derived from hidden Markov models. We demonstrate that the combination of several neural networks derived using different sequence-encoding schemes has a performance superior to neural networks derived using a single sequence-encoding scheme. The new method is shown to have a performance that is substantially higher than that of other methods. By use of mutual information calculations we show that peptides that bind to the HLA A*0204 complex display signal of higher order sequence correlations. Neural networks are ideally suited to integrate such higher order correlations when predicting the binding affinity. It is this feature combined with the use of several neural networks derived from different and novel sequence-encoding schemes and the ability of the neural network to be trained on data consisting of continuous binding affinities that gives the new method an improved performance. The difference in predictive performance between the neural network methods and that of the matrix-driven methods is found to be most significant for peptides that bind strongly to the HLA molecule, confirming that the signal of higher order sequence correlation is most strongly present in high-binding peptides. Finally, we use the method to predict T-cell epitopes for the genome of hepatitis C virus and discuss possible applications of the prediction method to guide the process of rational vaccine design.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Nielsen, M. (Intern), Lundegaard, C. (Intern), Worning, P. (Intern), Lauemoller, S. (Ekstern), Lamberth, K. (Ekstern), Buus, S. (Ekstern), Brunak, S. (Intern), Lund, O. (Intern)
Pages: 1007-1017
Publication date: 2003
Selecting informative data for developing peptide-MHC binding predictors using a query by committee approach

Strategies for selecting informative data points for training prediction algorithms are important, particularly when data points are difficult and costly to obtain. A Query by Committee (QBC) training strategy for selecting new data points uses the disagreement between a committee of different algorithms to suggest new data points, which most rationally complement existing data. In order to evaluate this QBC approach on a real-world problem, we compared strategies for selecting new data points. We trained neural network algorithms to obtain methods to predict the binding affinity of peptides binding to the MHC class I molecule, HLA-A2. We show that the QBC strategy leads to a higher performance than a baseline strategy where new data points are selected at random from a pool of available data. Most peptides bind HLA-A2 with a low affinity, and as expected using a strategy of selecting peptides that are predicted to have high binding affinities also lead to more accurate predictors than the baseline strategy. The QBC value is shown to correlate with the measured binding affinity. This demonstrates that the different predictors can easily learn if a peptide will fail to bind, but often conflict in predicting if a peptide binds. Using a carefully constructed computational setup, we demonstrate that selecting peptides with a high QBC performs better than low QBC peptides independently from binding affinity. When predictors are trained on a very limited set of data they cannot be expected to disagree in a meaningful way and we find a data limit below which the QBC strategy fails. Finally, it should be noted that data selection strategies similar to those used here might be of use in other settings in which generation of more data is a costly process.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Christensen, J. (Ekstern), Lamberth, K. (Ekstern), Nielsen, M. (Intern), Lundegaard, C. (Intern), Worning, P. (Intern), Lauemoller, S. (Ekstern), Buus, S. (Ekstern), Brunak, S. (Intern), Lund, O. (Intern)
Pages: 2931-2942
Publication date: 2003
Main Research Area: Technical/natural sciences

Publication information
Journal: Neural Computation
Volume: 15
Issue number: 12
ISSN (Print): 0899-7667
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): SNIP 1.069 SJR 0.896
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 2.47 SJR 0.833 SNIP 1.175
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 1.107 SNIP 1.143 CiteScore 2.5
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 0.964 SNIP 1.133 CiteScore 2.52
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 0.827 SNIP 1.009 CiteScore 2.39
ISI indexed (2013): ISI indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 0.853 SNIP 1.36 CiteScore 2.48
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 1.312 SNIP 1.385 CiteScore 2.59
From lanosterol to cholesterol: Structural evolution and differential effects on lipid bilayers

Cholesterol is an important molecular component of the plasma membranes of mammalian cells. Its precursor in the sterol biosynthetic pathway, lanosterol, has been argued by Konrad Bloch (Bloch, K. 1965. Science. 150:19-28; 1983. CRC Crit Rev. Biochem. 14:47-92; 1994. Blonds in Venetian Paintings, the Nine-Banded Armadillo, and Other Essays in Biochemistry. Yale University Press, New Haven, CT.) to also be a precursor in the molecular evolution of cholesterol. We present a comparative study of the effects of cholesterol and lanosterol on molecular conformational order and phase equilibria of lipid-bilayer membranes. By using deuterium NMR spectroscopy on multilamellar lipid-sterol systems in combination with Monte Carlo simulations of microscopic models of lipid-sterol interactions, we demonstrate that the evolution in the molecular chemistry from lanosterol to cholesterol is manifested in the model lipid-sterol membranes by an increase in the ability of the sterols to promote and stabilize a particular membrane phase, the liquid-ordered phase, and to induce collective order in the acyl-chain conformations of lipid molecules. We also discuss the biological relevance of our results, in particular in the context of membrane domains and rafts.

General information
State: Published
Organisations: Department of Chemistry
Pages: 1429-1444
Publication date: 2002
Main Research Area: Technical/natural sciences

Publication information
Journal: Biophysical journal
Volume: 82
Issue number: 3
ISSN (Print): 0006-3495
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
**Web-based Tools for Vaccine Design**

**General information**
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Center for Biological sequence analysis, Technical University of Denmark
Authors: Lund, O. (Intern), Nielsen, M. (Intern), Kesmir, C. (Intern), Christensen, J. K. (Ekstern), Lundegaard, C. (Intern), Worning, P. (Intern), Brunak, S. (Intern)
Pages: 45-51
Publication date: 2002

**Host publication information**
Title of host publication: HIV molecular immunology 2002
Publisher: Theoretical Biology & Biophysics
Main Research Area: Technical/natural sciences

**Bibliographical note**
Published by Theoretical Biology & Biophysics. Group T-10, Mail Stop K710. Los Alamos National Laboratory, Los Alamos, New Mexico 87545 U.S.A.
http://hiv-web.lanl.gov/immunology
Publication: Research - peer-review › Book chapter – Annual report year: 2002

**Sterol evolution and the physics of membranes**

**General information**
State: Published
Organisations: Department of Chemistry
Pages: 368-374
Publication date: 2000
Main Research Area: Technical/natural sciences

**Publication information**
Journal: Europhysics Letters
Volume: 52
ISSN (Print): 0295-5075
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): SNIP 0.569 SJR 0.498
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 1.18 SJR 0.549 SNIP 0.603
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 0.625 SNIP 0.593 CiteScore 1.12
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 0.555 SNIP 0.579 CiteScore 1.04
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 0.542 SNIP 0.539 CiteScore 1
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Lipid bilayers exhibit a phase behavior that involves two distinct, but coupled, order-disorder processes, one in terms of lipid-chain crystalline packing (translational degrees of freedom) and the other in terms of lipid-chain conformational ordering (internal degrees of freedom). Experiments and previous approximate theories have suggested that cholesterol incorporated into lipid bilayers has different microscopic effects on lipid-chain packing and conformations and that cholesterol thereby leads to decoupling of the two ordering processes, manifested by a special equilibrium phase, "liquid-ordered phase," where bilayers are liquid (with translational disorder) but lipid chains are conformationally ordered. We present in this paper a microscopic model that describes this decoupling phenomena and which yields a phase diagram consistent with experimental observations. The model is an off-lattice model based on a two-dimensional random triangulation algorithm and represents lipid and cholesterol molecules by hard-core particles with internal (spin-type) degrees of freedom that have nearest-neighbor interactions. The phase equilibria described by the model, specifically in terms of phase diagrams and structure factors characterizing different phases, are calculated by using several Monte Carlo simulation techniques, including histogram and thermodynamic reweighting techniques, finite-size scaling as well as non-Boltzmann sampling techniques (in order to overcome severe hysteresis effects associated with strongly first-order phase transitions). The results provide a consistent interpretation of the various phases of phospholipid-cholesterol binary mixtures based on the microscopic dual action of cholesterol on the lipid-chain degrees of freedom. In particular, a distinct small-scale structure of the liquid-ordered phase has been identified and characterized. The generic nature of the model proposed holds a promise for a unifying description for a whole series of different lipid-sterol mixtures. [S1063-651X(99)09305-8].

Off-lattice model for the phase behavior of lipid-cholesterol bilayers

Lipid bilayers exhibit a phase behavior that involves two distinct, but coupled, order-disorder processes, one in terms of lipid-chain crystalline packing (translational degrees of freedom) and the other in terms of lipid-chain conformational ordering (internal degrees of freedom). Experiments and previous approximate theories have suggested that cholesterol incorporated into lipid bilayers has different microscopic effects on lipid-chain packing and conformations and that cholesterol thereby leads to decoupling of the two ordering processes, manifested by a special equilibrium phase, "liquid-ordered phase," where bilayers are liquid (with translational disorder) but lipid chains are conformationally ordered. We present in this paper a microscopic model that describes this decoupling phenomena and which yields a phase diagram consistent with experimental observations. The model is an off-lattice model based on a two-dimensional random triangulation algorithm and represents lipid and cholesterol molecules by hard-core particles with internal (spin-type) degrees of freedom that have nearest-neighbor interactions. The phase equilibria described by the model, specifically in terms of phase diagrams and structure factors characterizing different phases, are calculated by using several Monte Carlo simulation techniques, including histogram and thermodynamic reweighting techniques, finite-size scaling as well as non-Boltzmann sampling techniques (in order to overcome severe hysteresis effects associated with strongly first-order phase transitions). The results provide a consistent interpretation of the various phases of phospholipid-cholesterol binary mixtures based on the microscopic dual action of cholesterol on the lipid-chain degrees of freedom. In particular, a distinct small-scale structure of the liquid-ordered phase has been identified and characterized. The generic nature of the model proposed holds a promise for a unifying description for a whole series of different lipid-sterol mixtures. [S1063-651X(99)09305-8].

General information
State: Published
The effect of cholesterol on the rupture of lipid membrane

General information
State: Published
Organisations: Department of Chemistry
Authors: Nielsen, M. (Intern), Ipsen, J. H. (Intern), Miao, L. (Intern), Mouritsen, O. G. (Intern), Zuckermann, M. J. (Ekstern)
Pages: A140
Publication date: 1999
Main Research Area: Technical/natural sciences

Publication information
Journal: Biophysical Journal
Volume: 76
Issue number: 1
ISSN (Print): 0006-3495
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.979 SJR 1.949
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.06 SJR 1.988 SNIP 1.005
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 2.13 SNIP 1.134 CiteScore 3.3
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 2.21 SNIP 1.15 CiteScore 3.33
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 2.245 SNIP 1.156 CiteScore 3.64
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 2.361 SNIP 1.143 CiteScore 3.57
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 2.357 SNIP 1.202 CiteScore 3.75
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 2.695 SNIP 1.303
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 3.016 SNIP 1.357
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 2
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 3.161 SNIP 1.413
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 2.857 SNIP 1.419
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 2.715 SNIP 1.403
Web of Science (2005): Indexed yes
Scopus rating (2004): SJR 2.538 SNIP 1.503
Web of Science (2004): Indexed yes
Scopus rating (2003): SJR 2.669 SNIP 1.444
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 2.415 SNIP 1.452
Web of Science (2002): Indexed yes
Scopus rating (2001): SJR 2.445 SNIP 1.368
Web of Science (2001): Indexed yes
Scopus rating (2000): SJR 2.91 SNIP 1.384
Web of Science (2000): Indexed yes
Scopus rating (1999): SJR 2.861 SNIP 1.439
Original language: English
Source: orbit
Source-ID: 174164
Publication: Research - peer-review › Journal article – Annual report year: 1999

Projects:

**T Cell Immunoinformatics**
Department of Bio and Health Informatics
Period: 01/04/2018 → 31/03/2021
Number of participants: 3
Phd Student:
Reynisson, Birkir (Intern)
Supervisor:
Marcatili, Paolo (Intern)
Main Supervisor:
Nielsen, Morten (Intern)

**Financing sources**
Source: Internal funding (public)
Name of research programme: Samfinansieret - Andet
Project: PhD

**Towards accurate prediction of T cell targets: Learning the rules of T cell receptor interaction**
Department of Micro- and Nanotechnology
Period: 01/10/2017 → 30/09/2020
Number of participants: 3
Phd Student:
Holm, Jeppe Sejerø (Intern)
Supervisor:
Nielsen, Morten (Intern)
Main Supervisor:
Hadrup, Sine Reker (Intern)

Financing sources
Source: Internal funding (public)
Name of research programme: Forskningsrådsfinansiering
Project: PhD

Pre-clinical exploration of cancer neoepitope immunotherapy
Department of Bio and Health Informatics
Period: 01/01/2017 → 31/12/2019
Number of participants: 3
Phd Student:
Jappe, Emma Christine (Intern)
Supervisor:
Kringelum, Jens Vindahl (Intern)
Main Supervisor:
Nielsen, Morten (Intern)

Financing sources
Source: Internal funding (public)
Name of research programme: Industrial PhD
Project: PhD

Metagenomic Data Stratified using Artificial Intelligence
Department of Bio and Health Informatics
Period: 01/11/2016 → 21/03/2020
Number of participants: 3
Phd Student:
Nissen, Jakob Nybo (Intern)
Supervisor:
Nielsen, Morten (Intern)
Main Supervisor:
Rasmussen, Simon (Intern)

Financing sources
Source: Internal funding (public)
Name of research programme: Institut stipendie (DTU)
Project: PhD

B-cell immunoinformatics
Department of Bio and Health Informatics
Period: 01/08/2016 → 31/07/2019
Number of participants: 3
Phd Student:
Jespersen, Martin Closter (Intern)
Supervisor:
Marcatili, Paolo (Intern)
Main Supervisor:
Nielsen, Morten (Intern)

Financing sources
Source: Internal funding (public)
Name of research programme: Samfinansieret - Andet
Project: PhD
Development of Immunoinformatics prediction methods for improved understanding of TCR-peptide-MHC interactions

Department of Bio and Health Informatics  
Period: 01/08/2016 → 31/07/2019  
Number of participants: 3  
Phd Student:  
Jensen, Kamilla Kjærgaard (Intern)  
Supervisor:  
Marcatili, Paolo (Intern)  
Main Supervisor:  
Nielsen, Morten (Intern)

Financing sources  
Source: Internal funding (public)  
Name of research programme: Samfinansieret - Andet  
Project: PhD

The gut microbiota and bone dynamics

Department of Bio and Health Informatics  
Period: 01/02/2015 → 31/08/2017  
Number of participants: 4  
Phd Student:  
Bresciani, Anne Gæther (Intern)  
Supervisor:  
Nielsen, Henrik Bjørn (Intern)  
Nielsen, Morten (Intern)  
Main Supervisor:  
Nielsen, Morten (Intern)

Financing sources  
Source: Internal funding (public)  
Name of research programme: Institut stipendie (DTU)  
Project: PhD

High Performance Machine Learning Methods applied within Bioinformatics

Department of Bio and Health Informatics  
Period: 15/12/2014 → 22/01/2018  
Number of participants: 7  
Phd Student:  
Jurtz, Vanessa Isabell (Intern)  
Supervisor:  
Lund, Ole (Intern)  
Winther, Ole (Intern)  
Main Supervisor:  
Nielsen, Morten (Intern)  
Examiner:  
Petersen, Thomas Nordahl (Intern)  
Buus, Søren (Ekstern)  
Gfeller, David (Ekstern)

Financing sources  
Source: Internal funding (public)  
Name of research programme: Samfinansieret - Andet  
Project: PhD

Combating Methicillin Resistant Staphylococcus aureus (MRSA) with bacteriophages

Department of Bio and Health Informatics
Isolation and characterization of bacteriaophages with therapeutic potential

Department of Bio and Health Informatics
Period: 01/12/2013 → 28/02/2018
Number of participants: 7
Phd Student:
Villarroel, Julia (Intern)
Supervisor:
Kilstrup, Mogens (Intern)
Larsen, Mette Voldby (Intern)
Main Supervisor:
Nielsen, Morten (Intern)
Examiner:
Nielsen, Henrik (Intern)
Lavigne, Rob (Ekstern)
Nielsen, Dennis Sandris (Ekstern)

Financing sources
Source: Internal funding (public)
Name of research programme: Institut stipendie (DTU)

Human Personality - Identification of important genotypic and phenotypic factors for the development of an individual's personality

Department of Systems Biology
Period: 15/12/2012 → 29/02/2016
Number of participants: 6
Phd Student:
Wolffhechel, Karin Marie Brandt (Intern)
Supervisor:
Pedersen, Anders Gorm (Intern)
Main Supervisor:
Jarmer, Hanne Østergaard (Intern)
Examiner:
Nielsen, Morten (Intern)
Lindgren, Cecilia Margareta (Ekstern)

Financing sources
Source: Internal funding (public)
Name of research programme: Institut stipendie (DTU)
Project: PhD
Reuter, Martin (Ekstern)

**Financing sources**
*Source:* Internal funding (public)  
*Name of research programme:* Institut stipendie (DTU)  
*Project:* PhD

**Identification of the features contributing to immunodominance**
*Department of Systems Biology*  
*Period:* 15/12/2012 → 22/03/2016  
*Number of participants:* 6  
*Phd Student:*  
*Trolle, Thomas (Intern)*  
*Supervisor:*  
*Lund, Ole (Intern)*  
*Main Supervisor:*  
*Nielsen, Morten (Intern)*  
*Examiner:*  
*Eklund, Aron Charles (Intern)  
Buus, Søren (Ekstern)  
Ternette, Nicola M. N. (Ekstern)

**Financing sources**
*Source:* Internal funding (public)  
*Name of research programme:* Institut stipendie (DTU) Samf.  
*Project:* PhD

**Prediction of antigen_BCR interactions based on high throughput peptide chip data and prediction of MHC-peptide-TCR Interactions**
*Department of Systems Biology*  
*Period:* 01/08/2012 → 25/11/2015  
*Number of participants:* 6  
*Phd Student:*  
*Hansen, Christian Skjødt (Intern)*  
*Supervisor:*  
*Lund, Ole (Intern)*  
*Main Supervisor:*  
*Nielsen, Morten (Intern)*  
*Examiner:*  
*Eklund, Aron Charles (Intern)  
Agüero, Fernán (Ekstern)  
Andersen, Peter Sejer (Ekstern)

**Financing sources**
*Source:* Internal funding (public)  
*Name of research programme:* Institut stipendie (DTU) Samf.  
*Project:* PhD

**Udvikling af featurebaseret in silico-metoder til diskrimination af immunogene og beskyttende antigener**
*Department of Bio and Health Informatics*  
*Period:* 15/12/2011 → 21/04/2016  
*Number of participants:* 4  
*Phd Student:*  
*Mattsson, Andreas Holm (Intern)*  
*Supervisor:*  
*Møller, Niels Iversen (Ekstern)  
Poznansky, Mark C. (Ekstern)
Epitope prediction methods

Department of Systems Biology  
Period: 01/09/2010 → 27/11/2013  
Number of participants: 6  
PhD Student:  
Karosiene, Edita (Intern)  
Supervisor:  
Lund, Ole (Intern)  
Main Supervisor:  
Nielsen, Morten (Intern)  
Examiner:  
Sicheritz-Pontén, Thomas (Intern)  
Andersen, Peter Sejer (Ekstern)  
Asquith, Becca (Ekstern)  

Financing sources  
Source: Internal funding (public)  
Name of research programme: Institut stipendie (DTU) Samf.  
Project: PhD

Bioinformatics for high-density peptide microarrays

Department of Systems Biology  
Period: 01/10/2009 → 17/12/2012  
Number of participants: 6  
PhD Student:  
Andreatta, Massimo (Intern)  
Supervisor:  
Nielsen, Morten (Intern)  
Main Supervisor:  
Lund, Ole (Intern)  
Examiner:  
Winther, Ole (Intern)  
Bader, Gary (Ekstern)  
Willemoes, Martin (Intern)  

Financing sources  
Source: Internal funding (public)  
Name of research programme: 1/3 DTU-stip, 2/3 FUR/andet  
Project: PhD

T Cell Reactive Tetramers for Virus Infections in Pigs

Virus infections in livestock are constant threats to animal welfare and productivity all over the world. In this project we will deliver new advanced technological reagents for measurement of the cytotoxic cells of the immune system with activity against virus infected cells in pigs. This will be an extremely important tool in the development of new efficacious vaccines against diseases like foot- and mouth disease and influenza. All cells op the body exhibit small fragments of their contents on the cell surface. A virus infected cell will therefore display fragments of virus proteins which, like a key in a lock on the cytotoxic cell, will activate killing of the infected cell. This will stop replication of the virus and this cell-mediated immunity is therefore a crucial part of the host defence against virus infections. The virus-key is made up of host tissue-type molecules displaying a small virus peptide (a chain of 8 to 11 amino acids). We will produce luminescent recombinant virus-keys, MHC class I tetramers, for pigs, which will enable us to directly stain and identify host cell with cytotoxic activity against virus. With these tetramers we can determine exactly which peptides in the virus proteins that mediate the desired immune response, and thereby which virus components that can be used in new targeted vaccines. Furthermore, we will be able to
measure if vaccines have induced the desired cytotoxic effector cells, and we can develop computer models to predict peptide antigens of new viruses. The project group consists of scientists from Technical University of Denmark, Copenhagen University and leading American scientists in virus infections and MHC molecules in pigs.

Adaptive Immunology & Parasitology
Division of Veterinary Diagnostics and Research
National Veterinary Institute
Center for Biological Sequence Analysis
Department of Systems Biology
University of Copenhagen
United States Department of Agriculture
Period: 01/07/2009 → 31/12/2012
Number of participants: 5
Project ID: 22380 X-1
Project participant:
Pedersen, Lasse Eggert (Intern)
Nielsen, Morten (Intern)
Jungersen, Gregers (Intern)
Buus, Søren (Ekstern)
Golde, William T. (Ekstern)

Financing sources
Source: Forskningsrådene - Andre
Name of research programme: Forskningsrådene - Andre
Amount: 4,657,800.00 Danish Kroner

Development and Application of Potentials for Protein-Protein Interactions
Department of Systems Biology
Period: 15/10/2006 → 20/10/2010
Number of participants: 7
Phd Student:
Zhang, Hao (Intern)
Supervisor:
Lund, Ole (Intern)
Lundegaard, Claus (Intern)
Main Supervisor:
Nielsen, Morten (Intern)
Examiner:
Petersen, Thomas Nordahl (Intern)
Jørgensen, Flemming Steen (Ekstern)
Kesmir, Can (Intern)

Financing sources
Source: Internal funding (public)
Name of research programme: DTU-lønnet stipendie
Project: PhD

Materialedata og plasticitetsteori for COMPRECIT
Department of Civil Engineering
Period: 01/10/1992 → …
Number of participants: 5
Phd Student:
Nielsen, Morten (Intern)
Supervisor:
Jensen, Bjarne Christian (Ekstern)
Nepper-Christensen, Palle (Ekstern)
Main Supervisor:
Nielsen, Mogens Peter (Intern)
Examiner:
Sørensen, Hans-Christian (Intern)

**Financing sources**
Source: Internal funding (public)
Name of research programme: ATV- Gammel ordning
Project: PhD