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.

General information
State: Accepted/In press
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Universidad Nacional de San Martin, University of Copenhagen, La Jolla Institute for Allergy & Immunology
Authors: Jensen, K. K. (Intern), Andreatta, M. (Ekstern), Marcatili, P. (Intern), Buus, S. (Ekstern), Greenbaum, J. A. (Ekstern), Yan, Z. (Ekstern), Sette, A. (Ekstern), Peters, B. (Ekstern), Nielsen, M. (Intern)
Number of pages: 28
Publication date: 2018
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
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.74 SJR 1.888 SNIP 0.937
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 2.079 SNIP 0.975 CiteScore 3.83
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 2.029 SNIP 1.05 CiteScore 3.61
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 2.038 SNIP 1.083 CiteScore 3.97
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.902 SNIP 1.047 CiteScore 3.94
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.841 SNIP 0.993 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 0.897
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 0.122 SNIP 0.928
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 0.111 SNIP 0.921
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 0.122 SNIP 0.964
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 0.122 SNIP 0.887
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 1.64 SNIP 0.921
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.
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 cocktails. 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

Life-Course Genome-wide Association Study Meta-analysis of Total Body BMD and Assessment of Age-Specific Effects
Bone mineral density (BMD) assessed by DXA is used to evaluate bone health. In children, total body (TB) measurements are commonly used; in older individuals, BMD at the lumbar spine (LS) and femoral neck (FN) is used to diagnose osteoporosis. To date, genetic variants in more than 60 loci have been identified as associated with BMD. To investigate the genetic determinants of TB-BMD variation along the life course and test for age-specific effects, we performed a meta-analysis of 30 genome-wide association studies (GWASs) of TB-BMD including 66,628 individuals overall and divided across five age strata, each spanning 15 years. We identified variants associated with TB-BMD at 80 loci, of which 36 have not been previously identified; overall, they explain approximately 10% of the TB-BMD variance when combining all age groups and influence the risk of fracture. Pathway and enrichment analysis of the association signals showed clustering within gene sets implicated in the regulation of cell growth and SMAD proteins, overexpressed in the musculoskeletal system, and enriched in enhancer and promoter regions. These findings reveal TB-BMD as a relevant trait for genetic studies of osteoporosis, enabling the identification of variants and pathways influencing different bone compartments. Only variants in ESR1 and close proximity to RANKL showed a clear effect dependency on age. This most likely indicates that the majority of genetic variants identified influence BMD early in life and that their effect can be captured throughout the life course.

General information
State: Published
Organisations: Department of Bio and Health Informatics, National Veterinary Institute, Immunoinformatics and Machine Learning, T-cells & Cancer, Wake Forest School of Medicine, Winston-Salem, NC 27101, USA., Erasmus University Rotterdam, University of Queensland, University of Cambridge, The Children's Hospital of Philadelphia, University of Copenhagen, Leiden University, Federal University of Pelotas, California Pacific Medical Center, University of Eastern Finland, McGill University, Sir Charles Gairdner Hospital, Boston University, University of Gothenburg, deCODE/Amgen, Wayne State University, National Institute on Aging, University of Western Australia, University of Edinburgh, Hebrew SeniorLife, Lund University, University of California at Davis, King's College London, The Ohio State University, The National University Hospital of Iceland, Oregon Health and Science University, University of Washington, University of Ioannina, Mayo Clinic, Imperial College London, Garvan Institute of Medical Research, University of Rochester, University of Bristol
BMD, CREB3L1, ESR1, GWASs, RANKL, Age-dependent effects, Bone mineral density, Fracture, Genetic correlation, Genome-wide association studies, Meta-regression, Total-body DXA

DOIs:
10.1016/j.ajhg.2017.12.005

Source: FindIt
Source-ID: 2395194600

Publication: Research - peer-review › Journal article – Annual report year: 2018
3D protein-structure-oriented discovery of clinical relation across chronic lymphocytic leukemia patients

Chronic lymphocytic leukemia (CLL) is the most common adult leukemia with still unclear etiology. Indications of antigenic pressure have been hinted, using sequence and structure-based reasoning. The accuracy of such approaches, and in particular of the ones derived from 3D models obtained from the patients’ antibody amino acid sequences, is intimately connected to both the reliability of the models and the quality of the methods used to compare and group them. The proposed work provides a sophisticated method for the classification of CLL patients based on clustering the amino acid sequences of the clonotypic B-cell receptor immunoglobulin, which is the ideal clone-specific marker, critical for clonal behavior and patient outcome. A novel CLL patient clustering method is hereby proposed, combining bioinformatics methods with the extraction of 3D object descriptors, used in machine learning applications. The proposed methodology achieved an efficient and highly informative grouping of CLL patients in accordance to their biological and clinical properties.

General information
State: Published
Organisations: Department of Biotechnology and Biomedicine, Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Technical University of Denmark, Center For Research And Technology - Hellas, Carlsberg Research Laboratory
Authors: Mochament, K. (Ekstern), Agathangelidis, A. (Ekstern), Polychronidou, E. (Ekstern), Palaskas, C. (Ekstern), Kalamaras, E. (Ekstern), Moschonas, P. (Ekstern), Stamatopoulos, K. (Ekstern), Chailyan, A. (Ekstern), Overby, N. (Ekstern), Marcatili, P. (Intern), Hadzidimitriou, A. (Ekstern), Tzovaras, D. (Ekstern)
Pages: 139-150
Publication date: 2017
Conference: 5th International Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2017, Granada, Spain, April 26–28, 2017, Granada, Spain, 26/04/2017 - 26/04/2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Lecture Notes in Computer Science
Volume: 10209
ISSN (Print): 0302-9743
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 0.67 SJR 0.315 SNIP 0.552
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 0.328 SNIP 0.618 CiteScore 0.37
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 0.325 SNIP 0.678 CiteScore 0.42
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 0.329 SNIP 0.699 CiteScore 0.49
ISI indexed (2013): ISI indexed no
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 0.323 SNIP 0.708 CiteScore 0.49
ISI indexed (2012): ISI indexed no
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 0.325 SNIP 0.721 CiteScore 0.49
ISI indexed (2011): ISI indexed no
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 0.314 SNIP 0.634
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 0.305 SNIP 0.548
Web of Science (2009): 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 neopeptide-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 neopeptides 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
Web of Science (2017): Indexed yes
Scopus rating (2016): CiteScore 5.37 SJR 2.963 SNIP 1.483
An introduction to Deep learning on biological sequence data - Examples and solutions

Deep neural network architectures such as convolutional and long short-term memory networks have become increasingly popular as machine learning tools during the recent years. The availability of greater computational resources, more data, new algorithms for training deep models and easy to use libraries for implementation and training of neural networks are the drivers of this development. The use of deep learning has been especially successful in image recognition; and the development of tools, applications and code examples are in most cases centered within this field rather than within biology. Here, we aim to further the development of deep learning methods within biology by providing application examples and ready to apply and adapt code templates. Given such examples, we illustrate how architectures consisting of convolutional and long short-term memory neural networks can relatively easily be designed and trained to state-of-the-art performance on three biological sequence problems: prediction of subcellular localization, protein secondary structure and the binding of peptides to MHC Class II molecules. All implementations and datasets are available online to the scientific community at https://github.com/vanessajurtz/lasagne4bio. Supplementary data are available at Bioinformatics online.

General information

State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Department of Applied Mathematics and Computer Science, Department of Electrical Engineering, Disease Intelligence and Molecular Evolution, Copenhagen Center for Health Technology, Cognitive Systems, University of Copenhagen
Pages: 3685-3690
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information

Journal: Bioinformatics
Volume: 33
Issue number: 22
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
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.
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.
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.
Comparative proteomics of oxidative stress response of Lactobacillus acidophilus NCFM reveals effects on DNA repair and cysteine de novo synthesis

Probiotic cultures encounter oxidative conditions during manufacturing, yet protein abundance changes induced by such stress have not been characterized for some of the most common probiotics and starters. This comparative proteomics investigation focuses on the response by Lactobacillus acidophilus NCFM to H2O2, simulating an oxidative environment. Bacterial growth was monitored by BioScreen and batch cultures were harvested at exponential phase for protein profiling of stress responses by 2D gel-based comparative proteomics. Proteins identified in 19 of 21 spots changing in abundance due to H2O2 were typically related to carbohydrate and energy metabolism, cysteine biosynthesis, and stress. In particular, increased cysteine synthase activity may accumulate a cysteine pool relevant for protein stability, enzyme catalysis and the disulfide-reducing pathway. The stress response further included elevated abundance of biomolecules reducing damage such as enzymes from DNA repair pathways and metabolic enzymes with active site cysteine residues. By contrast, a protein-refolding chaperone showed reduced abundance, possibly reflecting severe oxidative protein destruction that was not overcome by refolding. The proteome analysis provides novel insight into resistance mechanisms in lactic acid bacteria against reactive oxygen species and constitutes a valuable starting point for improving industrial processes, food design or strain engineering preserving microorganism viability.

General information
State: Published
Organisations: Enzyme and Protein Chemistry, Department of Biotechnology and Biomedicine, Immunoinformatics and Machine Learning, University of Torino
Authors: Calderini, E. (Ekstern), Celebioglu, H. U. (Intern), Villarroel, J. (Intern), Jacobsen, S. (Intern), Svensson, B. (Intern), Pessione, E. (Ekstern)
Number of pages: 10
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Proteomics
Volume: 17
Issue number: 5
Article number: 1600178
ISSN (Print): 1615-9853
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.85 SJR 1.492 SNIP 0.89
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.464 SNIP 0.978 CiteScore 3.7
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.436 SNIP 0.981 CiteScore 3.73
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.48 SNIP 0.985 CiteScore 3.88
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.489 SNIP 1.099 CiteScore 4.1
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.677 SNIP 1.182 CiteScore 4.49
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.494 SNIP 1.127
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
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 this thesis focuses on applying genomics and mathematical modelling to questions
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.
High-density peptide microarray exploration of the antibody response in a rabbit immunized with a neurotoxic venom fraction

Polyvalent snakebite antivenoms derive their therapeutic success from the ability of their antibodies to neutralize venom toxins across multiple snake species. This ability results from a production process involving immunization of large...
mammals with a broad suite of toxins present in venoms. As a result of immunization with this wide range of toxins, many polyvalent antivenoms have a high degree of cross-reactivity to similar toxins in other snake venoms - a cross-reactivity which cannot easily be deconvoluted. As a proof of concept, we aimed at exploring the opposite scenario by performing a high-throughput evaluation of the extent of cross-reactivity of a polyclonal mixture of antibodies that was raised against only a single snake venom fraction. For this purpose, a venom fraction containing short neurotoxin 1 (SN-1; Uniprot accession number P01416, three-finger toxin (3FTx) family), which is the medically most important toxin from the notorious black mamba (Dendroaspis polylepis), was employed. Following immunization of a rabbit, a specific polyclonal antibody response was confirmed by ELISA and immunodiffusion. Subsequently, these antibodies were investigated by high-density peptide microarray to reveal linear elements of recognized epitopes across 742 3FTxs and 10 dendrotoxins. This exploratory study demonstrates in a single immunized animal that cross-reactivity between toxins of high similarity may be difficult to obtain when immunizing with a single 3FTx containing venom fraction. Additionally, this study explored the influence of employing different lengths of peptides in high-density peptide microarray experiments for identification of toxin epitopes. Using 8-mer, 12-mer, and 15-mer peptides, a single linear epitope element was identified in SN-1 with high precision.

**General information**

State: Published
Organisations: Network Engineering of Eukaryotic Cell factories, Department of Bio and Health Informatics, Genomic Epidemiology, Immunoinformatics and Machine Learning, Department of Biotechnology and Biomedicine, Universidad de Costa Rica
Authors: Engmark, M. (Intern), Jespersen, M. C. (Intern), Lomonte, B. (Ekstern), Lund, O. (Intern), Laustsen, A. H. (Intern)
Pages: 151-158
Publication date: 2017
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Toxicon
Volume: 138
ISSN (Print): 0041-0101
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.33 SJR 0.752 SNIP 1.056
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 0.903 SNIP 1.05 CiteScore 2.47
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 0.972 SNIP 1.104 CiteScore 2.48
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.008 SNIP 1.244 CiteScore 2.9
ISI indexed (2013): ISI indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.022 SNIP 1.352 CiteScore 2.85
ISI indexed (2012): ISI indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 0.899 SNIP 1.06 CiteScore 2.54
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 0.866 SNIP 1.138
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 0.745 SNIP 0.979
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 0.887 SNIP 1.068
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 0.821 SNIP 1.106
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-1-immune cattle and hence are prime-candidate antigens for the generation of a subunit vaccine.
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  
Web of Science (2017): Indexed Yes  
BFI (2016): BFI-level 1  
Scopus rating (2016): CiteScore 3.74 SJR 1.888 SNIP 0.937  
Web of Science (2016): Indexed yes  
BFI (2015): BFI-level 1  
Scopus rating (2015): SJR 2.079 SNIP 0.975 CiteScore 3.83  
Web of Science (2015): Indexed yes  
BFI (2014): BFI-level 1  
Scopus rating (2014): SJR 2.029 SNIP 1.05 CiteScore 3.61  
Web of Science (2014): Indexed yes  
BFI (2013): BFI-level 1  
Scopus rating (2013): SJR 2.038 SNIP 1.083 CiteScore 3.97  
ISI indexed (2013): ISI indexed yes  
Web of Science (2013): Indexed yes  
BFI (2012): BFI-level 1  
Scopus rating (2012): SJR 1.902 SNIP 1.047 CiteScore 3.94  
ISI indexed (2012): ISI indexed yes  
Web of Science (2012): Indexed yes  
BFI (2011): BFI-level 1  
Scopus rating (2011): SJR 1.841 SNIP 0.993 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 0.897  
Web of Science (2010): Indexed yes  
BFI (2009): BFI-level 2  
Scopus rating (2009): SJR 0.122 SNIP 0.928  
Web of Science (2009): Indexed yes  
BFI (2008): BFI-level 1  
Scopus rating (2008): SJR 0.111 SNIP 0.921  
Web of Science (2008): Indexed yes  
Scopus rating (2007): SJR 0.122 SNIP 0.964  
Web of Science (2007): Indexed yes  
Scopus rating (2006): SJR 0.122 SNIP 0.887  
Web of Science (2006): 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
Web of Science (2017): Indexed Yes
Scopus rating (2016): CiteScore 3.6 SJR 1.699 SNIP 1.018
Web of Science (2016): Indexed yes
Scopus rating (2015): SJR 1.834 SNIP 1.059 CiteScore 3.74
Web of Science (2015): Indexed yes
Scopus rating (2014): SJR 1.904 SNIP 1.089 CiteScore 3.8
Scopus rating (2013): SJR 1.621 SNIP 0.985 CiteScore 3.41
Scopus rating (2012): SJR 1.138 SNIP 0.689 CiteScore 2.67
Scopus rating (2011): SJR 0.72 SNIP 0.442 CiteScore 1.63
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
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 4.34 SJR 1.705 SNIP 1.002
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 1.934 SNIP 1.092 CiteScore 4.45
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 1.945 SNIP 1.185 CiteScore 4.64
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 2.002 SNIP 1.256 CiteScore 5.16
ISI indexed (2013): ISI indexed yes
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)
Number of pages: 2
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
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.
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.
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.
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.
Positive diversifying selection is a pervasive adaptive force throughout the Drosophila radiation
The growing genomic information on non-model organisms eases exploring the evolutionary history of biodiversity. This is particularly true for Drosophila flies, in which the number of sequenced species doubled recently. Because of its outstanding diversity of species, Drosophila has become one of the most important systems to study adaptive radiation. In this study, we performed a genome-wide analysis of positive diversifying selection on more than 2000 single-copy orthologous groups in 25 species using a recent method of increased accuracy for detecting positive diversifying selection. Adopting this novel approach enabled us to find a consistent selection signal throughout the genus Drosophila, and a total of 1342 single-copy orthologous groups were identified with a putative signal of positive diversifying selection, corresponding to 1.9% of all loci. Specifically, in lineages leading to D. grimshawi, a strong putative signal of positive diversifying selection was found related to cell, morphological, neuronal, and sensorial development and function. A recurrent signal of positive diversifying selection was found on genes related to aging and lifespan, suggesting that selection had shaped lifespan diversity in Drosophila, including extreme longevity. Our study, one of the largest and most comprehensive ones on genome-wide positive diversifying selection to date, shows that positive diversifying selection has promoted species-specific differentiation among evolutionary lineages throughout the Drosophila radiation. Acting on the same biological processes via different routes, positive diversifying selection has promoted diversity of functions and adaptive divergence.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, University of Innsbruck
Authors: Cicconardi, F. (Ekstern), Marcatili, P. (Intern), Arthofer, W. (Ekstern), Schlick-Steiner, B. C. (Ekstern), Steiner, F. M. (Ekstern)
Pages: 230-243
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Molecular Phylogenetics and Evolution
Volume: 112
ISSN (Print): 1055-7903
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 4.34 SJR 2.194 SNIP 2.12
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 2.267 SNIP 1.759 CiteScore 3.85
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 2.331 SNIP 1.929 CiteScore 3.99
BFI (2013): BFI-level 2
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 hydrolases 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.
PROVIDE a project aiming at protein valorization through informatics, hydrolysis, and separation

General information
State: Published
Organisations: National Food Institute, Research Group for Gut Microbiology and Immunology, Research Group for Bioactives – Analysis and Application, Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Technical University of Denmark
Authors: Hansen, E. B. (Intern), Jacobsen, C. (Intern), Lund, O. (Ekstern), Marcatili, P. (Intern), García Moreno, P. J. (Intern)
Number of pages: 1
Publication date: 2017

Host publication information
Title of host publication: Book of Abstracts Sustain 2017
Place of publication: Kgs. Lyngby, Denmark
Publisher: Technical University of Denmark (DTU)
Article number: Sustain Abstract F-5
Main Research Area: Technical/natural sciences
Conference: Sustain 2017, Kgs. Lyngby, Denmark, 06/12/2017 - 06/12/2017
Electronic versions:
The CGE Tool Box
As whole genome sequence data of microorganisms are becoming easily accessible and cheap to produce, a transformation of the traditional methods used for typing, phenotyping and phylogenetic analysis of microorganisms is on the way. Following the anticipation that most clinical microbiological and food safety laboratories will soon have a sequencer in use on a daily basis, there is a growing need for easy-to-use bioinformatics methods that can quickly convert the sequence data into useful information on, e.g., the type of bacteria, whether it is resistant towards any types of antibiotics, and whether it is part of an outbreak. The Center for Genomic Epidemiology, which is located at the Technical University of Denmark, has since its beginning in 2010 developed such bioinformatics methods and made them freely available as web-services. These web-services and their use is the focus of this chapter.

General information
State: Published
Organisations: Department of Systems Biology, Department of Bio and Health Informatics, Center for Biological Sequence Analysis, Genomic Epidemiology, National Food Institute, Research Group for Genomic Epidemiology, Immunoinformatics and Machine Learning, Metagenomics, Statens Seruminstitute, Osaka University
Pages: 65-90
Publication date: 2017

Host publication information
Title of host publication: Applied Genomics of Foodborne Pathogens
Place of publication: Switzerland
Publisher: Springer
Chapter: 5
Main Research Area: Technical/natural sciences
Life Sciences, Food Microbiology, Food Science, Bioinformatics, Microbial Genetics and Genomics, Applied Microbiology, Whole genome sequencing, Web-services
DOIs: 10.1007/978-3-319-43751-4_5
Source: FindIt
Source-ID: 2372561475
Publication: Research - peer-review › Book chapter – Annual report year: 2017

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
Biotechnological Trends in Spider and Scorpion Antivenom Development

Spiders and scorpions are notorious for their fearful dispositions and their ability to inject venom into prey and predators, causing symptoms such as necrosis, paralysis, and excruciating pain. Information on venom composition and the toxins present in these species is growing due to an interest in using bioactive toxins from spiders and scorpions for drug discovery purposes and for solving crystal structures of membrane-embedded receptors. Additionally, the identification and isolation of a myriad of spider and scorpion toxins has allowed research within next generation antivenoms to progress at an increasingly faster pace. In this review, the current knowledge of spider and scorpion venoms is presented, followed by a discussion of all published biotechnological efforts within development of spider and scorpion antitoxins based on small molecules, antibodies and fragments thereof, and next generation immunization strategies. The increasing number of discovery and development efforts within this field may point towards an upcoming transition from serum-based antivenoms towards therapeutic solutions based on modern biotechnology.
Evaluating prediction strategies for identification of T cell responsive mutation-derived neoepitopes in cancer

Increasing evidences point to an important role of mutation-derived antigens in immune recognition of cancer. Current strategies for prediction of immunogenic neoepitopes results in large personalized peptide libraries, but only a minority (<1%) elicit T cell responses at detectable levels. Neoepitopes are of potential valuable as predictors of response to therapy and targets for personalized immunotherapeutic approached. Consequently, there is an unmet need to understand the rules identifying immunogenic neoepitopes. Both tumor mutation mapping via exome sequencing and mass-spectrometry-based elution for MHC class I presented peptides has been applied in different studies, combined with RNA sequencing to determine the expression level of relevant transcripts. Additionally, neoepitopes may be defined based on either autologeous tumor cell lines or snapfrozen tumor material. We present here a study in which all the above mentioned strategies are assessed in three melanoma patients. Predicted large peptide libraries matching the HLA expression of the patients was identified and selected based on any of the strategies given above. This resulted in a total of ~3000 peptides for the three patients. We investigated the T cell recognition of these personalized peptide libraries using a new technology based on DNA-barcode labeled MHC multimers to detect multiple, potentially >1000, different neoepitope specific T cell populations in a single sample. Through this unbiased comparison, we evaluate selection strategies for prediction of immunogenic cancer-associated neoepitopes, and identify rules for precise prediction. Precise prediction is essential for future application of neoepitopes both as predictors of responses to therapy and immunotherapeutic targets.

General information
State: Published
Organisations: National Veterinary Institute, Virology, Department of Bio and Health Informatics, Cancer Genomics, Immunoinformatics and Machine Learning, T-cells & Cancer, Philochem AG, Technical University of Denmark, University Hospital Herlev
Pages: 862-862
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.

Identification of common bacterial antigenic markers from bovine digital dermatitis lesions using meta-transcriptomics in combination with high-density peptide-microarrays

Bovine digital dermatitis (DD) is the most important infectious cause of lameness in dairy cattle, and a major contributing factor to welfare problems and economic losses in the dairy cattle industry worldwide. DD is a disease that involves chronic dermal inflammatory processes and destruction of collagenous and connective tissues. Multiple Treponema species, many of which are not-yet-cultivable, are strongly implicated in disease progression. Despite the economic and welfare importance of this disease, no effective vaccine is available; and there is presently very little knowledge concerning efficacious immunoprophylactic antigens against DD.

It is highly likely that DD-associated treponemes possess considerable antigenic variation, as cows exhibit a variable humoral response against different isolates of Treponema. Hence, combinations of antigens from multiple Treponema species should be used for the development of disease prevention measures. As treponemes from DD lesions are
extremely difficult to culture, identification of these antigens is challenging. To circumvent this problem, we studied the in situ gene expression patterns of the microbiome in DD-affected skin lesions and the host antibody response directed at the site of infection. By metatranscriptomics we measured the in situ genome-wide transcriptome of the bacterial population in DD-affected skin lesions from 21 dairy cows. From the transcriptome data, we identified a panel of Treponema genes that were highly expressed in multiple animals, and we monitored the host immune response to these target genes using high-density peptide microarrays. By this approach, we identified a small group of antigenic proteins, which were expressed in the majority of the samples, and demonstrated antigenicity when screened against sera from infected animal. Future studies will show if these proteins represent candidates for the development of novel biomarkers or vaccines.

VAR2CSA Epitope Identification – Finding All the Needles in All the Haystacks

VAR2CSA Epitope Identification – Finding All the Needles in All the Haystacks

VAR2CSA Epitope Identification – Finding All the Needles in All the Haystacks

VAR2CSA Epitope Identification – Finding All the Needles in All the Haystacks

VAR2CSA Epitope Identification – Finding All the Needles in All the Haystacks