Computational selection of novel antigenic targets in the *Mycobacterium bovis* proteome.

Damien Farrell[1], Kerri M. Malone[1], Kevin Rue-Albrecht[1], Anthony J. Chubb[2] and Stephen V. Gordon[1,2]
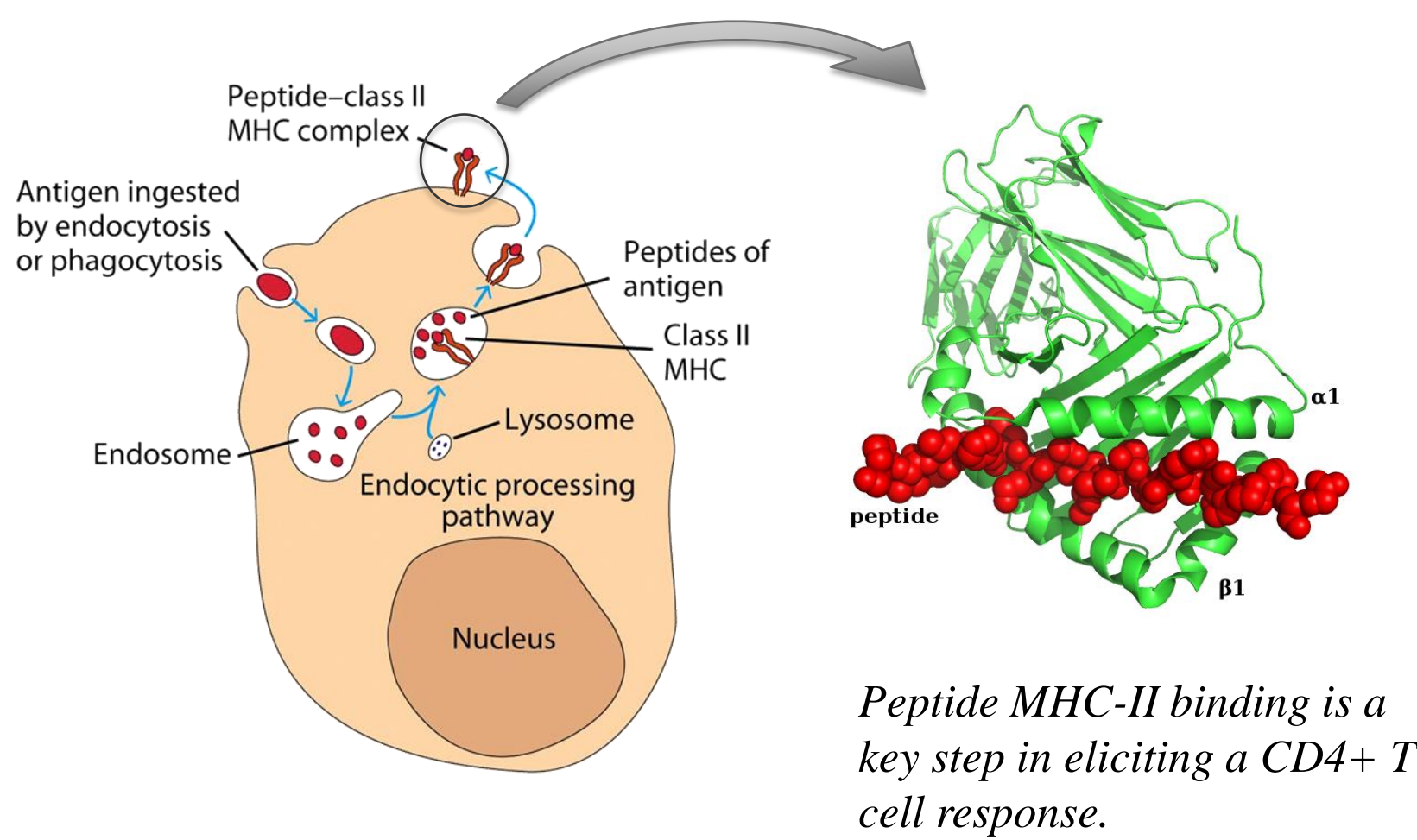
[1]School of Veterinary Medicine, University College Dublin

[2]Conway Institute of Biomolecular & Biomedical Research, University College Dublin   contact: **damien.farrell@ucd.ie**
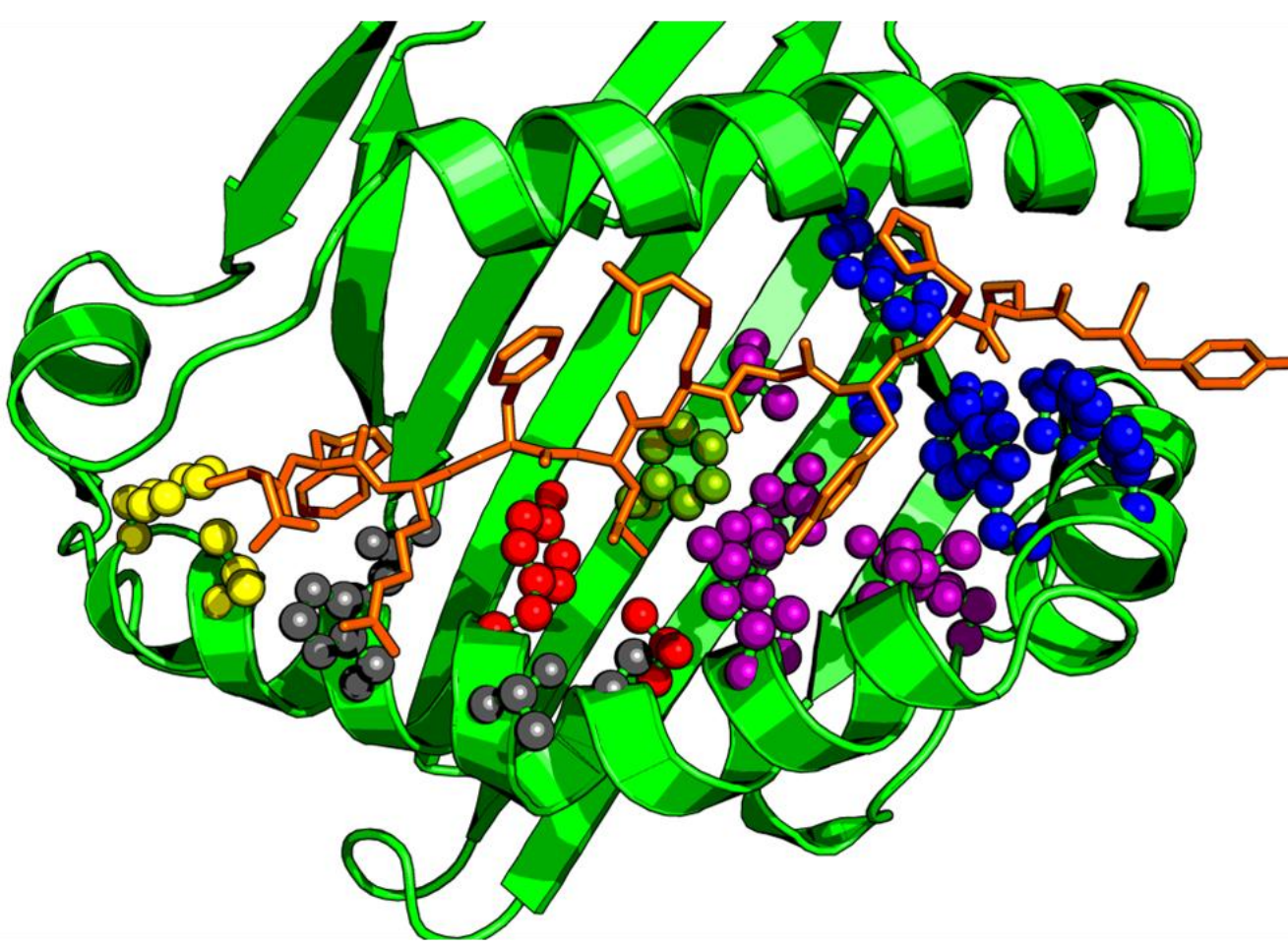
1. Introduction

The discovery of novel antigens is an essential requirement in devising new diagnostics for use in both *M. tuberculosis* (Mtb) and *M. bovis* control programmes. Reverse vaccinology is now a feasible method of extracting potential immunogenic epitopes from bacterial genomes to reduce the cost of experimental screening of antigens for anamnestic responses in infected hosts. Since a significant focus has been on the role of CD4+ T cells, the ability to predict peptide binding to MHC-II molecules is seen as a key step in discovery.

Previous antigen-mining experiments for identification of novel diagnostic or vaccine candidates for human and bovine TB follow a targeted approach, where specific groups of proteins suspected to contain likely candidates are identified and evaluated for immunogenicity. A disadvantage of those approaches is that they are restricted to a relatively small set of proteins biased by the initial selection criteria. Our objective was to computationally select antigens in a less biased manner.

Peptide MHC-II binding is a key step in eliciting a CD4+ T cell response.

2. T cell epitope prediction

There are multiple factors that promote recruitment of T cells but most are not well understood enough for prediction. Therefore prediction methods for MHC-II epitope prediction have thus far concentrated largely on the MHC-peptide binding step. This may be accomplished by structure based modelling using crystal structures of known MHC-II complexes or with sequence based methods that rely on training with large amounts of binding affinity data covering multiple alleles. The second methods are far more practical and successful. High false positive rates makes prediction of T cell epitopes challenging however. Additionally, most studies are on human MHC alleles, we have ascertained those most similar to bovine alleles for our purposes.
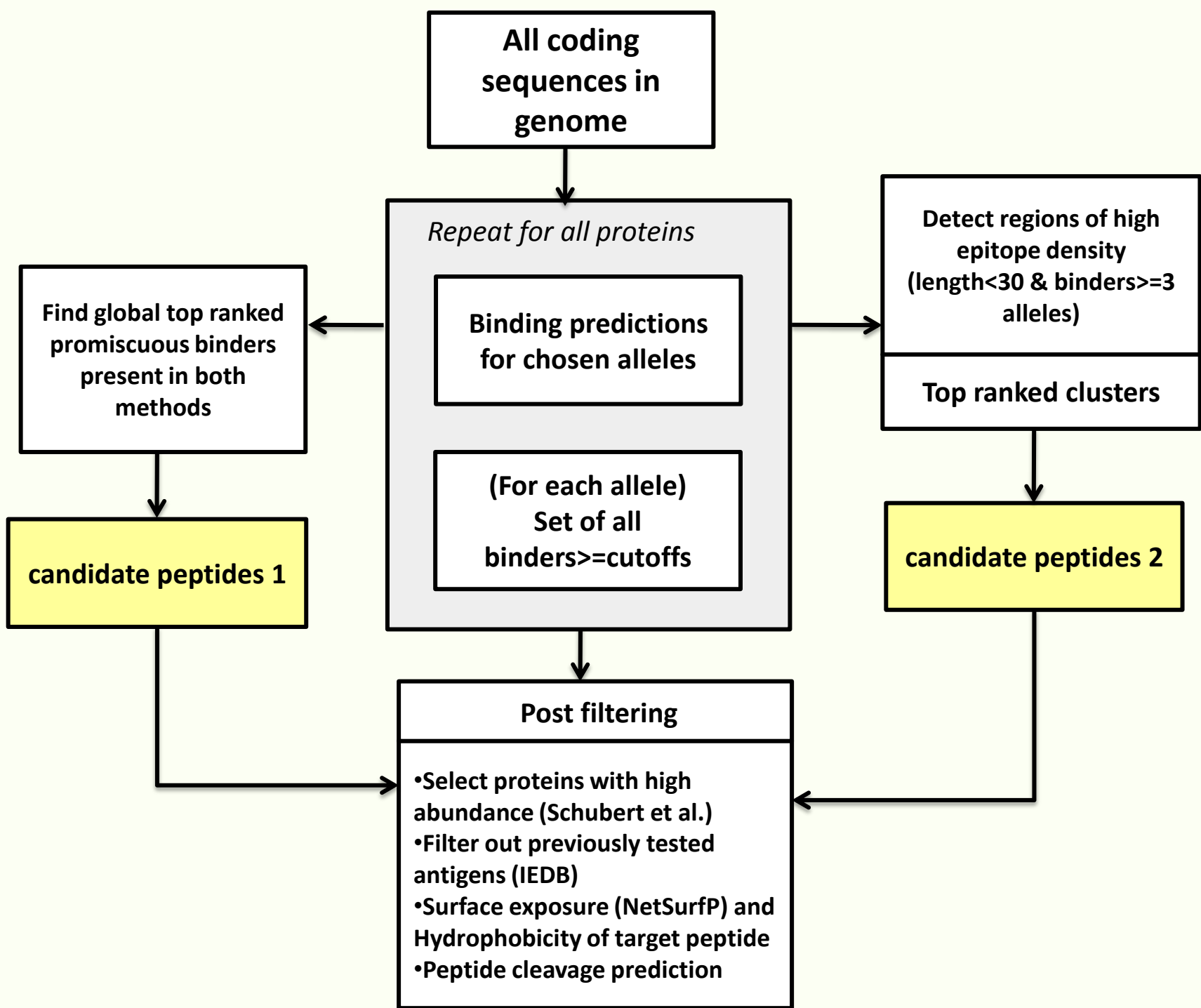
MHC binding pockets are used to model allelic differences so that predictions can be extrapolated to many alleles by simply 'mixing and matching' pocket profiles of known alleles.

3. Methods

We apply two prediction methods, TEPITOPE and NetMHCIIPan, to the *M. bovis* proteome using a subset of 8 known human DRB alleles to approximate the bovine DRB3 immune response. Two different strategies are then applied to filter the resulting set of binders: 1) **shared global binder ranking** and 2) **detecting areas of high epitope density**. This gives a specific 20-mer peptide region inside each protein predicted to be immunogenic thus in theory avoiding the costly synthesis of surplus peptides. The peptide list is then refined based on several metrics such as hydrophobicity, potential for cleavage and amino acid content (i.e. removal of peptides with repetitive sequences). Two additional criteria are finally applied based on the source protein:

1. All proteins > 400 amino acids in length are excluded
2. We select only those proteins detected in Mtb by the recent proteogenomic study of Schubert et al.
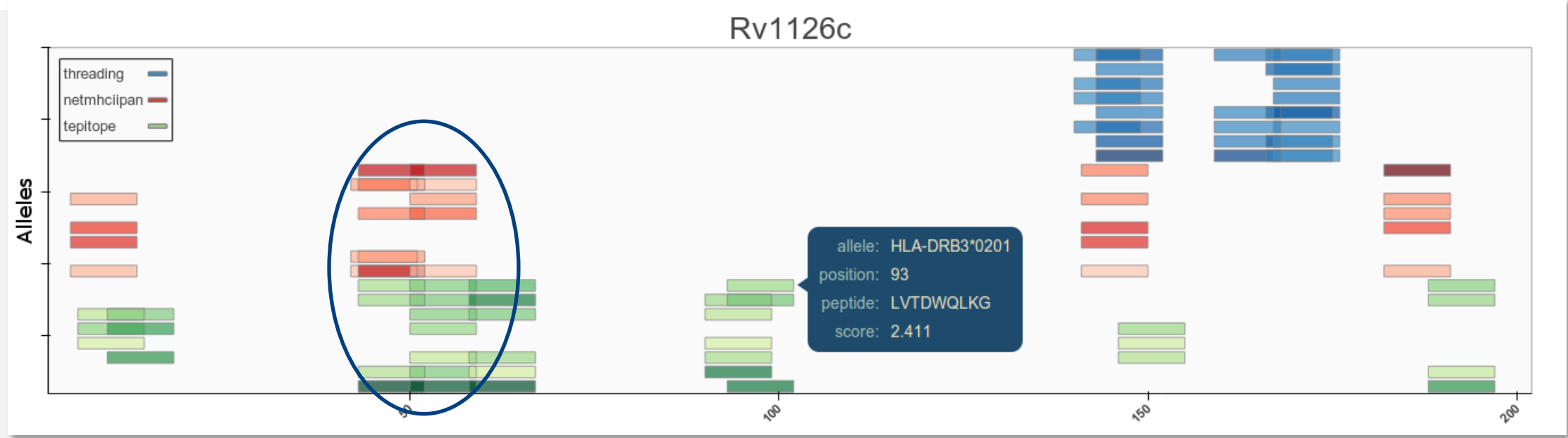
Our *M. bovis* predictions will shortly be tested by synthesising these peptides (along with a positive control) and measuring their ability to induce a interferon-γ response from infected cattle.

Methodology used for selecting proteins with possible antigenicity from a genome.
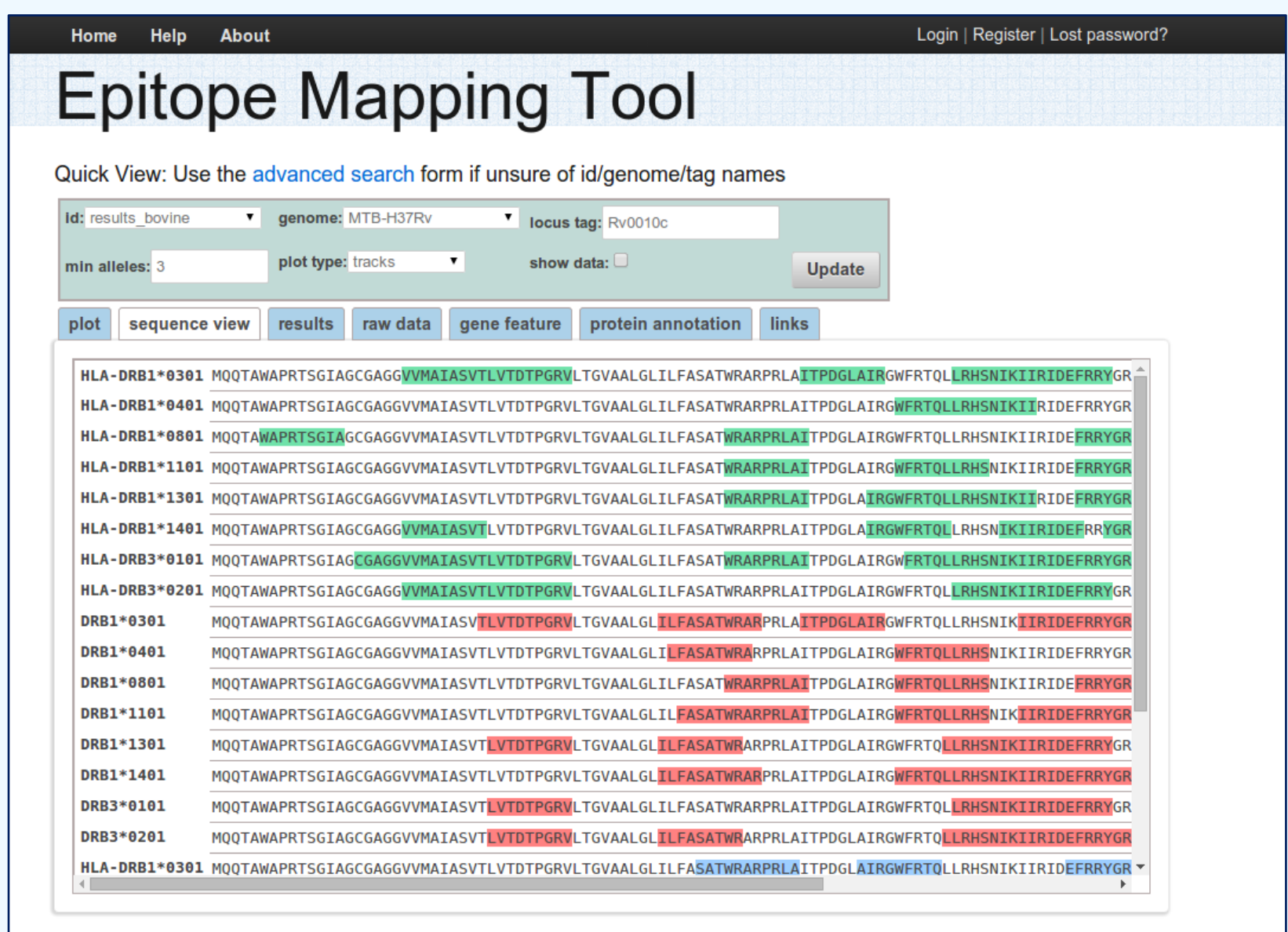
4. Results

After filtering a noticeable amino acid content bias was evident in some candidate peptides detected from regions of epitope density. These are likely due to false positives being picked up by the predictor. TEPITOPE shows a bias for hydrophobic residues in certain binding pockets. Cutoffs for the various filters can be varied depending on how many peptides are required. We rank our peptides using a score based on the measured Mtb absolute abundances, mean predictor score and the number of predicted binders in the 20-mer sequence to be synthesized.

*Example of predicted epitope regions along the sequence of **Rv1126c/Mb1157c**, a conserved membrane protein. Our approach detects areas with multiple predicted epitopes and then filters the resulting peptide regions. Each rectangle represents the predicted score for a 9-mer region of the sequence for a single allele. The colors represent three different prediction methods.*

Our method is largely unbiased with respect to functional category as would be expected. However we do filter out large (>400 aa) proteins to narrow the search space.

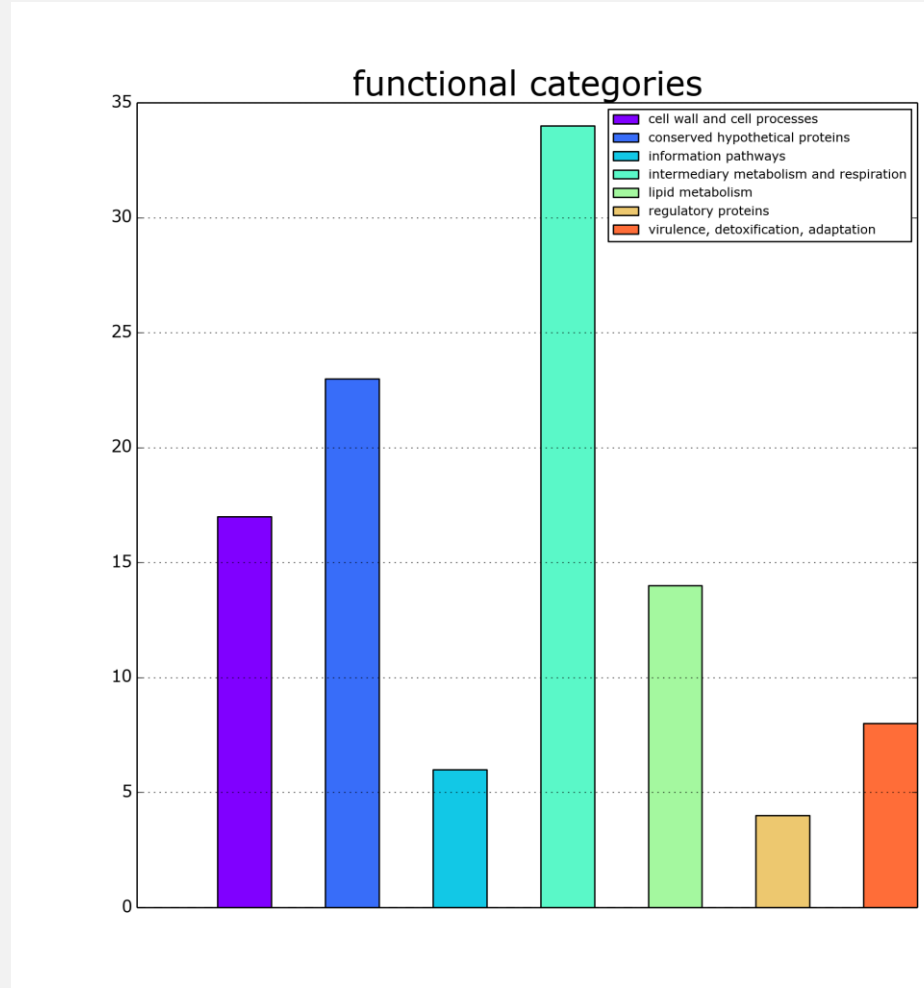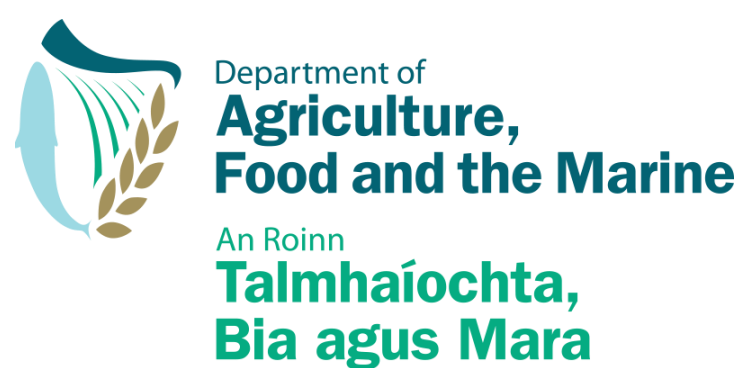*Distribution of functional categories for proteins with top hits.*

5. A web application for Epitope mapping in a genomic context

Most current T cell epitope prediction tools provide a web page for making individual sequence/protein predictions. However in studies of the immune response to a single organism, the whole genome context is important. We have developed a web tool that integrates gene information with pre-computed epitope predictions and also allows browsing and searching through a chosen genome to visualize predictions for a number of methods with multiple alleles in a compact way.

Example of web app interface.

Conclusions

- We have explored the use of suitable structure-based methods and concluded that they are too time consuming
- Our method is a large filtering step only but helps to address the large number of false positives that impede the search for likely antigens
- There are multiple ways to narrow down the list of candidates, we have selected two simple approaches
- This technique can be applied to the genome of any pathogen
- Peptides are yet to be tested. We expect 5-10% enrichment of responders in the candidates.

Selected References

1. Jones, G. J., Bagaini, F., Hewinson, R. G., & Vordermeier, H. M. (2011). The use of binding-prediction models to identify M. bovis-specific antigenic peptides for screening assays in bovine tuberculosis. Veterinary immunology and immunopathology, 141(3-4), 239–45. doi:10.1016/j.vetimm.2011.03.006
2. Chaves, F. a, Lee, A. H., Nayak, J. L., Richards, K. a, & Sant, A. J. (2012). The utility and limitations of current Web-available algorithms to predict peptides recognized by CD4 T cells in response to pathogen infection. Journal of Immunology (Baltimore, Md. : 1950), 188(9), 4235–48. doi:10.4049/jimmunol.1103640
3. Jørgensen, K. W., Buus, S., & Nielsen, M. (2010). Structural properties of MHC class II ligands, implications for the prediction of MHC class II epitopes. PloS One, 5(12), e15877. doi:10.1371/journal.pone.0015877
4. Schubert, O. T., Mouritsen, J., Ludwig, C., Röst, H. L., Rosenberger, G., Arthur, P. K., ... Aebersold, R. (2013). The Mtb proteome library: a resource of assays to quantify the complete proteome of Mycobacterium tuberculosis. Cell Host & Microbe, 13(5), 602–12. doi:10.1016/j.chom.2013.04.008