Mature Epitope Density – A strategy for target selection based on immunoinformatics and exported prokaryotic proteins

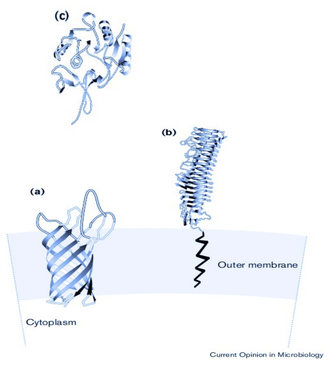
Anderson R. Santos[1,5], Vanessa Bastos Pereira[1], Eudes Barbosa[1,2], Jan Baumbach[2], Josch Pauling[2,4], Richard Röttger[2,4], Meritxell Zurita Turk[1], Artur Silva[3], Anderson Miyoshi[1], Vasco Azevedo[1]§
[1]Molecular and Cellular Genetics Laboratory, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; [2]Computational Biology Research Group, Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej, Odense M, Denmark; [3]DNA Polymorphism Laboratory, Universidade Federal do Pará, Campus do Guamá, Belém, Pará, Brazil; [4]Computational Systems Biology group, Max Planck Institute for Informatics, Campus E2.1, Saarbrücken, Germany; **[5]Faculty of Computing, Universidade Federal de Uberlândia, Campus Santa Mônica, Uberlândia, Minas Gerais, Brazil**

§ Corresponding author: vasco@icb.ufmg.br

ABSTRACT

Background: Current immunological bioinformatic approaches focus on the prediction of allelespecific epitopes capable of triggering immunogenic activity. The prediction of major histocompatibility complex (MHC) class I epitopes is well studied, and various software solutions exist for this purpose. However, currently available tools do not account for the concentration of epitope products in the mature protein product and its relation to the reliability of target selection.

Results: We developed a computational strategy based on measuring the epitope's concentration in the mature protein, called Mature Epitope Density (MED). Our method, though simple, is capable of identifying promising vaccine targets. Our online software implementation provides a computationally light and reliable analysis of bacterial exoproteins and their potential for vaccines or diagnosis projects against pathogenic organisms. We evaluated our computational approach by using the Mycobacterium tuberculosis (Mtb) H37Rv exoproteome as a gold standard model. A literature search was carried out on 60 out of 553 Mtb's predicted exoproteins, looking for previous experimental evidence concerning their possible antigenicity. Half of the 60 proteins were classified as highest scored by the MED statistic, while the other half were classified as lowest scored. Among the lowest scored proteins, ~13% were confirmed as not related to antigenicity or not contributing to the bacterial pathogenicity, and 70% of the highest scored proteins were confirmed as related. There was no experimental evidence of antigenic or pathogenic contributions for three of the highest MED-scored Mtb proteins. Hence, these three proteins could represent novel putative vaccine and drug targets for Mtb. A web version of MED is publicly available online at http://med.mmci.uni-saarland.de/.



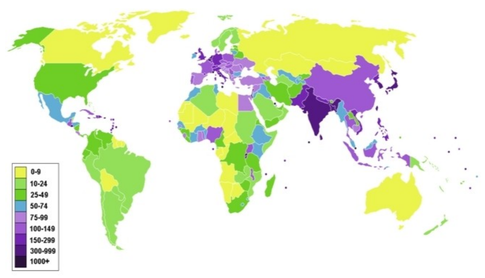
OBJECTIVE

Catalog and characterize bacterial exported proteins according to their possibility to produce cellular immunological responses in hosts.

RESULTS AND DISCUSSIONS

THE MED SCORE CONCEPT IS SIMILAR TO OTHER BROADLY KNEW CONCEPT: THE DEMOGRAPHIC DENSITY ...

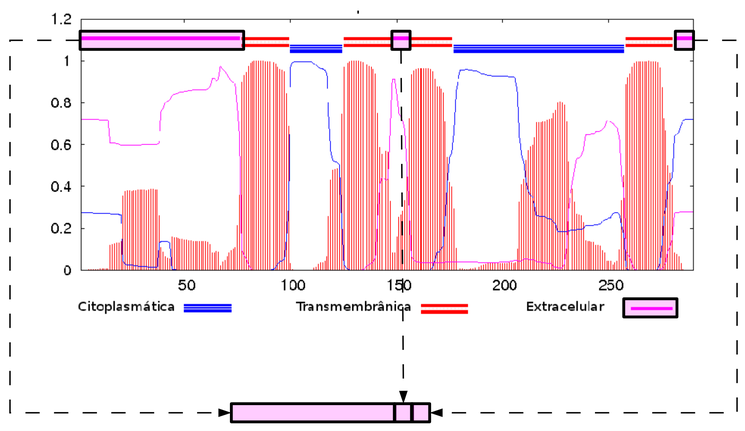
$$Demographic\ Density = \frac{Number\ of\ inhabitants}{Area}$$



Just substitute "inhabitants" for *in silico* predicted epitopes and "Area" for possible 9mer epitopes;

$$MED \simeq \frac{Amount\ of\ predicted\ epitopes * Average\ MHC\ affinity}{Possible\ epitopes\ predicted}$$

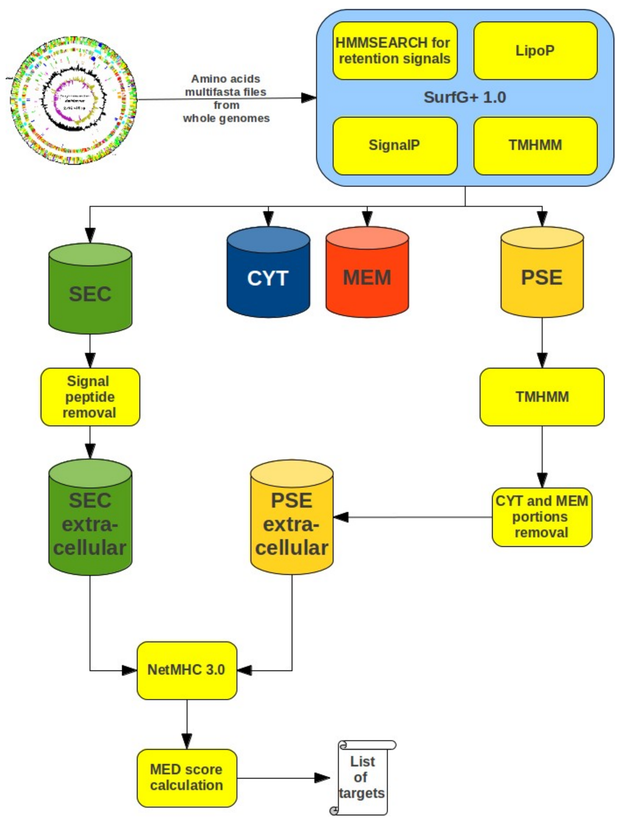
... BUT THE WHOLE AREA DO NOT NEED TO BE CONSIDERED, JUST THE MATURE PORTION OF A PROTEIN.



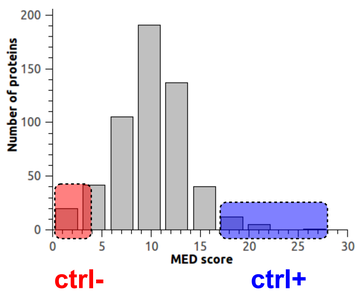
$$MED \simeq \frac{Amount\ of\ predicted\ epitopes * Average\ MHC\ affinity}{Possible\ epitopes\ predicted}$$

The above graph represents the prediction of sub-cellular location from a *C. pseudotuberculosis*' protein portions, strain 1002. For each amino acid a probability of being part of a trans-membrane region is assigned. This graph represents a protein of about 300 amino acids classified as potentially surface exposed (PSE) by the program TMHMM 2.0. Epitopes predicted in the extracellular portion are used to create an artificial protein for the MED score calculations.

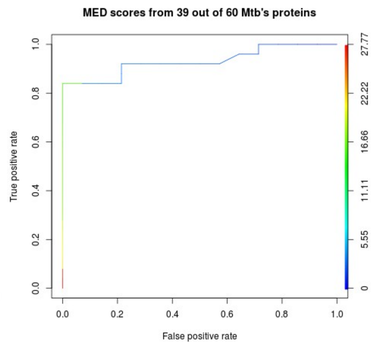
THE MED SCORE CALCULATION WAS AUTOMATED ACCORDING TO THIS MEDPIPE FLOWCHART AND IS AVAILABLE AT THE SITE http://med.mmci.uni-saarland.de/



THE CONCEPT WAS VERIFIED FOR *Mycobacterium tuberculosis* SHOWING A GOOD SENSITIVITY RATE ACCESSED BY LITERATURE SEARCH ...



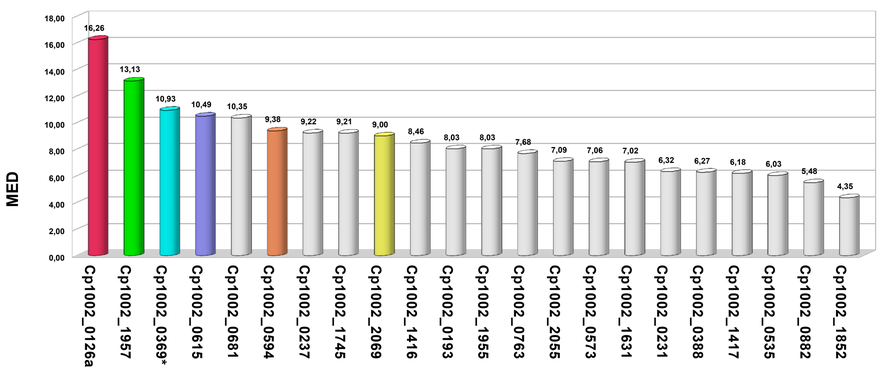
MED scores histogram of 553 *in silico* predicted exoproteins. Sixty proteins were chose for literature search, being half from the left (ctrl-) or lowest scored and other half from right (ctrl+) highest scored.



***ROC* curve for 39 out 60 proteins gathered from the aside histogram. From these 39 proteins, there are 18 and 21 in the ctrl- and ctrl+, respectively, achieving a sensitivity (Tpr) of 84% against a specificity (Fpr) of 7%.**

... AND ALSO FOR *Corynebacterium pseudotuberculosis* WHERE THREE FROM THE TOP FOUR SCORED PROTEINS WERE CONFIRMED AS ANTIGENIC BY DISTINCT EXPERIMENTS.

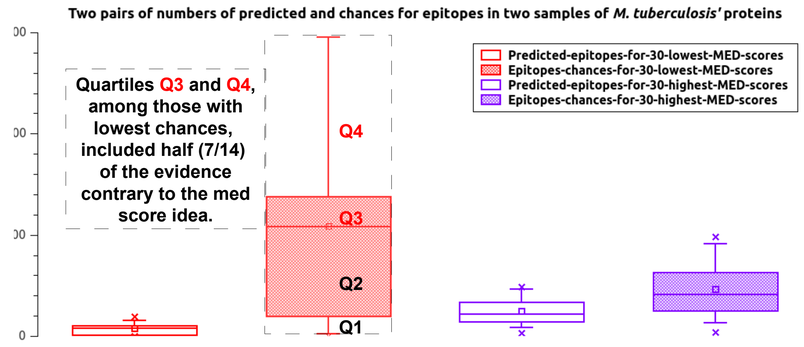
MED scores for the predicted C. pseudotuberculosis pan secretome found simultaneosly in strains 1002 and C231



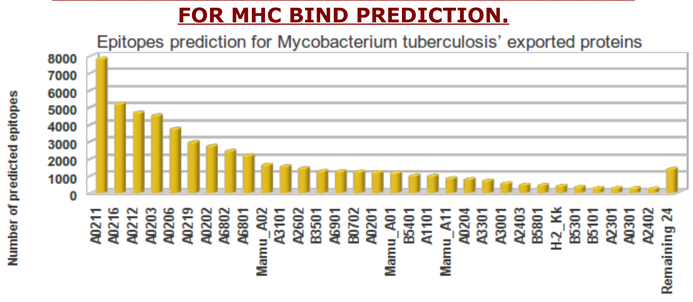
Identifier	*C. pseudotuberculosis* product	MED (nM/mer)	*In silico* and *in vitro* corroborative results
Cp1002_0126a	*Hypothetical protein*	16,26	*Surfg+ ∩ TTP/LC-MS*
Cp1002_1957	*Trehalose corynomycolyl transferase B*	13,13	*Surfg+ ∩ TTP/LC-MS ∩ Antigenicity*
Cp1002_0369*	*Phosphoesterase PA-phosphatase related protein*	10,93	*Surfg+ ∩ TTP/LC-MS ∩ Antigenicity*
Cp1002_0681	*Resuscitation-promoting factor RpfB*	10,35	*Surfg+ ∩ TTP/LC-MS ∩ Immunoproteomic*
Cp1002_0237	*Surface layer protein A*	9,22	*Surfg+ ∩ TTP/LC-MS ∩ Immunoproteomic*
Cp1002_1416	*NlpC/P60 protein*	8,46	*Surfg+ ∩ TTP/LC-MS ∩ Immunoproteomic*

HOWEVER, MED SCORE HAS IT'S LIMITATIONS, FOR INSTANCE, A FOLD GREATHER THAN FOUR BETWEEN PREDICTED EPITOPES AND POSSIBLE 9MER EPITOPES PRODUCES AN UNDER ESTIMATION OF THE MED SCORE.

Two pairs of numbers of predicted and chances for epitopes in two samples of *M. tuberculosis'* proteins



... AND ALSO THE SMALL NUMBER OF MHC CLASS I ALLELES OFFERED BY THE CURRENT COMPUTER PROGRAMS FOR MHC BIND PREDICTION.



MHC alleles in the software NetMHC and the number of predicted strong binders to epitopes from *Mtb* H37Rv exported proteins.

CONCLUSIONS

Conclusions: The software presented here offers a practical and accurate method to identify potential vaccine and diagnosis candidates against pathogenic bacteria by "reading" results from well-established reverse vaccinology software in a novel way, considering the epitope's concentration in the mature portion of the protein.

REFERENCES

[1] Rowland R, McShane H: Tuberculosis vaccines in clinical trials. Expert Rev Vaccines 2011, 10:645-658. [2] Rylance J, Pai M, Lienhardt C et al.: Priorities for tuberculosis research: a systematic review. Lancet Infect Dis 2010, 10:886-892. [3] Thaiss CA, Kaufmann SHE: Toward novel vaccines against tuberculosis: current hopes and obstacles. Yale J Biol Med 2010, 83:209-215. [4] Santos AR, Ali A, Barbosa E et al.: The reverse vaccinology – A contextual overview. IIOABJ 2011, 2:8-15. [5] Gaseitsiwe S, Valentini D, Mahdavifar S et al.: Peptide microarray-based identification of Mycobacterium tuberculosis epitope binding to HLA-DRB1*0101, DRB1*1501, and DRB1*0401. Clin Vaccine Immunol 2010, 17:168-175. [6] Kovjazin R, Volovitz I, Daon Y et al.: Signal peptides and trans-membrane regions are broadly immunogenic and have high CD8+ T cell epitope densities: Implications for vaccine development. Mol Immunol 2011, 48:1009-1018. [7] Barinov A, Loux V, Hammani A et al.: Prediction of surface exposed proteins in Streptococcus pyogenes, with a potential application to other Gram-positive bacteria. Proteomics 2009, 9:61-73. [8] Krogh A, Larsson B, von Heijne G et al.: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001, 305:567-580. [9] Lundegaard C, Lund O, Nielsen M: Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. Bioinformatics 2008, 24:1397-1398. [10] Larsen MV, Lundegaard C, Lamberth K et al.: Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 2007, 8:424. [11] Gupta SK, Smita S, Sarangi AN et al.: In silico CD4+ T-cell epitope prediction and HLA distribution analysis for the potential proteins of Neisseria meningitidis Serogroup B—A clue for vaccine development. Vaccine 2010, 28:7092-7097. [12] Mills AE, Mitchell RD, Lim EK: Corynebacterium pseudotuberculosis is a cause of human necrotising granulomatous lymphadenitis. Pathology 1997, 29:231-233. [13] Liu DTL, Chan W, Fan DSP et al.: An infected hydrogel buckle with Corynebacterium pseudotuberculosis. Br J Ophthalmol 2005, 89:245-246. [14] Baird GJ, Fontaine MC: Corynebacterium pseudotuberculosis and its role in ovine caseous lymphadenitis. J Comp Pathol 2007, 137:179-210. [15] Yeruham I, Elad D, Friedman S et al.: Corynebacterium pseudotuberculosis infection in Israeli dairy cattle. Epidemiol Infect 2003, 131:947-955. [16] Williamson LH: Caseous lymphadenitis in small ruminants. Vet Clin North Am Food Anim Pract 2001, 17:359-71, vii. [17] Aleman M, Spier SJ, Wilson WD et al.: Corynebacterium pseudotuberculosis infection in horses: 538 cases (1982-1993). J Am Vet Med Assoc 1996, 209:804-809. [18] Peel MM, Palmer GG, Stacpoole AM et al.: Human lymphadenitis due to Corynebacterium pseudotuberculosis: report of ten cases from Australia and review. Clin Infect Dis 1997, 24:185-191. [19] Join-Lambert OF, Ouache M, Canioni D et al.: Corynebacterium pseudotuberculosis necrotizing lymphadenitis in a twelve-year-old patient. Pediatr Infect Dis J 2006, 25:848-851. [20] Selim SA: Oedematous skin disease of buffalo in Egypt. J Vet Med B Infect Dis Vet Public Health 2001, 48:241-258. [21] Stavrinides J, McCann HC, Guttman DS: Host-pathogen interplay and the evolution of bacterial effectors. Cell Microbiol 2008, 10:285-292. [22] Sibbald MJJB, van Dij JML: Secretome Mapping in Gram-Positive Pathogens.. In Bacterial secreted protein: secretory mechanisms and role in pathogenesis. Volume . . Edited by Karl Wooldridge. Caister Academic Press; 2009:193-225. [23] Simeone R, Bottai D, Brosch R: ESX/type VII secretion systems and their role in host-pathogen interaction. Curr Opin Microbiol 2009, 12:4-10. [24] Doytchinova IA, Flower DR: Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine 2007, 25:856-866. [25] Tung C, Ziehm M, Kämper A et al.: POPISK: T-cell reactivity prediction using support vector machines and string kernels. BMC Bioinformatics 2011, 12:446. [26] Wang H, Lin Y, Pai T et al.: Prediction of B-cell linear epitopes with a combination of support vector machine classification and amino acid propensity identification. J Biomed Biotechnol 2011, 2011:432830. [27] Lundegaard C, Lamberth K, Harndahl M et al.: NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res 2008, 36:W509-12. [28] Hacker J: Pathogenomics: Insights into Tuberculosis and Related Mycobacterial Diseases. Wiley VCH Verlag GmbH & Co. KGaA Weinheim ; 2006. [29] Hacker J: Pathogenomics: Insights into Tuberculosis and Related Mycobacterial Diseases. In Pathogenomics: genome analysis of pathogenic microbes. Volume 2. 2007. Edited by Dobrindt U, Gobel W. Wiley VCH Verlag GmbH & Co. KGaA Weinheim ; 2006:616. [30] UniProt C: Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 2012, 40:D71-5. [31] Ansari HR, Flower DR, Raghava GPS: AntigenDB: an immunoinformatics database of pathogen antigens. Nucleic Acids Res 2010, 38:D847-53.

Supported by: CNPq