SciForum
Mol2Net

Title of the paper

Gilberto Yong[1,*], Viridiana Camacho Sierra [2] , Esvieta Tenorio Borroto [3]
[1] Universidad Autónoma del Estado de México, Esvieta@gmail.com
[2] Universidad Autónoma del Estado de México,
* Author to whom correspondence should be addressed; E-Mail: author1@email;
Tel.: +1-111-111-111 (ext. 123); Fax: +1-111-111-112.

Received: / Accepted: / Published:

Chagas disease is one of the most important in America transmitted by Trypanosoma cruzi diseases with approximately 7 million people at risk, most of them from Latin American. Due the non-availability of an ideal drug or treatment, development of an effective, and affordable vaccine could be a solution for control and prevention of this disease. In this study, use an bioinformatic approach to predict possible epitopes of the candidates with help of MHC-II Binding Predictor from IEDB,using the prediction method of recommended in IEDB and the set from Allele Class II from DbMHC and allelefrequencies.net with maximal population coverage, we analyze 10 sequence of surface protein expressed in *Trypanosoma cruzi* in its three different stages present in the human body ,and keep the only ones with allotypes referring to the Latin American. A prediction of 70,000 epitopes per protein was obtained which were classified into three groups according to the shared epitopes, where the cruzipain belongs to a single group as it does not present similar epitopes with the other proteins. The first group contains the proteins Asp-3, Asp-2, Gp85, Gp90, Tc85, Sa85 with 17 shared epitopes and a population coverage of 87.89%. The second group Asp-3, Gp82, Gp83 with 31 shared epitopes and 87.89% population coverage. Because Cruzipain is not sharing any epitope, was selected the largest number of replicates contained in the same protein with a coverage above of 80 %. The selected epitopes are going to be synthesized to evaluate their potential as a possible vaccine against *Trypanosoma cruzi.*

Keywords: Trypanosoma; Epitope Prediction; IEDB; Chagas disease; MHC II

1. Introduction

Trypanosoma cruzi is the causative agent of Chagas disease an endemic pathology in Latin America and a huge public health problem, it is transmitted by insects from the family Triatoma, Panstrongylus y Rhodnius. This protozoan is a hemoflagellate parasite that develops in three different cellular forms: amastigote, trypomastigote and epimastigote. Approx 7 to 8 million peoples are infected and 40 million are in risk of take the Chagas disease (Coura, 2007; Yun et al, 2009; WHO,2014) *T.cruz*i is one of the most successful pathogens due to its capacity for infection, survival and persistence in mammalian hosts. A key step for *T. cruzi* persistence is the stage of invasion express different adhesion molecules on its surface, such as mucins, trans-sialidases and other glycoproteins that allow it to

enter the host cell.

These molecules are expressed in the different cellular forms of the parasite and are essential for the host-parasite interaction. *T.cruzi* have in their different stages changes the molecules in their surface this characteristic allows the parasite infect a differences types of cells and brings protection from the immune system of the host (Buscaglia, Campo et al. 2006). In the first stage *T.cruzi* enter to the host in the form of trypomastigote Metacyclic rising the Ca+ surrounding. In this form penetrate the tissue through_surface glycoproteins with a negative charge (Scharfstein, Schmitz et al. 2000); while the invasion of the host cell is realized the trypomastigote create a vesicle becoming in amastigote form and replicate in the cytoplasm. The cell invasion can be classified in three stages: adhesion and recognize - signaling, and invasion (Málaga and Yoshida 2001).

They are currently known different glycoproteins that are expressed in the surface and have adhesion properties expressed in trypomastigote metacyclic like gp90, gp82, gp30 and gp83/50(Yoshida 2006) this type of proteins represents 1% of genome of *T.cruzi* and can be found in other stages of the life cycle of this parasite, also this protein can be classified like trans-sialidases. Trans-sialidases are expressed by trypomastigote and are anchored by glycosylphosphatidylinositol (GPI) to the parasite plasma membrane.

In the form of epimastigote the gp85 is expressed in the membrane this superfamily gp85/TS, have a subgroup of glycoproteins that have a role in the process of adhesion and invasion (VALENZUELA, SEPÚLVEDA et al.).

Stage	Protein
Trypomastigote	Gp90
	Tc85
	Gp85
	Gp82
	Gp83
Metacyclic trypomastigote	SA85
Amastigote	ASP-2
	ASP-3
	ASP-4

Epitope prediction MHC II

The innate immune system reacts quickly against several compounds supposed to be foreign or very rare in a healthy and uninfected individual is able to very specifically react against proteins and peptides specific for pathogenic cells and foreign organisms.

Epitopes were originally defined as the part of an antigen that defines the binding to an immunoglobulin (Huang and Honda 2006). Antigen is generally a processed part of a protein in complex with an MHC protein.is which part of a protein (peptide) is responsible for an immune response. Thus, often this part is referred to as the epitope and the native protein from which the epitope originated as the antigen.

MHC binding prediction methods are today of a very high quality and can predict MHC binding peptides with high accuracy. This is possible for a large range of MHC alleles and relevant length of binding peptides (Lundegaard, Lund et al. 2012).

There are several programs for the prediction of affinity of the epitopes with the MHC but the Immune Epitope Database is considered the most complete. The Immune Epitope Database (IEDB) incorporates more than 120,000 curated epitopes, most of which are extracted from scientific publications and, in contrast to SYFPEITHI, includes also a lot of data on synthetic peptides (Vita, Overton et al. 2015).

The identification of specific epitopes can define the most important fragments of sequence in a protein to lead new treatments. The accuracy of this tools must do with the increase in data volume in the past years that improve the machine learning methods. To improve predictions in machine learning, multiple predictors can be combined to perform a consensus prediction. The most frequently used consensus methods are CONSENSUS, which is hosted on the IEDB website (Moutaftsi, Peters et al. 2006).

The MHC class II binding groove has special pockets that will fit defined amino acids of the binding peptide, and have a major influence on the binding energy(Bordner 2010). HLA class II ligand prediction is more difficult than class I prediction owing to the unknown position of the binding core within the generally longer peptides. This turns out to be an interesting combinatorial optimization problem: select the minimal set of epitopes maximizing the coverage on the whole world population (represented by its global allele frequencies) (Toussaint, Maman et al. 2011). As epitopes are a true subset of what can bind the MHCs of a given individual, the high degree of polymorphism imposes a big challenge on epitope discovery. Fortunately, there are allele frequency databases in web, like allelefrequencies.net (Gonzalez-Galarza, Christmas et al. 2011).

The knowledge of which strong epitopes a protein contains has further importance when considering the use of proteins and peptides as therapeutic drugs. For MHC class II binding, it is inherently harder to go from peptide binding data to a defined motif of the binding core as this is a continuous stretch of nine amino acid residues placed somewhere in a larger peptide usually in the range of 12–20 residues in length. In human's MHC class II chains are encoded by genes in the HLA-DR, -DQ and -DP loci. Knowledge of the allotypes is thus essential for predicting HLA-presented peptides.

.

.

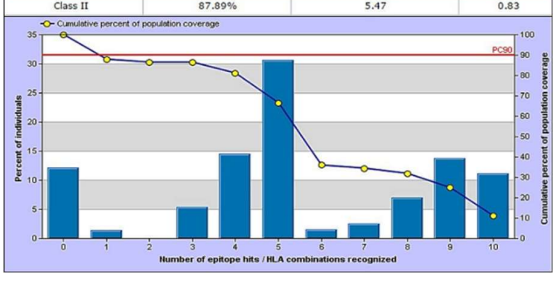
2. Results and Discussion

Results

We obtain 68 epitopes divided in three groups, the first group contains the proteins Asp-3, Asp-2, Gp85, Gp90, Tc85, Sa85 with 17 shared epitopes and a population coverage of 87.89%.

Graph. 1 Group A Protein Asp-3, Asp-2, Gp85, Gp90, Tc85, Sa85

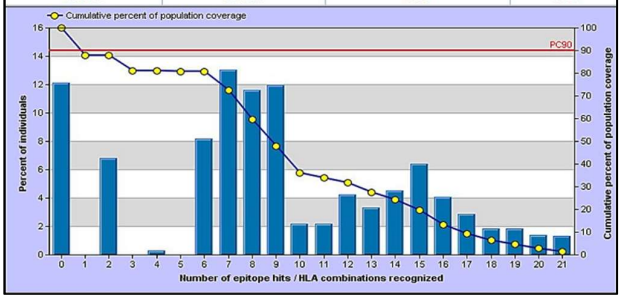
MHC Class	Coverage [a]	Average hit [b]	PC90 [c]
Class II	87.89%	5.47	0.83



The second group Asp-3, Gp82, Gp83 with 31 shared epitopes and 87.89% population coverage.

Graph. 2 Group B Protein Asp-3, Gp82, Gp83

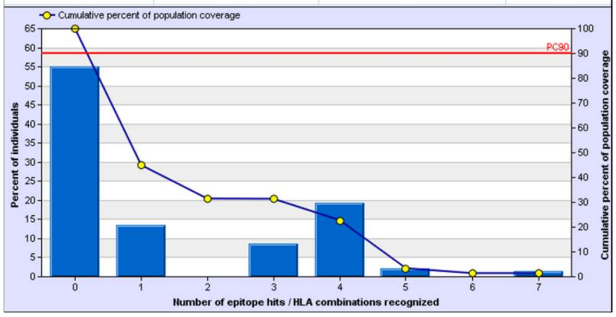
MHC Class	Coverage [a]	Average hit [b]	PC90 [c]
Class II	87.89%	8.90	1.65



The groups obtained and the protein integrate them correspond to the type of proteins in case of

the group 1 all proteins are glycoproteins and the second group all are trans-lidase, this verify the well conserved in the superfamily's and must take in consideration in the future of develop vaccine. Cruzipain is not sharing any epitope, that was reflect in the coverage obtain of 44.91%.

MHC Class	Coverage [a]	Average hit [b]	PC90 [c]
Class II	44.91%	1.36	0.18



3. Materials and Methods

. Understanding the different stages of *T. cruzi* inside the host and the different proteins involved in the process, we chose the proteins according to literature, those has a key function in the process of adhesion and invasion, 10 proteins fit with this requirement to cover a big spectrum of the parasite surface proteins.

Adhesion Molecules in Metacyclic trypomastigote gp90, gp82, gp30 and gp35/50 (Yoshida 2006) and mucins that correspond to 1% of the whole genome of *T. cruzi* and has a function in relation parasite - host (Freitas-Junior, Briones et al. 1998).

Trans - sialidases are hardly related with the invasion process (Mattos, Tonelli et al. 2014) gp82, gp80, gp35/50 y gp85 (Barrias, de Carvalho et al. 2013) use for transportation of sialic acid to parasite mucins, as *T. cruzi* cannot synthesize (Osorio, Ríos et al. 2012). The Superfamily Tc-85 are found in the trypomastigote membrane but is not present in the epimastigote form, are known as Gp85 (Mattos, Tonelli et al. 2014).

Cruzipain or gp57/51 is a lysosomal cysteine protease from *T.cruzi* that can also be found to a lower quantity in the parasitic plasma membrane(Alvarez, Niemirowicz et al. 2012). It is the best characterized cysteine protease and plays a fundamental role in the progression of Chagas disease (Gea, Guinazu et al. 2006). This is expressed mainly in trypomastigote and amastigotes. In trypomastigote is in the flagellar pocket while in amastigotes is located on the cell surface, probably to interact with the host cell cytoplasm (Gea, Guinazu et al. 2006).

IEDB (MHC II Binding Predictor)

A prediction that identifies the share epitopes could help to develop a better treatment and give us information to understand how *T. cruzi* protects against the immune system. In IEDB MHC II we found possible epitopes that can activated a response immune system. The MHC II set was obtained from the allelefrequencies.net with maximus global coverage to the initial process(Greenbaum, Sidney et al. 2011). Every protein was submitted to the same parameters, and filter according the consensus score, in this case we eliminate all epitopes with a score above 25. This guarantees the high affinity to the MHC II. We obtain a matrix with 700000 approx. epitopes for each protein sorted by consensus score.

Share (core) epitopes

Using the previous matrix sorted by consensus score, we search for the core of each epitope per the SMM method. The SMM-align method was shown to outperform other state of the art MHC class II prediction methods. The method predicts quantitative peptide :MHC binding affinity values, making it ideally suited for rational epitope discovery. The method has been trained and evaluated on the, to our knowledge, largest benchmark data set publicly available and covers the nine HLA-DR super types suggested (Nielsen, Lundegaard et al. 2007).

We search the repeats of each core in their respective protein with a length between 8 and 10 amino acids. This method allowed to find the most expressed epitopes in each protein and discard the epitopes with low frequency. We assumed that a core with a high expression ensures a best chance to MHC II to detect this region in the protein and trigger the immune response.

A cluster analysis was performed to group the core of the epitopes of the different proteins to discard those with a low frequency. to make sure this epitope has a significant population coverage in Latin America we use Population Coverage Calculation from IEDB Analysis Resource(Bui, Sidney et al. 2006).

4. Conclusions

The disease of Chagas is one of the main parasitic diseases in Latin America and at present does not have a vaccine.
Thanks to the new amount of data, the methods of machine learning and to databases such as allelefrequencies.net, the predictions are each more reliable with the capacity to cover a larger population and reduce the time in the development of new vaccines. The results obtained show that it is possible to find epitopes that fulfill these qualities, despite the different stages of T. cruzi. Subsequently the epitopes will be synthesized to evaluate their tripanolititic activity against the disease of Chagas.

Conflicts of Interest

The authors declare no conflict of interest

References and Notes

1. Alvarez, V. E., G. T. Niemirowicz and J. J. Cazzulo (2012). "The peptidases of Trypanosoma cruzi: digestive enzymes, virulence factors, and mediators of autophagy and programmed cell death." Biochim Biophys Acta 1824(1): 195-206.
2. Barrias, E. S., T. M. U. de Carvalho and W. De Souza (2013). "Trypanosoma cruzi: Entry into Mammalian Host Cells and Parasitophorous Vacuole Formation." Frontiers in Immunology 4: 186.
3. Bordner, A. J. (2010). "Towards universal structure-based prediction of class II MHC epitopes for diverse allotypes." PLoS One 5(12): e14383.
4. Bui, H. H., J. Sidney, K. Dinh, S. Southwood, M. J. Newman and A. Sette (2006). "Predicting population coverage of T-cell epitope-based diagnostics and vaccines." BMC Bioinformatics 7: 153.
5. Buscaglia, C. A., V. A. Campo, A. C. C. Frasch and J. M. Di Noia (2006). "Trypanosoma cruzi surface mucins: host-dependent coat diversity." Nat Rev Micro 4(3): 229-236.
6. Freitas-Junior, L. H., M. R. Briones and S. Schenkman (1998). "Two distinct groups of mucin-like genes are differentially expressed in the developmental stages of Trypanosoma cruzi." Mol Biochem Parasitol 93(1): 101-114.
7. Gea, S., N. Guinazu, A. Pellegrini, E. Carrera Silva, L. Giordanengo, R. Cano and M. Aoki (2006). "Cruzipain a major Trypanosoma cruzi cystein protease in host-parasite interplay." Inmunología España 25: 225-238.
8. Gonzalez-Galarza, F. F., S. Christmas, D. Middleton and A. R. Jones (2011). "Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations." Nucleic Acids Res 39(Database issue): D913-919.
9. Greenbaum, J., J. Sidney, J. Chung, C. Brander, B. Peters and A. Sette (2011). "Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes." Immunogenetics 63(6): 325-335.
10. Huang, J. and W. Honda (2006). "CED: a conformational epitope database." BMC Immunol 7: 7.
11. Lundegaard, C., O. Lund and M. Nielsen (2012). "Predictions versus high-throughput experiments in T-cell epitope discovery: competition or synergy?" Expert Rev Vaccines 11(1): 43-54.
12. Málaga, S. and N. Yoshida (2001). "Targeted Reduction in Expression ofTrypanosoma cruzi Surface Glycoprotein gp90 Increases Parasite Infectivity." Infection and Immunity 69(1): 353-359.
13. Mattos, E. C., R. R. Tonelli, W. Colli and M. J. Alves (2014). "The Gp85 surface glycoproteins from Trypanosoma cruzi." Subcell Biochem 74: 151-180.
14. Moutaftsi, M., B. Peters, V. Pasquetto, D. C. Tscharke, J. Sidney, H.-H. Bui, H. Grey and A. Sette (2006). "A consensus epitope prediction approach identifies the breadth of murine TCD8+-cell responses to vaccinia virus." Nat Biotech 24(7): 817-819.
15. Nielsen, M., C. Lundegaard and O. Lund (2007). "Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method." BMC Bioinformatics 8: 238-238.

16. Osorio, L., I. Ríos, B. Gutiérrez and J. González (2012). "Virulence factors of Trypanosoma cruzi: who is who?" Microbes and Infection 14(15): 1390-1402.

17. Scharfstein, J., V. Schmitz, V. Morandi, M. M. Capella, A. P. Lima, A. Morrot, L. Juliano and W. Muller-Esterl (2000). "Host cell invasion by Trypanosoma cruzi is potentiated by activation of bradykinin B(2) receptors." J Exp Med 192(9): 1289-1300.

18. Toussaint, N. C., Y. Maman, O. Kohlbacher and Y. Louzoun (2011). "Universal peptide vaccines - optimal peptide vaccine design based on viral sequence conservation." Vaccine 29(47): 8745-8753.

19. VALENZUELA, L., S. SEPÚLVEDA and G. CABRERA "Moléculas de adhesión involucradas en el proceso de invasión de Trypanosoma cruzi a células hospederas." Ibero-Latinoamericana de Parasitología: 105.

20. Vita, R., J. A. Overton, J. A. Greenbaum, J. Ponomarenko, J. D. Clark, J. R. Cantrell, D. K. Wheeler, J. L. Gabbard, D. Hix, A. Sette and B. Peters (2015). "The immune epitope database (IEDB) 3.0." Nucleic Acids Res 43(Database issue): D405-412.

21. Yoshida, N. (2006). "Molecular basis of mammalian cell invasion by Trypanosoma cruzi." An Acad Bras Cienc 78(1): 87-111.