In silico characterization of non-synonymous substitutions and immune recognition of SARS-CoV-2: potential vaccine target for development in the Americas

Andrés F. Cuspoca-Orduz[1], Laura L. Díaz-Lache[1], Álvaro F. Acosta-Costilla[1], Yardany R. Méndez-Fandiño[1]
Escuela de Medicina, Universidad Pedagógica y Tecnológica de Colombia (UPTC), Tunja – Colombia

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) is an emerging pathogen that appeared towards the end of 2019 in Wuhan, China.

It mainly affects the respiratory tract and is transmitted through respiratory droplets, aerosols, and direct contact with the mucosa [1].

In March 2020, the WHO declared SARS-CoV-2 as the etiological agent of the current pandemics. To date, millions of infections and more than 1 million deaths have been reported [2].

Whole-genome sequencing repositories of the SARS-CoV-2 available online allow us to identify frequent non-synonymous mutations (NSMs) and infer changes in pathogenicity and immunological recognition.

OBJECTIVES

General:

- To propose "in silico" multi-epitope vaccine candidate for SARS-CoV-2 considering the NSMs of circulating viruses in Latinoamerica (LATAM) and the world.

Specifics:

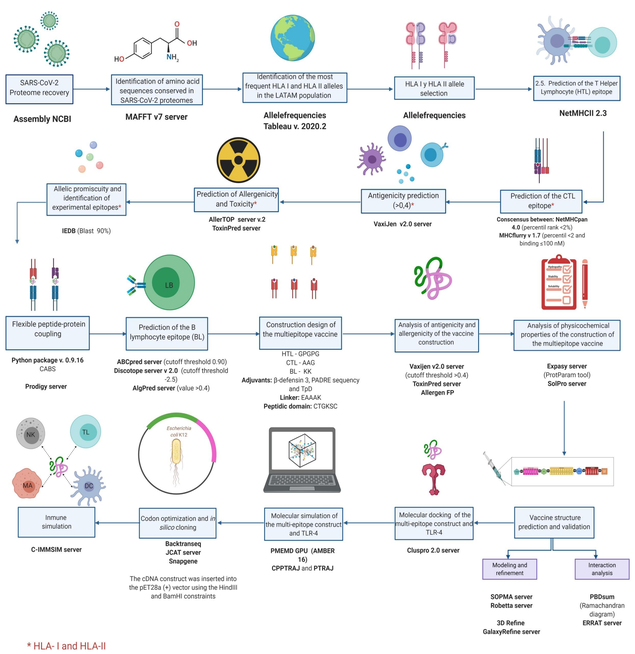
- To identify most frequent NSMs retrieved from whole-genome sequencing in the world through the CNCB and GISAID repositories.
- To identify potential promiscuous vaccine peptides (PPVPs), with a prediction of strong agglutination towards Human Leukocyte Antigen (HLA) class I and II molecules more frequent in LATAM.
- To identify conformational and linear epitopes of B Lymphocyte (BL) present on aminoacid structures of Spike protein (SP) and nucleoprotein (NP).
- To identify immunogenic amino acid sequences as possible adjuvants of immune response that apply in immunogenic, thermostable, and safe construct.

Through the repositories:

GISAID Initiative — Sharing All Influenza Data

China National Center for Bioinformation — 2019 Novel Coronavirus Resource (2019nCoVR)

The most frequently NSMs were in the search potential epitopes through 92 SARS-CoV-2 proteomes recovered from NCBI from December 2019 to March 2020

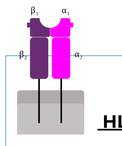


* HLA- I and HLA-II

Results

We used a reverse engineering approach to design a multi-epitope vaccine candidate, based on the identification of PPVPs characterized by:

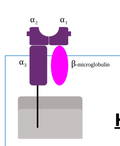
- Antigenic
- Non toxic
- Non allergenic
- Immunogenic



HLA-II

199 peptides came, with strong binding capacity (percentile rank <2.0% NetMHCPan v4.0). Which belong to:

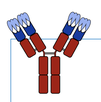
SP (n=114)
NP (n=62)
MG (n=23)



HLA-I

944 peptide with strong binding capacity. Which come from the following proteins:

ORF 1 (n = 573)
SP (n = 163)
MG (n = 51)
ORF 3a (n=43)
NP (n=43)



BL

Crystal structure of NP and SP

- Discotop v2.0: 3D structure-based prediction by consensus.
- ABCPred: by neural networks for linear epitope selection criteria.

In the following table, we exposed the selected PPVPs. In the last column was described previous experimental evidence found out in Immune Epitope Database and Analysis Resource (IEBD).

Character	Number of Peptide (P)	Peptide	Position	Protein of Origin	IEBD
HLA-II epitopes	1	AAAYYVGYLQPRTFL	262-276	SP	In validated region
	2	DDSEPVLKGVKLHYT	1259-1273	SP	Validated peptide. trial in B cell and HLA.
	3	LVIGAVILRGHLRIA	138-152	MG	
	4	QSIIAYTMSLGAENS	690-704	SP	In validated region
	5	QSLLIVNNATNVVIK	115-129	SP	In validated region
	6	SFRLFARTRSMWSFN	99-113	MG	
	7	VLSFELLHAPATVCG	512-526	SP	In validated region
HLA-I epitopes	8	CISTKHFYW	3147-3155	ORF1-NSP4	
	9	FAMQMAYRF	898-906	SP	In validated region
	10	FLLNKEMYL	3183-3191	ORF1-NSP4	Validated peptide. trial in T cell.
	11	GYKSVNITF	835-843	ORF1-NSP3	
	12	ITLCFTLKR	110-118	ORF7a	
	13	KRAKVTSAM	4022-4030	ORF1-NSP8	
	14	KVKYLYFIK	4225-4233	ORF1-NSP9	Validated peptide. trial in T cell.
	15	LEMELTPVV	1012-1020	ORF1-NSP3	
	16	MPYFFTLLL	2169-2177	ORF1-NSP3	
	17	VMYASAVVL	3684-3692	ORF1-NSP6	
	18	WTAGAAAYY	258-266	SP	In validated region

Table 1. PPVPs predicted across most frequent HLA-I and HLA-II alleles in LATAM.

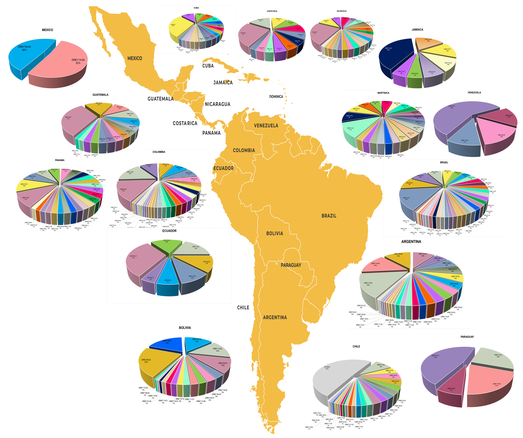


Figure 1: HLA molecules class I and II over median by locus from 18 LATAM countries

The most frequent NSMs reported up to October 28, 2020, in the CNCB and GISAID repositories were taken into account for the elaboration of the multi-epitope construct to maintain a viable and actual proposal.

TOP	WORLD (GISAID)	WORLD (CNCB)	EE.UU (Wang, et al)	URUGUAY(Elizondo et al.)
1	D614G, SP	A222V, SP	P323L, NSP12(RdRp)	D614G, SP
2	P323L, NSP12	L18I, SP	D614G, SP	P323L, NSP12
3	R203K, NP	D614G, SP	Q57H, ORF3a	L84S, ORF8
4	G204R, NP	S477T, SP	T85I, NSP2	R203K, NP
5	Q57H, ORF 3a	Q57H, ORF3a	L84S, ORF8	G204R, NP
6	T85I, NSP2	I300V, ORF1ab	Y541C, NSP13(Helicase)	L37F, NSP12
7	A222V, SP	T265I, ORF1ab	P504L, NSP13(Helicase)	A152S, NP
8	V30L, ORF10	P4715H, ORF1ab	S24L, ORF8	G15S, NSP5

Table 2. Non-synonymous mutations [28-30] .

PPVPs were incorporated into the multi-epitope construct adjuvants and spacers, are presented in Figure 2:

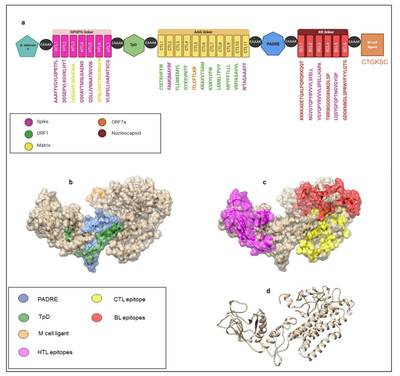
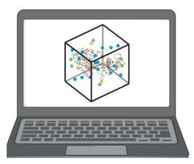
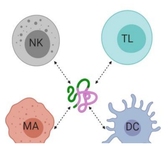


Figure 2. Multi-epitope construct

To generate a safe construct, we evaluate physicochemical properties through the characterization of amino acids, proving to be satisfactory.



Molecular dynamics simulation of the construct with TLR-4/ myeloid differentiation factor 2 (MD2): Start of stabilization at 20 ns and a duration of 68 ns. Stabilization started at a stable conformation from beta-defense 3 to Pan DR T helper epitope (TpD) of the construction of the vaccine.



Immune simulation: We observed the general activation of CD4 + and CD8 + T cells, immunity based on HTL-1 cells also predominated. Stimulation of various immunoglobulins (IgG1 + IgG2, IgM and IgG + IgM) was observed after the first injection with a robust response to the third injection.

Discussion

Our vaccine approach differs from other proposed "*in silico*" for SARS-CoV-2 due to :

1. We use conserved peptides from all proteins of SARS-CoV-2, which were antigenic, non-allergenic, non-toxic, immunogenic without NSMs, instead of complete proteins because it gives greater specific recognition and a lower likelihood of adverse reaction [3] Table 1.
2. Our approach blocks, at least in a theoretical way, with different phases of the viral cycle and not only the entry process [3][4].
3. Linear antibodies towards SP are included in the N terminal and conform to a disorder domain, and do not have N-glycosylation.
4. In case of a natural selection pressure within the amino acid sequences in pandemic ongoing, antibodies from NP can be useful, specifically in a possible mutation process in RBD from SP, acting in severity of coronavirus disease 2019 (COVID-19).
5. We analyzed strong binding prediction towards HLA molecules class I and II over median by locus from 18 LATAM countries (Figure 1).

Peptides with experimental evidence were P1, P2, P4, P5, P7, P9, P10, P14, P18. While P3, P6, P8, P11-P13, 915-P17 do not have any validation yet (Table 1).

ADJUVANTS

PADRE
Improves immune response in vaccines against human papillomavirus and hepatitis B virus [5–6].

β- defensin 3
It induces the maturation of Langerhans cells to dendritic cells, stimulates their proliferation and production of IFN-γ by TL-CD4 +[7–9].

TpD
Trigger increased TLC and robust generation of antibodies [10].

M cell ligant
CTGKSC peptide domain
Increases theoretical multi-epitope solubility because it has amphiphilic properties [11], delivering the construct to TLR-4/ MD2 in M cells by polymer nanoparticles [11-13]. With this delivery option, it would be possible to reach all remote populations, where access to medical care is arduous.

MULTI-EPITOPE VACCINES

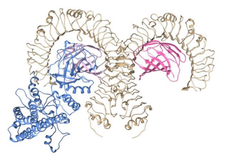
In other infectious and non-infectious agents:

- Multi-epitope peptide vaccine proposals obtained with a similar approach: DEN-2 [14], Hendra [15], Nipah [16], *Pseudomonas aeruginosa* [17], *Klebsiella pneumoniae* [18], and *Plasmodium spp* [19], and non-infectious Kaposi's sarcoma [20].
- *In vivo* models have tested specific and safe multi-epitope vaccines for influenza virus type A [21], Ebola virus[22], HPV-16 [23], and *uropathogenic Escherichia coli (E. coli)* [24].

In SARS-CoV-2:

- In SARS-CoV-2, some predictions over immunogenic peptides have been validated from *"in silico"* studies with methodology similar to ours. In phase 1 of the study (NCT04546841) immunogenic peptides, were predicted from structural, non-structural, and accessories proteins. In TL-CD4 and TL-CD8 from convalescent patients, immunogenic peptides interacted and activated TL producing IFN-γ [25].

Interaction between the multi-epitope vaccine candidate and TLR-4 was stable with probable non-canonical interaction.
In nature, other non-canonical interactions have been described, as in the case of the Human Immunodeficiency Virus (HIV) protein *Tat* [27]



Conclusions

- Based on the NSMs and our bioinformatics pipeline, we propose a vaccine with a global reach that meets the requirements for in vivo and in vitro experimental trials.
- It was possible to identify NSMs from SARS-CoV-2 in Latam and worldwide. Any of the proposed epitopes in showed frequent mutations on the sequences.
- A rational and optimized multi-epitope construct was proposed based on the prediction of strong epitope binding to the most frequent HLA I and II in LATAM.
- The linear and conformational epitopes were established in BLs, on amino acid structures of Spike protein (SP) and nucleoprotein (NP) of SARS-CoV-2.
- It was shown that our vaccine construct was antigenic, non-allergenic, and non-toxic in the in silico approach as well as its safety and the evaluation of the construct through molecular dynamics showed its safety and stability.

Acknowledgment

The authors would like to express their gratitude for the comments made by Dra Nury Olaya Galan and Bioinformatics support to Dr Juvenal Yosa Reyes, Director of Molecular Simulation and Bioinformatics Laboratory at Simon Bolivar University - Barranquilla - Colombia

References