In silico EsxG·EsxH Rational Epitope Selection: Candidate Epitopes for Vaccine Design against Pulmonary Tuberculosis

Constanza Estefania Martinez-Olivares[1], Rogelio Hernández-Pando[1]*, Edgar Mixcoha[2]*

[1]Experimental Pathology Laboratory, Instituto Nacional de Ciencias Médicas y Nutrición "Salvador Zubirán", Experimental Pathology Department, Mexico City, Mexico

[2]Addiction Immunology Laboratory, Instituto Nacional de Psiquiatría "Ramón de la Fuente Muñiz" - CONACyT, Mexico City, Mexico

*** Corresponding author**

E-mail: emixcohahe@conacyt.mx (EM)

E-mail: rhdezpando@hotmail.com (RHP)

Keywords: pulmonary tuberculosis; EsxG·EsxH; vaccine; humoral response; Molecular Dynamics simulations; epitope

Abstract

Rational design of new vaccines against pulmonary tuberculosis is imperative. EsxG·EsxH complex is involved in metal uptake, drug resistance, and immune response evasion. These characteristics make it an ideal target for rational vaccine development. The aim of this study is to show the rational design of epitope-based peptide vaccines by using bioinformatics and structural vaccinology tools. A total of 4.15 μs of Molecular Dynamics simulations were carried out to describe the behavior in solution of heterodimer, single epitopes, and epitopes loaded into MHC-II complexes. In order to predict T and B cell epitopes for antigenic activation, bioinformatic tools were used. Hence, we propose three epitopes with the potential to design pulmonary tuberculosis vaccines. The possible use of the proposed epitopes includes subunit vaccines, as a booster in BCG vaccination to improve its immune response, as well as the generation of antibodies that interfere with the mycobacterium tuberculosis homeostasis, affecting its survival.

Introduction

Tuberculosis (TB) is the leading bacterial cause of death in the world. Mycobacterium tuberculosis (Mtb), the pathogen responsible for the infection, spreads by air route into droplets or aerosol sprays from sick people, affecting mainly the lungs (pulmonary TB). With the COVID-19 pandemic, the number of TB deaths increased by 5.6%, reversing the existing progress in reducing mortality (1). Mycobacterium bovis bacille Calmette-Guérin (BCG) is

the only available vaccine against TB. BCG vaccine was developed 100 years ago, it shows efficient protection against miliary and meningeal TB, but its protection against pulmonary TB ranges from 0 to 80 % (2).

Vaccination is one of the most successful public health initiatives to prevent infectious diseases (3). For TB disease, the development of new vaccines is necessary to reduce its worldwide incidence; hence, a rational design of vaccines is imperative. Rational vaccine design makes use of bioinformatics tools and computational chemistry approaches to identify protein regions based on physicochemical and structural information to trigger a protective immune response. The main advantage of these tools is the reduction of development in time and cost (3,4).

One of the vaccine approaches is the use of epitope vaccines (5). Epitope-based vaccines are considered safe because they exclude allergens, toxins, or functional domains of the pathogen. Therefore, the involved immune response is restricted to defined antigenic regions (3). A disadvantage of subunit vaccines constituted by epitopes is the duration of the evoked immune memory due to their small size, short half-life, limited activity on immune cells stimulation, and their limited immune memory. Although BCG is an attenuated vaccine, it shows a relatively limited immune memory, but it is possible to extend it by using subunit vaccines as boosters after BCG vaccination. Vaccination is also necessary to stimulate crucial immune targets from innate immunity, as well as acquired or adaptive cellular and humoral immunity. BCG is efficient to stimulate innate immunity, being macrophage activation and training immunity a clear example (6). BCG stimulates adaptive cellular immunity, but it seems that it induces a limited activation of humoral immunity. Antibodies also contribute to

47 the protection against Mtb, being essentially mediated by bacterial opsonization, and
48 improving its intracellular killing (7). B lymphocytes can produce bactericidal antibodies but
49 it has not been demonstrated for TB yet. However, it is quite possible that antibodies could
50 neutralize or interfere with significant bacterial proteins related to essential metabolic
51 processes that could be bactericidal antibodies.
52
53 The selection of an optimal antigen is a critical step in rational vaccine design (3).
54 EsxG and EsxH proteins from Mtb were selected for this work. Mainly, EsxG·EsxH complex
55 belongs to the ESX-3 secretion system, which is essential for bacteria viability, and it is
56 metabolically important due to its role in iron uptake and in the adaptation to a low zinc
57 environment (metal homeostasis) (8–11). EsxG and EsxH proteins are secreted by the ESX-
58 3 secretory system as a heterodimer. EsxG·EsxH heterodimer not only has significant activity
59 in the bacterial metabolism, but it also has a remarkable immunological activity response to
60 TB (12,13). Phagolysosome formation is a basic mechanism for bacterial elimination in
61 phagocytic cells and is efficiently blocked by Mtb, after interacting with the host endosomal
62 sorting complex (ESCRT) that prevents lysosome fusion. The ESCRT complex classifies the
63 ubiquitin-labeled surface receptors to be degraded in the lysosome and loads peptides derived
64 from the pathogens into MHC class II molecules. The EsxG·EsxH complex inhibits
65 phagosome maturation by binding Hrs protein and delaying T cell activation during infection.
66 Therefore, it is important to identify mechanisms that promote a rapid and robust activation
67 of protective T cells (14–17). Furthermore, the genes that encode EsxG and EsxH proteins
68 have been associated with Mtb first-line drug resistance mechanisms (18). These
69 characteristics make EsxG and EsxH proteins an ideal target for rational vaccine
70 development.

71 The structure of the EsxG·EsxH complex has been solved by Nuclear Magnetic
72 Resonance. The heterodimer folds into a helix-turn-helix-turn structure of each monomer,
73 which are arranged antiparallel to each other, forming a complex of four helices according to
74 Protein Data Bank (PDB: 2KG7). These helices are defined as follows. EsxG: residue 16 to
75 40 forms helix g1a, residue 41 to 42 forms helix g1b, and residue 48 to 77 forms helix g2.
76 EsxH: residue 18 to 40 forms helix h1a, residue 41 to 44 forms helix h1b, residues 61 to 72
77 forms helix h2a, and residues 75 to 81 forms helix h2b. The putative Zn^{2+} binding site has
78 been identified in H14, H70, and H76 residues of EsxH protein, and the side chain of E77 is
79 hypothesized to form the fourth Zn^{2+} coordination group (19). The EsxG·EsxH complex
80 model under code 2KG7 lacks Zn^{2+}; there are no other experimental coordinates of dimer
81 that contain Zn^{2+}.
82
83 The aim of this study is to show the rational design of epitope-based peptide vaccines
84 by using bioinformatics and structural vaccinology tools. Three epitopes of the EsxG·EsxH
85 complex were selected. A total of 4.15 μs of Molecular Dynamic simulations (MD) were
86 used to describe the behavior of epitopes in solution or loaded into a MHC-II complex. B cell
87 epitopes were predicted for humoral activation. Even knowing that Mtb is an intracellular
88 pathogen, it must have entered from outside, so if the immune system has developed specific
89 and avid antibodies against any of the TB proteins, it will be advantageous to inhibit the
90 infection. Furthermore, if one of these proteins has demonstrated to be necessary for the
91 survival of the bacteria, such as the EsxG·EsxH complex, then the antibodies generated by
92 active immunization with a rationally designed conjugate vaccine could protect against TB
93 infection. Hence, we propose three promising candidate epitopes for vaccine design against
94 pulmonary TB.

Methods

MHC class II binding epitope prediction

Major histocompatibility complex (MHC) class II binding epitope prediction was carried out by the ProPred (http://crdd.osdd.net/raghava/propred/) (20) and the NetMHCII 2.3 (http://www.cbs.dtu.dk/services/NetMHCII/) (21) servers. In order to locate binding regions with default parameters, 51 and 25 HLA-DR alleles were selected, respectively.

B cell epitope prediction

A Linear B cell epitope prediction from EsxG (UniProt ID: O53692) and EsxH (UniProt ID: P9WNK3) sequences were carried out by BepiPrep 2.0 server (http://www.cbs.dtu.dk/services/BepiPred/) (22), using a threshold of 0.46. The DiscoTope 2.0 server was used for discontinuous B cell epitope prediction (23), with default parameters. The antigenic determinant prediction was performed by the IEDB analysis resource server (http://tools.iedb.org/bcell/) using the Kolaskar and Tongaonkar antigenicity scale (24), with default parameters.

EsxG·EsxH Complex Molecular Dynamics simulation

116 A quaternary structure formed by tuberculosis virulence factors EsxG and EsxH (ID:
117 2KG7) (19) was obtained from RCSB Protein Data Bank (http://www.rcsb.org) (25). For this
118 study, out of thirty conformers fetched from PDB, the first model was selected. Protein
119 coordinates were centered in a dodecahedron box with a 1.5 nm distance between protein and
120 unit cell walls. AMBER99SB was used as a force field to describe the whole system (26).
121 Protein was solvated using the SPC water model (27). The system was neutralized by
122 replacing water molecules until reaching 0.15 M NaCl mimicking physiological conditions.
123 The neutralized system was energetically minimized for 1000 steps using the steepest descent
124 algorithm. The system was equilibrated at 310 K with an NVT assembly for 1 ns.
125 Subsequently, 5 ns of an NPT equilibration at 310 K and 1 atm of pressure was used. At the
126 end of the equilibration stages, Molecular Dynamics simulations were carried out using the
127 GROMACS software package version 5.1 (28,29). The Parrinello-Rahman barostat (30) and
128 the Nosé-Hoover thermostat (31) was used to maintain constant pressure and temperature.
129 Long-range electrostatic interactions were calculated using the particle-mesh Ewald
130 algorithm (32), under periodic boundary conditions, with a 0.15 nm grid. We simulated the
131 system for 500 ns, motion equations were integrated every 2 ps, coordinates and velocities
132 were saved every 10 ps.
133
134 In order to study the monomer's molecular behavior in solution, 250 ns MD
135 simulations of each monomer were calculated at 310 K under the same conditions. Monomers
136 coordinates were taken from 2KG7 model 1.
137

Thermal unfolding simulations

A thermal unfolding pathway was simulated to understand the relationship between the structure and thermal stability of the EsxG·EsxH complex. The system was treated according to Section 2.3. Simulations were conducted for 50 ns at 350, 400, 450, 500, and 550 K, respectively. The initial structure was the same for all simulated systems and corresponds to the structure after the final NVT simulation.

Molecular Docking

MHC class II molecules are crucial for initiating an antigenic-specific immune response to presented antigens (33), as well as for the activation, proliferation, and differentiation of B cells (34). A peptide-protein docking of the selected epitopes was performed to predict their binding conformation into MHC class II molecules.

The crystal structure of the human leukocyte antigen (HLA) class II histocompatibility DRB1 allele (ID: 6BIZ) (35) was obtained from RCSB Protein Data Bank (25). HLA-DRB1 molecule was selected because it binds to most Mtb epitopes and is the allele with the highest superficial expression (36). Predicted epitope sequences and HLA-DRB1 crystal structure were entered in the CABS-Dock server (http://biocomp.chem.uw.edu.pl/CABSdock) (37) for epitope-protein docking, so to obtain coordinates of the complexes. The most populated cluster model for each complex with the smallest RMSD value was selected. In order to improve the quality of interactions and

peptide presentation, a flexible docking was performed in FlexPepDock server (http://flexpepdock.furmanlab.cs.huji.ac.il) (38), using default parameters; the best resultant model was selected to perform MD simulations.

MD simulations: epitopes and HLA-epitopes complexes

Coordinates of the four chosen epitopes (section 3.2) were selected and extracted from the 2KG7 file (model 1). MD simulations of epitopes were carried out along 600 ns for each epitope, under the same conditions described in Section 2.3. MD simulations of HLA-epitopes complexes were carried out along 250 ns under the same conditions. Motion equations were integrated every 2 ps, coordinates and velocities were stored every 10 ps.

Molecular Dynamics analysis

Analysis of unfolding trajectories was done using GROMACS package tools (28,29). The following parameters were analyzed: Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration (RoG), Hydrogen bonds (Hb), secondary structure time evolution, Solvent Accessible Surface Area (SASA), clusterization, distance matrix, and Principal Component Analysis (PCA). Clusterization was performed using the gromos method (39), using RMSD cutoff 0.2 nm over the backbone atoms. Hydrogen bond analysis was calculated when donor and acceptor atoms were at a 0.3 nm distance, forming an angle of $180^{o} \pm 30^{o}$ for at least 10 ps. Alignments for RMSD, RMSF, RoG were performed using backbone atoms, PCA was performed using alpha carbons coordinates of residues.

Images were made in PyMOL (40) molecular visualization software and Gnuplot graph package (41).

Results

MHC class II and B cell epitope prediction

Considering that TB is an intracellular infection in which antigen presentation mediated by dendritic cells and macrophages is quite important to activate cell-mediated immunity, the prediction of MHC class II epitopes is a crucial element. In this regard, ProPred (20) and NetMHCII 2.3 (21) threw seven epitopes from EsxG protein and fourteen epitopes for EsxH protein. Promiscuous regions that could bind to several HLA-DR alleles (S1 Table) were selected to identify residues that were bound to the MHC class II nucleus and nonamers epitopes.

One important aspect for an efficient anti-mycobacterial immune response is the production of antibodies that not only contribute to promoting macrophage opsonization, but they can mediate the neutralization of these proteins and induce substantial metabolic abnormalities during iron and zinc bacterial deprivation. Thus, the prediction of linear B cell epitopes was important. BepiPrep 2.0 prediction (22) resulted in four epitopes for EsxG, with more than five residues in length, and three residues located in positions 67, 68, and 79. For

205 EsxH, three epitopes were predicted with more than sixteen residues in length. As shown in
206 Fig 1A, the epitopes were homogeneously located along the sequence of both proteins.
207
208 Discontinuous B cell epitope prediction in DiscoTope 2.0 (23) yielded five residues
209 for EsxG protein: [M^{1}, S^{2}, L^{4}, F^{97}]; and the following residues for EsxH protein as
210 components of discontinuous B-cell epitopes: [^{1}MSQIMYNYPAM^{11}, H^{14}], [^{44}QGDTG^{48},
211 T^{50}], [^{79}NT^{80},^{83}MMARD^{87}], [E^{90}, A^{92}, ^{94}WGG^{96}] (Fig 1B).
212
213 Kolaskar and Tongaonkar's prediction, a semi-empirical method that uses
214 physicochemical properties of amino acids was used to predict antigenic determinants (24).
215 Protein regions that satisfied the threshold value (antigenic propensity threshold 1.00) were
216 predicted to be potential antigenic sites in which antibodies elicit antigen-specific responses
217 (Fig 1C).
218
219 **Fig 1. Epitope Prediction**. (A). Linear B cell epitope prediction. Prediction in BepiPred 2.0.
220 Residues that present a score higher than 0.46 are indicated with an E at the top of the
221 sequence. Residues with values greater than 0.46 are colored in dark orange. (B).
222 Discontinuous B cell epitope prediction. Prediction in DiscoTope 2.0. Epitopes predicted are
223 shown as spheres. (C). Antigenic determinant prediction. Kolaskar and Tongaonkar´s
224 prediction method was employed; x-axis - residue number; y-axis - antigenic propensity.
225 Residues with antigenic propensity greater than 1.000 were predicted as potential antigenic
226 sites.
227

Epitope selection

We selected three epitopes based on promiscuity to bind MHC class II molecules, antigenic propensity, structural stability, lineal and discontinuous prediction of B cells, and surface exposition. According to this selection criteria, the epitope G1 (^{11}LVASQSAFA^{19}) was selected from EsxG protein. Epitopes H1 (^{5}MYNYPAMLG^{13}) and H2 (^{65}LVRAYHAMSSTHE^{77}) were chosen from EsxH protein. H2 epitope with thirteen residues joint three nonameric epitopes and corresponded to the zinc-binding region (19), Table 1. One epitope from EsxG protein was selected, which did not meet the selection criteria. G2 (^{51}FQAAHARFVAAAAKVN^{66}), it was used to prove that our exclusion criteria was suitable for the selection.

Table 1. Epitopes selected. Amino acid sequence from epitopes selected of protein EsxG and EsxH.

Epitope number	Epitope ID	Epitope sequence	Selected epitope	Short name	Selected epitope sequence
1	ep1	^{11}LVASQSAFA^{19}	ep1	G1	^{11}LVASQSAFA^{19}
2	ep2	^{51}FQAAHARFV^{59}	ep2 - ep3	G2	^{51}FQAAHARFVAAAAKVN^{66}
3	ep3	^{58}FVAAAAKVN^{66}			
4	ep4	^{5}MYNYPAMLG^{13}	ep4	H1	^{5}MYNYPAMLG^{13}
5	ep5	^{65}LVRAYHAMS^{73}	ep5 - ep6 - ep7	H2	^{65}LVRAYHAMSSTHE^{77}
6	ep6	^{66}VRAYHAMSS^{74}			

7	ep7	^{69}YHAMSSTHE^{77}			

EsxG·EsxH Complex Molecular Dynamics Simulation

In order to understand EsxG·EsxH complex behavior in an aqueous solution, a MD simulation for 500 ns was performed. RMSD results showed that abrupt changes occurred from 0 to 9 ns, which might be related to the equilibration phase. The dimer presented two global conformations corresponding to two phases observed in the RMSD. The first one of them from 10 ns to 148 ns, the second one from 149 ns until 500 ns, keeping an average RMSD value of 0.995 nm along the simulation (S1A Fig). The temporary evolution of RoG was evaluated to understand the dimer's compactness degree. The system started with a RoG value of 2.245 nm, after 25 ns it fluctuated between 2.237 nm and 1.835 nm. At 149 ns the molecular compactness increased, it maintained an RMSD average of 1.887 nm (S1B Fig). These fluctuations coincided with the two RMSD stages previously described. Observing the trajectory, the compactness was mainly due to the N-terminal of the EsxG monomer, which adopted a conformation folded toward helix g1a. The folding of the C and N-terminal from EsxH towards helix h2 (Figs 2A-2C) yielded a more packed structure. RMSF analysis was performed to determine fluctuating regions of the protein, being residues 1 to 7 from N-terminal of EsxG monomer the most fluctuating region (1.465 nm), followed by residue 87 (0.718 nm), residue 88 (0.729 nm), residues 95 to 97 from C-terminal of EsxG (0.684 nm), residues 1 to 6 from N-terminal of EsxH (0.544 nm), and residues 85 to 89 from C-terminal of EsxH (0.523 nm) (S1C Fig). The results showed that EsxG monomer showed greater RMSF values than EsxH.

The secondary structure analysis revealed the following structural changes on the alpha-helix of the G monomer: i) residues 3 to 7, which initially formed a coil and bend structures, which structured into alpha-helix, from 378 ns to 374 ns. ii). Residue 18 lost its alpha-helix structure after 10 ns of simulation and was restructured again together with residue 17 from 206 to 382 ns and from 460 ns to 500 ns. iii) Residue 39 lost its alpha-helix structure from 41 to 78 ns and from 379 to 397 ns. iv) The alpha-helix structure elongated from residues 39 to 42 in time 2 to 41 ns. v). The g2 helix lengthens at residues 46 to 49 from 78 ns and was maintained up to 500 ns. vi). Residues 74 to 76 changed from turn to alpha-helix structure during 18 to 61 ns and from 126 to 159 ns. vii) Residues 81 to 84 get an intermittent structure between alpha-helix and turn from 122 ns until the end of the simulation. The alpha-helices of monomer EsxH had a more stable secondary structure than the two described before. In contrast, the following changes were observed in the secondary structure of the EsxH monomer: i) residues 5 to 9 of the EsxH chain started the simulation unstructured. They were structured to an alpha-helix from 58 ns and was maintained up to 500 ns of simulation. ii) There was a shortening of helix h1, being residues 21 to 23 the ones that lost their secondary structure after 103 ns of simulation. iii) Residues 38 to 42 were structured in alpha-helix from 24 ns to 138 ns, then the structure fluctuated between alpha-helix and turn until the end of the simulation. iv) Residue 77 to 81 lost its helix structure between 26 ns to 123 ns, later it recovered its structure. v) Residues 82 to 88 lengthened the helix intermittently during the simulation. Despite these changes, the dimer maintained its secondary structure with minimal fluctuations in both monomers (S1D Fig).

Clusterization showed that 80% of the total visited structures along the simulation were sorted in the first eight clusters, although the first cluster was the most populated with

288 47.9% (S1E Fig). Hydrogen bonds are important to preserve the native secondary and tertiary
289 structure of proteins; 140 intra-protein hydrogen bonds on average were kept through the
290 entire simulation, having a maximum of 169 hydrogen bonds at 320 ns (S1F Fig). A principal
291 component analysis was performed by using Cα atoms to obtain the principal component
292 covariance matrices. Analyzing the first one, it revealed correlated displacements between 1)
293 C and N-terminal from EsxG, 2) N-terminal from EsxG monomer and residues 40 to 55 from
294 EsxG, 3) both terminals from EsxH monomer and N-terminal from EsxG, and 4) the
295 displacements of the four helices of the heterodimer were correlated themselves (Fig 2D). A
296 map contact was performed to evaluate the distance among residues, the folding of the C and
297 N-terminal ends from EsxG and EsxH monomers to their respective helices was shown (Fig
298 2E). The g2 and h2 helices were the most stable motifs in heterodimer.
299
300 **Fig 2. Global structure behavior from EsxG·EsxH dimer at 310 K.** (A). 0 ns of
301 simulation. Arrows indicate regions that underwent folding. (B). 149 ns of simulation. (C).
302 500 ns of simulation. Blue depicts EsxG monomer; orange depicts EsxH monomer. (D). First
303 principal component covariance matrix. The matrix is color-coded, from red (correlated
304 displacements) to blue (non-correlated displacements). The diagonal line stands for the
305 correlation between the residues paired with themselves; the blue stands for the correlation
306 between each residue pair during the 500 ns simulation. (E). Contact map. The matrix is
307 color-coded, blue (farthest) to red (closest). The diagonal line represents the zero distance
308 between the residues paired with themselves, while spots represent the distances (nm) for
309 each residue pair during the 500 ns simulation (V1 Video in supporting information).

Monomers Molecular Dynamic Simulation

In order to study the monomer's molecular behavior in solution, MD simulations of each monomer at 310 K were carried out. The EsxG monomer RMSD value stabilized at 57 ns. After this time, one conformational structure with an average of 1.292 nm was presented (S2A Fig) and an average RoG of 1.459 nm (S2B Fig). An average of 62 intra-protein hydrogen bonds, with a maximum of 79 bonds at 185 ns, and a minimum of 47 bonds at 267 ns were found (S2C Fig). N and C-terminal were the most fluctuating regions, with an RMSF value of 0.823 nm in residue 1 and 1.223 nm in residue 97, respectively (S2D Fig). Both terminals folded toward g2 helix (Figs 3A-3C). Secondary structure analysis revealed that residues 15 to 19 fluctuated between a turn structure and alpha-helix after 37 ns, while residues 41 to 43 after 101 ns form a turn structure, residues 63 to 65 after 49 ns fold into a turn structure. Residues 75 and 76 fluctuated between turn and alpha-helix structure after 18 ns, residues 91 to 95 changed from coil to alpha-helix structure between 58 and 98 ns of simulation (S2E Fig). In this case, 75% of the structures were found in the first five clusters, being the first cluster the most populated one with 42.8% (S2F Fig). The first principal component covariance matrix showed correlation movements between the N and C-terminal. While an anticorrelation between the N and C ends with the g1 and g2 helices was found (Fig 3D); these movements caused the loss of the secondary structure.

On the other hand, the EsxH monomer structure stabilized at 66 ns, with an RMSD average of 0.926 nm and a RoG of 1.795 nm (S3A ang S3B Figs, respectively). An average of 63 intra-protein hydrogen bonds was maintained, with a maximum of 85 bonds at 7 ns and

a minimum of 44 bonds at 160 ns and 269 ns (S3C Fig). Residues 19, 29 to 36, 50 to 70, and the C-terminal presented the greatest fluctuation with a maximum RMSF value of 0.848 nm. The least fluctuating regions were residues 9 to 15 with an RMSF between 0.225 and 0.333 nm, and residues 76 to 88 with 0.121 to 0.203 nm (S3D Fig). Helix h1 had an important loss of secondary structure. Moreover, residues 24 to 27 were maintained throughout the simulation whereas residues 28 to 31 were only maintained until 66 ns, but the alpha-helix structure was later lost and restructured at 218 ns. After 188 ns of simulation, residues 34 to 44 were structured in an alpha-helix, which was maintained at the end of the simulation (S3E Fig). Helix h2 maintained a regular stability of secondary structure (Figs 3E-3G). Particularly, 59% of the structures were found in the first 15 clusters, the two initial clusters presented a larger population of structures (S3F Fig). Anticorrelated motions were observed between residues 29 to 41 and 17 to 27 and residues 27 to 47 and from 47 to 71 (Fig 3H). These results demonstrated that EsxH did not have a stable secondary structure as a monomer, as it was found for EsxG.

Fig 3. Global structure behavior of EsxG at 310 K. (A). at 0 ns of simulation. (B). at 57 ns of simulation. (C). at 250 ns of simulation. (D). First principal component covariance matrix. The matrix is color-coded, from red (correlated displacements) to blue (non-correlated displacements). The diagonal line stands for the correlation between the residues paired with themselves, while the color stands for the correlation between each residue pair during the 250 ns simulation. **Global structure of the first cluster from EsxH at 310 K.** (E). 0 ns of simulation. (F). 66 ns of simulation. (G). 250 ns of simulation. (H). First principal component covariance matrix.

Thermal unfolding

MD simulations supply data related to the unfolding pathway and structural stability at elevated temperatures. Therefore, MD simulations were carried out at different temperatures. Fig 4A shows the initial structure of the dimer. The folding of the N-terminal of EsxG occurred at 5 ns of simulation at 350 K. Residues 20 to 24 from helix h1a evolved to an irregular secondary structure, which remained until 50 ns. At 50 ns of simulation, both ends folded into the EsxG helices. These observations corresponded to the evolution of the RMSD value, which had variations between 0.7 nm and 1.1 nm, from 7 to 30 ns of the simulation. After this time, the structure stabilized and remained close to the mean (0.858 nm) (S4A Fig). The RoG value was close to the mean (1.897 nm) after 10 ns (S4B Fig). The most fluctuating region was the N-terminal of the EsxG monomer with an RMSF value of 0.97 nm, followed by the C-terminal of EsxG and ExH monomer with an RMSF value of 0.84 nm (S4C Fig). The average of the intra-protein hydrogen bonds was 132 (S4D Fig).

The dimer underwent relevant structural changes when applying kinetic energy by increasing temperature. As these changes occurred quickly, just the first 5 ns of the simulations from 400 K to 550 K were considered for analysis.

The average structure at 400 K (Fig 4D), 450 K (Fig 4E), 500 K (Fig 4F), and 550 K (Fig 4G) were calculated. Fig 4G shows a significant loss of native conformation. Regarding RMSD, the average values were 0.887 nm (S5A Fig), 0.810 nm (S5B Fig), 1.292 nm (S5C Fig), and 1.417 nm (S5D Fig), respectively. At 400 K, the radius of gyration was larger (2.142

380 nm) than 310 K and 350 K (S5E Fig) because the N-terminal of the EsxG monomer unfolded,
381 as did residues 76 and 77 (Thr and His) of the EsxH monomer. Despite this, the secondary
382 structure of the helix was maintained during the simulation (S6A Fig). However, the intra-
383 protein Hb decreased to 132 (S6E Fig).
384
385 **Fig 4. Average structures from EsxG·EsxH dimer thermal unfolding simulations.** (A).
386 0 ns of simulation. (B). 5 ns of simulation at 350 K. (C). 50 ns of simulation at 350 K. (D). 5
387 ns of simulation at 400 K. (E). 5 ns of simulation at 450 K. (F). 5 ns of simulation at 500 K.
388 (G). 5 ns of simulation at 550 K.
389
390 At 450 K, the compactness of the dimer was directly proportional to the increase in
391 temperature being RoG average values as follows: 1.962 nm at 450 K (S5F Fig), 1.920 nm
392 at 500 K (S5G Fig), and 1.768 nm at 550 K (S5H Fig). At 450 K, residues 30 to 40 from
393 helix g1 lost their regular secondary structure at 2 ns of simulation. Helix h1 also lost its
394 structure after 2 ns (S6B Fig). Furthermore, helix g1 preserved its structure until 2 ns, but it
395 completely lost its secondary structure. In the opposite way, residues 53 to 60 corresponding
396 to the EsxH monomer kept its structure. Helix g2 has higher stability in terms of secondary
397 structure (S6C Fig).
398
399 At 550 K, residues 20 to 30 of the EsxG monomer were maintained during 1.5 ns,
400 and residues 50 to 62 along 5 ns of simulation. EsxH monomer lost its regular secondary
401 structure, turns and a beta secondary structure predominated (S6D Fig).
402

As for Hydrogen bonds, when the temperature was increased to 450 K the intra-protein Hb average number decreased to 125 (S6F Fig), at 500 K, Hb decreased to 113 (S6G Fig), and at 550 K in the beginning of the simulation decreased to 110 then increased to 145 (S6H Fig). The increase can be explained by interactions generated with the change of motifs in the secondary structure.

MD simulations: epitopes and HLA-epitopes complexes

In order to study the behavior of the epitopes in solution, MD simulations of the selected peptides under physiological conditions were carried out. Furthermore, these epitopes were docked to the MHC-II HLA molecule (ID: 6BIZ) to predict how epitopes are bonded in the HLA to observe the epitope behavior while forming the complex, and to describe the intermolecular interactions formed in the HLA/epitope complexes. Moreover, MD simulations were carried out from the HLA/epitope complexes to compare them with the epitope solution dynamics.

For epitope G1, one global conformation was observed with a RMSD value of 0.310 nm (S7A Fig). The RMSD of the G2 epitope fluctuated throughout the simulation; the epitope got three conformations, the first one from 0 to 192 ns, the second one from 194 ns to 231 ns, and the third one from 232 ns to the end of the simulation (S7B Fig). H1 epitope adopted a global conformation throughout the simulation, which was maintained with a RMSD of 0.095 nm (S7C Fig). For H2 epitope, a global conformation was found with a RMSD of 0.297 nm (S7D Fig). G2 epitope presented several conformations throughout the simulation. When we compared the RMSD of G1, G2, and H2 epitopes loaded in the HLA, similar behavior

426 for the G1 complex was found, which indicated no change in the conformation (S7A Fig). In
427 G2, H1, and H2 complexes, we observed an increase in the RMSD. G2 complex had one
428 conformation until 234 ns. H1 complex showed three conformations from 0 ns to 32 ns, from
429 33 ns to 151 ns, and from 152 ns to 250 ns (S7C Fig). In the H2 complex, two conformations
430 were observed from 0 ns to 40 ns, the second one from 50 ns to 250 ns (S7D Fig).
431
432 The RoG of the epitopes that remained stable in solution throughout the simulation,
433 showed a mean of 0.833 nm (S7E Fig), 0.806 nm (S7F Fig), 0.709 nm (S7G Fig), and 1.065
434 nm (S7H Fig) for epitopes G1, G2, H1, and H2, respectively. When the epitopes were
435 coupled to the HLA molecule, the RoG in the G1 remained without major changes (S7E Fig).
436 The Rog in the G2 kept stable up to 237 ns, after this time, the structure lost compactness
437 (S7F Fig). The H1-HLA complex compacted from 18 ns to 132 ns, then lost its compactness
438 and compacts again at 192 ns (S7G Fig). The RoG decreased during the simulation in the H2
439 indicating compaction of the structure (S7H Fig).
440
441 The four epitopes in the solution did not present a regular secondary structure (S8A-
442 8D Figs). The N and C-terminals of the four epitopes in solution were the regions with the
443 highest RMSF (S9A-9D Figs). In the arrangement of the epitopes to the HLA, epitopes G1,
444 H1, and H2 continued without having a regular secondary structure. However, the coil and
445 bend motifs remained constant throughout the simulation (Figs 5A, 5C, 5D) (S8E, 8G and
446 8H Figs). Interestingly, G2 epitope, when loaded into the HLA, took a regular secondary
447 structure, residues 7 to 11 formed an alpha-helix structure that was maintained during the
448 entire simulation (Fig 5B) (S8F Fig). When the epitopes formed a complex with HLA, we
449 observed a decrease in the RMSF value of the N-terminal from G1 (S9A Fig), and in the C-

450 terminal from G2 (S9B Fig) and H1 (S9C Fig), compared to the solvated epitope. In the H2-
451 HLA complex, the RMSF of the N and C-ends were the most fluctuating regions, the central
452 residues of the epitope decreased their RMFS (S9D Fig).
453
454 **Fig 5. Average structures from epitopes in solution and HLA-epitopes complexes**. (A).
455 G1 epitope. (B). G2 epitope. (C). H1 epitope. (D). H2 epitope.
456
457 Concerning clusterization of epitopes in solution, we observed that H1 epitope had
458 one predominant conformation visited throughout the simulation (S10C Fig). G1, G2, and
459 H2 epitopes visited at least three conformations during the simulation (S10A, 10B and 10D
460 Figs). When G1 formed the complex with the HLA, we observed one predominant structure
461 (S10E Fig), while G2, H1, and H2 even though it has one predominant structure, visited at
462 least other two conformations during the simulation (S10F-10H Figs).
463
464 We evaluated the total intra-protein bound for each epitope in solution, being 1.68,
465 2.48, 0.46, and 6.09 for G1, G2, H1, and H2 respectively (S11A-11D Figs). As we expected,
466 hydrogen bonds increased to 6, 7.75, 4.41, and 9.68 as mean value, respectively, when the
467 epitopes were coupled to the HLA molecule (S11E-11H Figs). Additionally, we identified
468 the hydrogen bonds with an occupancy greater than 20% to establish which interactions were
469 maintained for the longest time during the simulation (Table 2).
470
471 **Table 2. Hydrogen bonds occupancy in HLA-epitopes complexes.**

Epitope	Occupancy	Residue from Epitope	Epitope	Occupancy	Residue from Epitope

		– Residue from HLA			– Residue from HLA
G1	36.44%	V2:m –N82:s (chain B)	H1	45.50%	Y2:m –N82:s (chain B)
	22.59%	V2:m –N82:s (chain B)		36.85%	G9:s –W61:s (chain B)
	38.18%	A3:m –Q9:s (chain A)		22.21%	G9:m –W61:s (chain B)
	40.57%	S4:m –Q9:s (chain A)		25.10%	G9:s –Q64:s (chain B)
	33.80%	S4:m –Q9:s (chain A)		28.29%	G9:s –N69:s (chain A)
	22.99%	S4:s –Q9:s (chain A)		57.93%	G9:s –R71:s (chain B)
	62.77%	F8:m –Q64:s (chain B)	H2	85.09%	R3:m –H81:s (chain B)
	32.78%	F8:m –Q64:s (chain B)		25.82%	Y5:s –H81:m (chain B)
G2	27.02%	F1:m –N82:s (chain B)		27.89%	H6:m –N62:s (chain B)
	24.98%	F1:m –N82:s (chain B)		40.67%	M8:m –Y30:s (chain B)
	28.78%	Q2:m –Q9:s (chain A)		56.06%	S9:m –Y47:s (chain B)
	26.06%	H5:s –E55:m (chain A)		42.93%	S9:s –R71:s (chain B)
	111.12%	R7:s –N66:s (chain B)		67.22%	S10:m –W61:s (chain B)
	49.02%	K14:m –R71:s (chain B)		38.34%	H12:s –Y60:m (chain B)
	29.20%	N16:s –R71:s (chain B)		225.11%	E13:s –R76:s (chain A)

m – main chain

s – side chain

Finally, the solvent-accessible surface area (SASA) of the epitope in solution or forming the complex with HLA molecule was evaluated to determine the residues and the difference in the exposure. The SASA of the epitopes in solution was subtracted from the SASA of the epitopes forming the complex. These values were denominated as ΔSASA (Fig 6). All residues had a ΔSASA greater than zero, so a greater area of exposure to solvent was found in the complex.

Fig 6. Solvent accessible surface area (SASA). (A). G1 epitope. (B). G2 epitope. (C). H1 epitope. (D). H2 epitope**.** The blue bars depict the SASA from the epitope in solution. The green bars depict the SASA from the epitope in the MHC. The yellow bars depict the ΔSASA, subtracting the SASA of the epitopes in solution from the SASA of the epitopes forming the complex.

Discussion

In this study, we describe a method to rationally select EsxG·EsxH epitopes with the potential to be vaccine candidates for pulmonary TB. Through bioinformatics tools, we selected three epitopes with the best immunogenic characteristics: 1) G1 (^{11}LVASQSAFA^{19}), H1 (^{5}MYNYPAMLG^{13}), and H2 (^{65}LVRAYHAMSSTHE^{77}). Table 3 summarizes the characteristics of the selected epitopes. G1 epitope prediction was consistent with previous work that evaluated the response of $CD4^+$ T cells from mice infected with Mtb to ^{6}AHIPQLVASQSAFAAKAGLM^{25} synthetic peptide in *in vitro* assays, associating this peptide with protection (42). H1 epitope prediction was supported with previous studies,

498 which has shown that the N-terminal region, included epitope
499 ^{1}MSQIMYNYPAMLGHAGDM^{18}, was the most immunogenic of the protein (13,43), but
500 weakly recognized by T cell (12,13). Interestingly, ^{3}QMYNYPAM^{11} or ^{4}IMYNYPAM^{11}
501 induced strong specific responses of $CD8^{+}$ T cells (44,45) and were highly promiscuous to
502 bind MHC class I alleles (46). ^{3}QMYNYPAM^{11} residues are the minimum required to
503 generate cytotoxic response against Mtb in the acute and chronic stage of infection. This
504 cytotoxic effect is through degranulation mechanism and CD95L-mediated apoptosis (47).
505 In addition, it has been shown that peptide ^{5}MYNYPAMLG^{1} binds with low affinity to
506 A*0201, A*2402, and A*3002, MHC class I alleles (46,48). H2 epitope prediction matches
507 with evidence showing that ^{61}AMEDLVRAYHAMSSTHEA^{78} residues induced a strong
508 $CD4^{+}$ T cell response (49). ^{63}EDLVRAYHAMSSTHEANTMAMMARDTAEAAKWGG^{96}
509 residues have been more immunogenic in mice vaccinated with BCG (12). In contrast,
510 epitope ^{65}LVRAYHAMS^{73} was specific to MHC class I molecules (46). Our H2 epitope
511 includes H70, H76, and E77 residues, involved in the Zn^{2+} binding (19). Additionally, we
512 used G2 epitope to validate our selection strategy. It should be noted that G2 contains
513 residues previously associated with protection against TB (42). Given that MHC molecules
514 have high specificity, the difference in a single amino acid can be crucial for T-cell
515 recognition or ablation (13). Previous studies have shown that the substitution in one amino
516 acid in EsxH ^{4}IMYNYPAM^{11} epitope, significantly influenced the response and binding to
517 T cells, such as protection against TB in murine models (43,50).
518
519 **Table 3. Characteristics of the selected epitopes.** Comparation of characteristics from
520 epitopes in solution and when they are presented into HLA complex

Analysis	G1	G2	H1	H2	G1-HLA	G2-HLA	H1-HLA	H2-HLA
RMSD (nm)	0.310	0.350	0.095	0.297	0.357	0.434	0.447	0.525
RoG (nm)	0.833	0.806	0.709	1.065	2.402	2.434	2.417	2.445
Clusters (to cover 60%)	13	161	2	3	1	2	4	7
Hbonds	1.68	2.48	0.46	6.09	6	7.75	4.41	9.68
B-lineal epitope	Yes	No	Yes	Yes	N/A	N/A	N/A	N/A
B-conformational epitope	No	No	Yes	No	N/A	N/A	N/A	N/A

Molecular Dynamics simulations, a tool of structural vaccinology (4) predicts the stability of EsxG·EsxH complex and describes the epitopes. Our results from EsxG·EsxH complex at 310 K showed that the N and C ends of EsxG monomer were the most fluctuating and flexible regions of the complex. However, after 149 ns of simulation, these helices folded into the core of the EsxG monomer (Fig 2A), and this folding remained stable until the end of the simulation (Figs 2B, 2C). Despite the secondary structure analysis revealing certain modifications during the simulation, these changes were not relevant for changing the secondary structure of the dimer. Thus, the fluctuation and flexibility, coupled with their high exposition to the solution and the absence of important changes in the secondary structure of the dimer, could suggest a biological relevance, which is consistent with the conservation of some residues among mycobacterial species, previously demonstrated by Ilghari and

533 collaborators. Analysis of correlation matrix and contact map (Figs 2D, 2E) were in line with
534 the characteristic helix-turn-helix hairpin structure, arranged in an antiparallel way, forming
535 a bundle of 4 helices (19), and we proved that these interactions were held on a 500 ns
536 simulation, even though Zn^{2+} ion was not present in the dimer structure. Although we
537 observed some structural changes during the simulation, these changes did not alter the
538 regular secondary structure in the dimer, which was corroborated by the fact that the
539 hydrogen bonds and the RMSD remained constant, indicating that the complex was highly
540 stable for at least 500 ns. Our results are consistent with previous studies showing that both
541 monomers form a stable 1:1 heterodimeric complex (51). We observed that the RMSD
542 increased proportionally with the increase in temperature. As expected, the RoG and
543 hydrogen bonds decreased, indicating that the protein begins to undergo denaturation.
544 However, a large part of the secondary structure from the helices was maintained at 450 K
545 (Figs 4A-4G). These results showed that dimer significantly resisted denaturation. It is
546 necessary to form Zn^{2+} ion-protein complex to compare stability and aqueous solution protein
547 behavior described in this work.
548
549 The simulation of the EsxG monomer behavior in solution confirms the high
550 fluctuation and flexibility of the N and C ends (Figs 3A-3C). The PCA results showed that
551 both terminals moved correlatively between them, and they moved in anticorrelation with
552 the g1 and g2 helices. Additionally, the monomer EsxH showed loss of secondary structure
553 in helix h1 (Figs 3D-3G), indicating that EsxH as a monomer does not have a stable structure.
554 Despite this loss of structure, EsxH maintained its helical structure; however, in the analysis
555 carried out by Lightbody and collaborators, the EsxH monomer did not resist chemical
556 denaturation, so that a molten globule structure was formed. The analysis of the molten

557 globule structure revealed the formation of a hydrophobic mini nucleus. The nucleus was
558 formed by tryptophan residues. In this study, the tryptophan SASA was evaluated to
559 determine the changes in their exposure to the solvent in the EsxH monomer while forming
560 the complex with EsxG protein to evidence the formation of the hydrophobic nucleus and of
561 the molten globule. When we evaluated the tryptophan environment, we found that the
562 solvent exposure of the four tryptophans decreases considerably when the EsxH protein
563 forms a 1:1 complex with the EsxG (Table 4). The results agree with what was previously
564 reported by Lightbody and collaborators, where tryptophan residues W43, W54, and W58
565 formed a hydrophobic core, located at the interface of the EsxG-EsxH complex, while
566 tryptophan W94 was more exposed. This hydrophobic nucleus is necessary to induce the
567 partial folding of EsxH and may explain why EsxH as a monomer forms a molten globule**.**
568 Our data are in accordance with previous results and support the idea that they functionally
569 act as a dimer and not as individual proteins. Likewise, this suggests that they could act as a
570 heterodimer in acidic environments, such as the phagosome, without altering its structure
571 (51).

572 **Table 4. Solvent Accessible Surface Area in tryptophan residues.**

Residue	SASA (nm) EsxG·EsxH	SASA (nm) EsxH
W43	0.353	0.914
W54	0.006	0.260
W58	0.023	0.490
W94	0.823	1.804

573

574 Unfortunately, we were unable to generate an EsxG·EsxH/Zn^{2+} complex for the
575 simulation, since residues involved in metal coordination, as well as the geometry of the
576 metallic site, are not yet well established. This limitation could have influenced the mobility
577 of this region. Ilghari and collaborators (19) reported that the implied Zn^{2+} binding site was
578 located at one end of the complex, near the site where a cleft forms, suggesting a potential
579 site for interaction with a target protein. In our study, we observed the formation of this cleft,
580 at the very beginning of the simulation, however, this structure lost this conformation a few
581 ns after the simulation started (Fig 7), which could indicate that the cleft conformation is
582 restricted to Zn^{2+} presence or other targets. Even with the lack of Zn^{2+}, the dimer was still
583 stable during all simulation time.
584
585 **Fig 7. Cleft formation in EsxG·EsxH dimer at 310 K.** (A). 0 ns. (B). 149 ns. (C). 500 ns**.**
586 Blue represents the EsxG monomer. Orange represents the EsxH monomer. Residues
587 involved in cleft formation are presented in light blue (EsxG) and yellow (EsxH).
588
589 Because epitopes in solution are highly flexible and disordered, compared to when they
590 are part of a protein complex, conformational variations in these can lead to vaccine failure
591 or success, the reason why we consider it is important to predict the stability of epitopes. Our
592 results suggest that the H1 epitope was the most stable, presenting one conformation,
593 suggesting that this was the preferred conformation when it is alone in solution. Despite G1
594 and H2 epitopes presenting one predominant conformation, they visited at least three
595 conformations. G2 epitope in solution showed great fluctuation having several conformations
596 during the 600 ns simulation, which suggests that it is the one with the most variable
597 conformations in solution. These epitopes bind to multiple alleles of MHC class II molecules,

598 which are useful features for a vaccine candidate in a human population with heterogeneous
599 HLA (52). HLA class II molecules present antigens to CD4+ T cells, thus being essential for
600 eliminating Mtb (53). For this reason, we performed a docking to predict the presentation of
601 selected epitopes. DRB1 allele was selected because this binds the majority of known Mtb
602 epitopes; furthermore, it is the allele with the greater surface expression (36). Our results
603 indicate that when G1 formed a complex with HLA presented a predominant structure.
604 Despite G2 and H2 as complex had a majority conformation, they presented at least three
605 conformations. Interestingly, G2 epitope folds into an alpha-helix at residues 7 to 12.
606 Remarkably, H1 epitope visited at least four conformations when the HLA complex was
607 formed, which suggests that the binding to the HLA is not completely stable (Figs 5A-5D),
608 which was corroborated by the residues involved in the formation of hydrogen bonds (Table
609 2).
610
611 The three epitopes proposed and considered in the prediction are B cell receptor
612 epitopes too (Table 3). Although our predictions included linear and discontinuous epitopes,
613 most of the B cell epitopes are discontinuous and depend on the conformation. Therefore *in*
614 *silico* prediction is challenging (3). H1 epitope contains seven residues that match the
615 discontinuous B cell prediction. Thus, B lymphocytes can produce antibodies against the
616 EsxG and EsxH proteins that are involved in metabolic processes in Mtb, which could be
617 both neutralizing or bactericidal antibodies. Finally, the lack of homology of our epitopes
618 with human and rodent proteins suggests that they might avoid an autoimmune response in
619 future clinical trials. The EsxG and EsxH sequence identities between species implied that
620 the predicted epitopes could be used specifically against several Mtb complex strains. These
621 results support the finding that high conservation between species of Mtb complex and nearly

related (13,54) could be used against nontuberculous strains such as *M. kansasii*, *M. szulgai*, *M. simiae* and *M. gastri*. In agreement with previous studies, mortality related to nontuberculous mycobacteria surpasses TB infections in developed countries. This added to the difficulty of eradicating and its substantial re-occurrence, finding new therapies is crucial (55,56).

Conclusion

This work presented an approach to the rational design of vaccines against pulmonary TB. Using bioinformatics tools and Molecular Dynamics simulations, we selected three epitopes that are B-cell or T-cell epitopes. The epitopes were selected from the EsxG·EsxH complex, proteins involved in metal uptake, drug resistance, and immune response evasion makes it an ideal target for rational vaccine development. Thus, the three epitopes proposed in the present work have high potential to be subunit vaccine candidates or conjugated vaccines. These epitopes can be used in conjunction with adjuvants, or as a boost in BCG vaccination to have a robust and highly specific immune response. As they are B-cell epitopes, the generated antibodies could block the binding of metals or proteins involved in the phagolysosome formation, and thus in the prevention of the immune evasion, which could compromise Mtb survival. Although Molecular Dynamics simulations are a very reliable tool to predict molecular behavior, it is mandatory to perform *in vivo* assays with these predicted epitopes in order to demonstrate their immunogenic activity.

Acknowledgments

MO-CE gratefully acknowledges the scholarship from CONACYT to pursue her graduate studies. She would also like to express her deep gratitude to the graduate program in Biochemical Sciences at the National Autonomous University in Mexico (UNAM), for their continued support in this endeavor. EM thanks to LANCAD supercomputing center for facilities and infrastructure for calculation and storage data.

Funding

This paper was supported by grant 413957, CVU 406611 from the National Council for Science and Technology of Mexico (Consejo Nacional de Ciencia y Tecnología-CONACyT). Project number 21-2019 and 13-2020 by LANCAD supercomputing center.

Author Contributions

Conceptualization: Edgar Mixcoha; Constanza Estefania, Martínez-Olivares; Rogelio Hernández-Pando.

Data curation: Edgar Mixcoha; Constanza Estefania, Martínez-Olivares.

Formal analysis: Edgar Mixcoha; Constanza Estefania, Martínez-Olivares.

Funding Acquisition: Edgar Mixcoha.

Investigation: Edgar Mixcoha; Constanza Estefania, Martínez-Olivares.

Methodology: Edgar Mixcoha; Constanza Estefania, Martínez-Olivares.

Resources: Edgar Mixcoha.

Visualization: Edgar Mixcoha; Constanza Estefania, Martínez-Olivares.

Writing – Original Draft: Constanza Estefania, Martínez-Olivares.

Writing-Original Draft Preparation: Edgar Mixcoha; Constanza Estefania, Martínez-Olivares; Rogelio Hernández-Pando.

References

1. World Health Organization. Global Tuberculosis Report 2021. Genova; 2021.
2. Andersen P, Doherty TM. The success and failure of BCG — implications for a novel tuberculosis vaccine. Nat Rev Microbiol. 2005;3(8):656–62.
3. Rueckert C, Guzmán CA. Vaccines: From Empirical Development to Rational Design.

688 PLoS Pathog. 2012;8(11).

689 4. Ribas-Aparicio RM, Castelán-Vega JA, Jiménez-Alberto A, Monterrubio-López GP,
690 Aparicio-Ozores G. The Impact of Bioinformatics on Vaccine Design and
691 Development. In: InTechOpen, editor. Vaccines [Internet]. 6th ed. Rijeka, Croatia;
692 2017. Available from: https://www.intechopen.com/books/advanced-biometric-
693 technologies/liveness-detection-in-biometrics

694 5. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of
695 peptides in vaccine design. Nat Rev Drug Discov. 2007;6(5):404–14.

696 6. van 't Wout JW, Poell R, van Furth R. The role of BCG/PPD-activated macrophages
697 in resistance against systemic candidiasis in mice. Scand J Immunol. 1992;36(5):713–
698 719.

699 7. Lu LL, Chung AW, Rosebrock TR, Ghebremichael M, Yu WH, Grace PS, et al. A
700 functional role for antibodies in tuberculosis. Cell. 2016;167(2):433–443.e14.

701 8. Serafini A, Boldrin F, Palù G, Manganelli R. Characterization of a Mycobacterium
702 tuberculosis ESX-3 conditional mutant: Essentiality and rescue by iron and zinc. J
703 Bacteriol. 2009;191(20):6340–4.

704 9. Serafini A, Pisu D, Palù G, Rodriguez GM, Manganelli R. The ESX-3 Secretion
705 System Is Necessary for Iron and Zinc Homeostasis in Mycobacterium tuberculosis.
706 PLoS One. 2013;8(10).

707 10. Sloan Siegrist M, Steigedal M, Ahmad R, Mehra A, Dragset MS, Schuster BM, et al.
708 Mycobacterial Esx-3 requires multiple components for iron acquisition. MBio.
709 2014;5(3):e01073-14.

710 11. Tinaztepe E, Wei JR, Raynowska J, Portal-Celhay C, Thompson V, Philipsa JA. Role
711 of metal-dependent regulation of ESX-3 secretion in intracellular survival of

712 Mycobacterium tuberculosis. Infect Immun. 2016;84(8):2255–63.
713 12. Hervas-Stubbs S, Majlessi L, Simsova M, Morova J, Rojas MJ, Nouzé C, et al. High
714 frequency of CD4+ T cells specific for the TB10.4 protein correlates with protection
715 against Mycobacterium tuberculosis infection. Infect Immun. 2006;74(6):3396–407.
716 13. Skjøt RLV, Brock I, Arend SM, Munk ME, Theisen M, Ottenhoff THM, et al. Epitope
717 Mapping of the Immunodominant Antigen TB10.4 and the Two Homologous Proteins
718 TB10.3 and TB12.9, Which Constitute a Subfamily of the esat-6 Gene Family. Infect
719 Immun. 2002;70(10):5446–53.
720 14. Mehra A, Zahra A, Thompson V, Sirisaengtaksin N, Wells A, Porto M, et al.
721 Mycobacterium tuberculosis Type VII Secreted Effector EsxH Targets Host ESCRT
722 to Impair Trafficking. PLoS Pathog. 2013;9(10).
723 15. Mittal E, Skowyra ML, Uwase G, Tinaztepe E, Mehra A, Köster S, et al.
724 Mycobacterium tuberculosis Type VII Secretion System Effectors Differentially
725 Impact the ESCRT Endomembrane Damage Response. MBio. 2018;9(6):e01765-18.
726 16. Ankley L, Thomas S, Olive AJ. Fighting persistence; how chronic infections with
727 Mycobacterium tuberculosis evade T cell-mediated clearance and new strategies to
728 defeat them. Infect Immun [Internet]. 2020;IAI.00916-19. Available from:
729 http://iai.asm.org/content/early/2020/02/19/IAI.00916-19.abstract
730 17. Portal-Celhay C, Tufariello JM, Srivastava S, Zahra A, Klevorn T, Grace PS, et al.
731 Mycobacterium tuberculosis EsxH inhibits ESCRT-dependent CD4+ T-cell
732 activation. Nat Microbiol. 2016;2(2):1–9.
733 18. Peñuelas-Urquides K, González-Escalante L, Villarreal-Treviño L, Silva-Ramírez B,
734 Gutiérrez-Fuentes DJ, Mojica-Espinosa R, et al. Comparison of gene expression
735 profiles between pansensitive and multidrug-resistant strains of mycobacterium

736 tuberculosis. Curr Microbiol. 2013;67(3):362–71.

737 19. Ilghari D, Lightbody KL, Veverka V, Waters LC, Muskett FW, Renshaw PS, et al.
738 Solution structure of the Mycobacterium tuberculosis EsxG·EsxH complex:
739 Functional implications and comparisons with other M. tuberculosis Esx family
740 complexes. J Biol Chem. 2011;286(34):29993–30002.

741 20. Singh H, Raghava GPS. ProPred: prediction of HLA-DR binding sites.
742 Bioinformatics [Internet]. 2001 Dec 1;17(12):1236–7. Available from:
743 https://academic.oup.com/bioinformatics/article-
744 lookup/doi/10.1093/bioinformatics/17.12.1236

745 21. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved
746 methods for predicting peptide binding affinity to MHC class II molecules.
747 Immunology. 2018;154(3):394–406.

748 22. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: Improving sequence-
749 based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res.
750 2017;45(W1):W24–9.

751 23. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B Cell Epitope
752 Predictions: Impacts of Method Development and Improved Benchmarking. PLoS
753 Comput Biol. 2012;8(12):e1002829.

754 24. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic
755 determinants on protein antigens. FEBS Lett. 1990;276(1–2):172–4.

756 25. Goodsell DS, Zardecki C, Di Costanzo L, Duarte JM, Hudson BP, Persikova I, et al.
757 RCSB Protein Data Bank: Enabling biomedical research and drug discovery. Protein
758 Sci. 2020;29(1):52–65.

759 26. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of

760 Multiple Amber Force Fields and Development of Improved Protein Backbone
761 Parameters. PROTEINS Struct Funct Bioinforma. 2006;65:712–25.
762 27. Berendsen HJC, Postma JPM, W.F. van G, J. H. Interaction Models for Water in
763 Relation to Protein Hydration. In: Intermolecular Forces The Jerusalem Symposia on
764 Quantum Chemistry and Biochemistry. Jerusalem, Israel: Springer, Dordrecht; 1981.
765 28. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing
766 parallel molecular dynamics implementation. Comput Phys Commun. 1995;91(1–
767 3):43–56.
768 29. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC.
769 GROMACS: Fast, flexible, and free. J Comput Chem. 2005;26(16):1701–18.
770 30. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular
771 dynamics method. J Appl Phys. 1981;52(12):7182–90.
772 31. Hünenberger PH. Thermostat Algorithms for Molecular Dynamics simulations. Adv
773 Polym Sci. 2005;173:105–49.
774 32. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald
775 sums in large systems. J Chem Phys. 1993;98(12):10089–92.
776 33. Holling TM, Schooten E, van Den Elsen PJ. Function and regulation of MHC class
777 II molecules in T-lymphocytes: Of mice and men. Hum Immunol. 2004;65(4):282–
778 90.
779 34. Katikanenia DS, Jina L. B cell MHC class II signaling: A story of life and death. Hum
780 Immunol. 2019;80(1):37–43.
781 35. Ting YT, Petersen J, Ramarathinam SH, Scally SW, Loh KL, Thomas R, et al. The
782 interplay between citrullination and HLA-DRB1 polymorphism in shaping peptide
783 binding hierarchies in rheumatoid arthritis. J Biol Chem. 2018;293(9):3236–51.

784 36. Knudsen NPH, Nørskov-Lauritsen S, Dolganov GM, Schoolnik GK, Lindenstrøm T,
785 Andersen P, et al. Tuberculosis vaccine with high predicted population coverage and
786 compatibility with modern diagnostics. Proc Natl Acad Sci U S A. 2014;111(3):1096–
787 101.
788 37. Kurcinski M, Jamroz M, Blaszczyk M, Kolinski A, Kmiecik S. CABS-dock web
789 server for the flexible docking of peptides to proteins without prior knowledge of the
790 binding site. Nucleic Acids Res. 2015;43(W1):W419–24.
791 38. London N, Raveh B, Cohen E, Fathi G, Schueler-Furman O. Rosetta FlexPepDock
792 web server - high resolution modeling of peptide-protein interactions. Nucleic Acids
793 Res. 2011;39(Web Server issue):W249–53.
794 39. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide
795 Folding: When Simulation Meets Experiment. Angew Chemie Int Ed.
796 1999;38(1/2):236–40.
797 40. DeLano W. The PyMOL Molecular Graphics System [Internet]. Palo Alto, CA, USA:
798 Schrödinger, LLC.; 2002. Available from: http://www.pymol.org.
799 41. Williams T, Kelley C. gnuplot 5.0. An Interactive Plotting Program [Internet]. 2017.
800 Available from: http://gnuplot.info
801 42. Coler RN, Dillon DC, Skeiky YAW, Kahn M, Orme IM, Lobet Y, et al. Identification
802 of Mycobacterium tuberculosis vaccine candidates using human CD4+ T-cells
803 expression cloning. Vaccine. 2009;27(2):223–33.
804 43. Lindenstrøm T, Aagaard C, Christensen D, Agger EM, Andersen P. High-frequency
805 vaccine-induced CD8+ T cells specific for an epitope naturally processed during
806 infection with M. tuberculosis do not confer protection. Eur J Immunol.
807 2014;44(6):1699–709.

808 44. Flórido M, Pillay R, Gillis CM, Xia Y, Turner SJ, Triccas JA, et al. Epitope-specific
809 CD4+, but not CD8+, T-cell responses induced by recombinant influenza A viruses
810 protect against Mycobacterium tuberculosis infection. Eur J Immunol.
811 2015;45(3):780–93.
812 45. Hoang TTKT, Nansen A, Roy S, Billeskov R, Aagaard C, Elvang T, et al. Distinct
813 Differences in the Expansion and Phenotype of TB10.4 Specific CD8 and CD4 T Cells
814 after Infection with Mycobacterium tuberculosis. PLoS One. 2009;4(6):e5928.
815 46. Axelsson-Robertson R, Ahmed RK, Weichold FF, Ehlers MM, Kock MM, Sizemore
816 D, et al. Human Leukocyte Antigens A*3001 and A*3002 Show distinct Peptide-
817 Binding Patterns of the Mycobacterium tuberculosis Protein TB10.4: Consequences
818 for Immune Recognition. Clin Vaccine Immunol. 2011;18(1):125–34.
819 47. Billeskov R, Vingsbo-Lundberg C, Andersen P, Dietrich J. Induction of CD8 T Cells
820 against a Novel Epitope in TB10.4: Correlation with Mycobacterial Virulence and the
821 Presence of a Functional Region of Difference-1. J Immunol. 2007;179(6):3973–81.
822 48. Axelsson-Robertson R, Weichold F, Sizemore D, Wulf M, Skeiky YAW, Sadoff J, et
823 al. Extensive major histocompatibility complex class I binding promiscuity for
824 Mycobacterium tuberculosis TB10.4 peptides and immune dominance of human
825 leucocyte antigen (HLA)-B*0702 and HLA-B*0801 alleles in TB10.4 CD8+ T-cell
826 responses. Immunology. 2010;129(4):496–505.
827 49. Dietrich J, Aagaard C, Leah R, Olsen AW, Stryhn A, Doherty TM, et al. Exchanging
828 ESAT6 with TB10.4 in an Ag85B Fusion Molecule-Based Tuberculosis Subunit
829 Vaccine: Efficient Protection and ESAT6-Based Sensitive Monitoring of Vaccine
830 Efficacy. J Immunol. 2005;174(10):6332–9.
831 50. Axelsson-Robertson R, Ju JH, Kim HY, Zumla A, Maeurer M. Mycobacterium

832 tuberculosis-specific and MHC class I-restricted CD8+ T-cells exhibit a stem cell
833 precursor-like phenotype in patients with active pulmonary tuberculosis. Int J Infect
834 Dis [Internet]. 2015;32:13–22. Available from:
835 http://dx.doi.org/10.1016/j.ijid.2014.12.017
836 51. Lightbody KL, Ilghari D, Waters LC, Carey G, Bailey MA, Williamson RA, et al.
837 Molecular features governing the stability and specificity of functional complex
838 formation by Mycobacterium tuberculosis CFP-10/ESAT-6 family proteins. J Biol
839 Chem. 2008;283(25):17681–90.
840 52. Mustafa AS, Skeiky YA, Al-Attiyah R, Alderson MR, Hewinson RG, Vordermeier
841 HM. Immunogenicity of Mycobacterium tuberculosis Antigens in Mycobacterium
842 bovis BCG-Vaccinated and M. bovis-Infected Cattle. Infect Immun.
843 2006;74(8):4566–72.
844 53. Hoft DF. Tuberculosis vaccine development: goals, immunological design, and
845 evaluation. Lancet. 2008;372(9633):164–75.
846 54. Davila J, Zhang L, Marrs CF, Durmaz R, Yang Z. Assessment of the genetic diversity
847 of mycobacterium tuberculosis esxA, esxH, and fbpB genes among clinical isolates
848 and its implication for the future immunization by new tuberculosis subunit vaccines
849 Ag85B-ESAT-6 and Ag85B-TB10.4. J Biomed Biotechnol. 2010;2010.
850 55. Johnson MM, Odell JA. Nontuberculous mycobacterial pulmonary infections. J
851 Thorac Dis. 2014;6(3):210–20.
852 56. Mirsaeidi M, Machado RF, Garcia JGN, Schraufnagel DE. Nontuberculous
853 Mycobacterial Disease Mortality in the United States, 1999-2010: A Population-Based
854 Comparative Study. PLoS One. 2014;9(3):1–9.
855

Supporting information

All Simulations carried out in this work are available to be downloaded in:

Mixcoha E, Martinez Olivares CE, Hernández-Pando R (2022) Native and Thermal Unfolding simulations EsxG·EsxH. https://doi.org/10.6084/m9.figshare.20063552.v1

Mixcoha E, Martinez Olivares CE, Hernández-Pando R (2022) Epitopes in solution Molecular Dynamics Simulations. https://doi.org/10.6084/m9.figshare.20055002.v1

Mixcoha E, Martinez Olivares CE, Hernández-Pando R (2022) Molecular Dynamics Simulation of MHCII-H2. https://doi.org/10.6084/m9.figshare.20054138.v2

Mixcoha E, Martinez Olivares CE, Hernández-Pando R (2022) Molecular Dynamics Simulations of MHCII-G1. https://doi.org/10.6084/m9.figshare.20054159.v3

Mixcoha E, Martinez Olivares CE, Hernández-Pando R, Martinez Olivares CE (2022) Molecular Dynamics Simulation of MHCII-G2. https://doi.org/10.6084/m9.figshare.20054402.v1

Mixcoha E, Martinez Olivares CE, Hernández-Pando R (2022) Molecular Dynamics Simulations of MHCII-H1.https://doi.org/10.6084/m9.figshare.20054423.v1

878 **S1 Table. MHC class-II binding epitope prediction.** 49 HLA-DR alleles were considered
879 to locate promiscuous binding regions. The output consists of a list of nonameric epitopes
880 capable of binding into a given HLA-DR allele.
881
882 **S1 Fig. A. Root Mean Square Deviation (RMSD) plot from EsxG·EsxH complex at 310**
883 **K.** Red line represents the RMSD value during 500 ns of simulation. Blue line represents the
884 RMSD average. Shaded bar represents standard deviation. **B. Radius of Gyration (RoG)**
885 **plot from EsxG·EsxH complex at 310 K.** Red line stand for RoG value during 500 ns of
886 simulation. Blue line stands for the RoG average. Shaded bar represents standard deviation.
887 **C. Root Mean Square Fluctuation (RMSF) plot from EsxG·EsxH complex at 310 K**. X
888 axis corresponds to residue number. The blue color depicts the EsxG monomer. The orange
889 color depicts the EsxH monomer. **D. Secondary structure plot from EsxG·EsxH complex**
890 **at 310 K.** Residues 1 to 97 represents EsxG monomer, residues 98 to 194 represents EsxH
891 monomer. **E. Clusterization plot from EsxG·EsxH complex at 310 K.** X axis stand for the
892 cluster ID. Y axis stand for the number of structures in the cluster. **F. Hydrogen bonds plot**
893 **from EsxG·EsxH complex at 310 K.** Black line stand for total hydrogen bonds during 500
894 ns of simulation.
895
896 **S1 Video. Contact map from EsxG·EsxH dimer at 310 K, 350K, 400K, 450K, 500K and,**
897 **550K.** All contact map videos are available to watch in:
898 https://doi.org/10.6084/m9.figshare.20063552
899
900 **S2 Fig. A. Root Mean Square Deviation (RMSD) plot from EsxG monomer at 310 K.**
901 Red line depict the RMSD value during 250 ns of simulation. Blue line depicts the RMSD

902 average. Shaded bar represents standard deviation. **B. Radius of Gyration (RoG) plot from**
903 **EsxG monomer at 310 K.** Red line represents RoG value during 250 ns of simulation. Blue
904 line represents the RoG average. Shaded bar represents standard deviation. **C. Hydrogen**
905 **bonds plot from EsxG monomer at 310 K.** Black line stand for total hydrogen bonds during
906 250 ns of simulation. **D. Root Mean Square Fluctuation (RMSF) plot from EsxG**
907 **monomer at 310 K**. X axis corresponds to residue number. **E. Secondary structure plot**
908 **from EsxG monomer at 310 K.** X axis corresponds to residue number. **F. Clusterization**
909 **plot from EsxG monomer at 310 K.** X axis represents the cluster ID. Y axis represents the
910 number of structures.
911
912 **S3 Fig. A. Root Mean Square Deviation (RMSD) plot from EsxH monomer at 310 K.**
913 Red line depict the RMSD value during 250 ns of simulation. Blue line depicts the RMSD
914 average. Shaded bar represents standard deviation. **B. Radius of Gyration (RoG) plot from**
915 **EsxH monomer at 310 K.** Red line represents RoG value during 250 ns of simulation. Blue
916 line represents the RoG average. Shaded bar represents standard deviation. **C. Hydrogen**
917 **bonds plot from EsxH monomer at 310 K.** Black line represents total hydrogen bonds
918 during 250 ns of simulation. **D. Root Mean Square Fluctuation (RMSF) plot from EsxH**
919 **monomer at 310 K**. X axis corresponds to residue number. **E. Secondary structure plot**
920 **from EsxH monomer at 310 K.** X axis corresponds to residue number. **F. Clusterization**
921 **plot from EsxH monomer at 310 K.** X axis represents the cluster ID. Y axis represents the
922 number of structures.
923
924 **S4 Fig. A. Root Mean Square Deviation (RMSD) plot from EsxG·EsxH complex at 350**
925 **K.** Red line stand for the RMSD value during 50 ns of simulation. Blue line stands for the

926 RMSD average. Shaded bar represents standard deviation. **B. Radius of Gyration (RoG)**
927 **plot from EsxG·EsxH complex at 350 K.** Red line depict RoG value during 50 ns of
928 simulation. Blue line depicts the RoG average. Shaded bar represents standard deviation. **C.**
929 **Root Mean Square Fluctuation (RMSF) plot from EsxG·EsxH complex at 350 K**. X axis
930 corresponds to residue number. The blue color depicts the EsxG monomer. The orange color
931 represents the EsxH monomer. **D. Hydrogen bonds plot from EsxG·EsxH complex at 350**
932 **K.** Black line represents total hydrogen bonds during 50 ns of simulation.
933
934 **S5 Fig. A. Root Mean Square Deviation (RMSD) plot from EsxG·EsxH complex at 400**
935 **K.** Red line represents the RMSD value during 5 ns of simulation. Blue line represents the
936 RMSD average. Shaded bar represents standard deviation. **B. Root Mean Square Deviation**
937 **(RMSD) plot from EsxG·EsxH complex at 450 K.** Red line stand for the RMSD value
938 during 5 ns of simulation. Blue line stands for the RMSD average. Shaded bar stands for
939 standard deviation. **C. Root Mean Square Deviation (RMSD) plot from EsxG·EsxH**
940 **complex at 500 K.** Red line depict the RMSD value during 5 ns of simulation. Blue line
941 depicts the RMSD average. Shaded bar depicts standard deviation. **D. Root Mean Square**
942 **Deviation (RMSD) plot from EsxG·EsxH complex at 550 K.** Red line represents the
943 RMSD value during 5 ns of simulation. Blue line represents the RMSD average. Shaded bar
944 represents standard deviation. **E. Radius of Gyration (RoG) plot from EsxG·EsxH**
945 **complex at 400 K.** Red line stand for RoG value during 5 ns of simulation. Blue line stands
946 for the RoG average. Shaded bar stands for standard deviation. **F. Radius of Gyration (RoG)**
947 **plot from EsxG·EsxH complex at 450 K.** Red line represents RoG value during 5 ns of
948 simulation. Blue line represents the RoG average. Shaded bar represents standard deviation.
949 **G. Radius of Gyration (RoG) plot from EsxG·EsxH complex at 500 K.** Red line depict

RoG value during 5 ns of simulation. Blue line depicts the RoG average. Shaded bar represents standard deviation. **H. Radius of Gyration (RoG) plot from EsxG·EsxH complex at 550 K.** Red line represents RoG value during 5 ns of simulation. Blue line represents the RoG average. Shaded bar represents standard deviation.

S6 Fig. A. Secondary structure plot from EsxG·EsxH complex at 400 K. Residues 1 to 97 depict EsxG monomer, residues 98 to 194 depict EsxH monomer. **B. Secondary structure plot from EsxG·EsxH complex at 450 K.** Residues 1 to 97 depict EsxG monomer, residues 98 to 194 represents EsxH monomer. **C. Secondary structure plot from EsxG·EsxH complex at 500 K.** Residues 1 to 97 stand for EsxG monomer, residues 98 to 194 stand for EsxH monomer. **D. Secondary structure plot from EsxG·EsxH complex at 550 K.** Residues 1 to 97 represents EsxG monomer, residues 98 to 194 represents EsxH monomer. **E. Hydrogen bonds plot from EsxG·EsxH complex at 400 K.** Black line depitc total hydrogen bonds during 5 ns of simulation. **F. Hydrogen bonds plot from EsxG·EsxH complex at 450 K.** Black line depict total hydrogen bonds during 5 ns of simulation. **G. Hydrogen bonds plot from EsxG·EsxH complex at 500 K.** Black line stand for total hydrogen bonds during 5 ns of simulation. **H. Hydrogen bonds plot from EsxG·EsxH complex at 550 K.** Black line stand for total hydrogen bonds during 5 ns of simulation.

S7 Fig. A. Root Mean Square Deviation (RMSD) plot from G1·HLA complex at 310 K. Red line represents the RMSD value from the G1-HLA complex during 250 ns of simulation. Black line represents the RMSD value from the G1 epitope in solution during 250 ns of simulation. Blue line represents the RMSD average. Shaded bar represents standard deviation. **B. Root Mean Square Deviation (RMSD) plot from G2·HLA complex at 310**

974 **K.** Red line depict the RMSD value from the G2-HLA complex during 250 ns of simulation.
975 Black line depicts the RMSD value from the G2 epitope in solution during 250 ns of
976 simulation. Blue line depicts the RMSD average. Shaded bar represents standard deviation.
977 **C. Root Mean Square Deviation (RMSD) plot from H1·HLA complex at 310 K.** Red line
978 stand for the RMSD value from the H1-HLA complex during 250 ns of simulation. Black
979 line stands for the RMSD value from the H1 epitope in solution during 250 ns of simulation.
980 Blue line stands for the RMSD average. Shaded bar stands for standard deviation. **D. Root**
981 **Mean Square Deviation (RMSD) plot from H2·HLA complex at 310 K.** Red line
982 represents the RMSD value from the H2-HLA complex during 250 ns of simulation. Black
983 line represents the RMSD value from the H2 epitope in solution during 250 ns of simulation.
984 Blue line represents the RMSD average. Shaded bar represents standard deviation. **E. Root**
985 **Mean Square Deviation (RMSD) plot from G1·HLA complex at 310 K.** Red line depict
986 the RMSD value from the G1-HLA complex during 250 ns of simulation. Green line depicts
987 the RMSD value from the G1 epitope in solution during 250 ns of simulation. Blue line
988 depicts the RMSD average. Shaded bar represents standard deviation. **F. Root Mean Square**
989 **Deviation (RMSD) plot from G1·HLA complex at 310 K.** Red line stand for the RMSD
990 value from the G2-HLA complex during 250 ns of simulation. Green line stands for the
991 RMSD value from the G2 epitope in solution during 250 ns of simulation. Blue line stands
992 for the RMSD average. Shaded bar represents standard deviation. **G. Root Mean Square**
993 **Deviation (RMSD) plot from H1·HLA complex at 310 K.** Red line depict the RMSD value
994 from the H1-HLA complex during 250 ns of simulation. Green line depicts the RMSD value
995 from the H1 epitope in solution during 250 ns of simulation. Blue line depicts the RMSD
996 average. Shaded bar represents standard deviation. **H. Root Mean Square Deviation**
997 **(RMSD) plot from H2·HLA complex at 310 K.** Red line represents the RMSD value from

998 the H2-HLA complex during 250 ns of simulation. Green line represents the RMSD value
999 from the H2 epitope in solution during 250 ns of simulation. Blue line represents the RMSD
1000 average. Shaded bar represents standard deviation.
1001
1002 **S8 Fig. A. Secondary structure plot from G1 epitope in solution at 310 K.** Residues 1 to
1003 9 stand for the G1 epitope LVASQSAFA. **B. Secondary structure plot from G2 epitope in**
1004 **solution at 310 K.** Residues 1 to 16 stand for the G2 epitope FQAAHARFVAAAAKVN. **C.**
1005 **Secondary structure plot from H1 epitope in solution at 310 K.** Residues 1 to 9 represents
1006 the H1 epitope MYNYPAMLG. **D. Secondary structure plot from H2 epitope in solution**
1007 **at 310 K.** Residues 1 to 13 represents the H2 epitope LVRAYHAMSSTHE. **E. Secondary**
1008 **structure plot from G1 epitope in HLA at 310 K.** Residues 1 to 9 depict the G1 epitope
1009 LVASQSAFA. **F. Secondary structure plot from G2 epitope in HLA at 310 K.** Residues
1010 1 to 16 depict the G2 epitope FQAAHARFVAAAAKVN. **G. Secondary structure plot**
1011 **from H1 epitope in HLA at 310 K.** Residues 1 to 9 depict the H1 epitope MYNYPAMLG.
1012 **H. Secondary structure plot from H2 epitope in HLA at 310 K.** Residues 1 to 13 depict
1013 the H2 epitope LVRAYHAMSSTHE.
1014
1015 **S9 Fig. A. Root Mean Square Fluctuation (RMSF) plot from G1 epitope in solution, and**
1016 **in HLA complex at 310 K**. X axis corresponds to residue number, residues 1 to 9 stand for
1017 the G1 epitope LVASQSAFA. Blue color stand for G1 epitope in solution. Black color stand
1018 for G1 epitope in complex with HLA. **B. Root Mean Square Fluctuation (RMSF) plot**
1019 **from G2 epitope in solution, and in HLA complex at 310 K**. X axis corresponds to residue
1020 number, residues 1 to 16 depict the G2 epitope FQAAHARFVAAAAKVN. Blue color
1021 represents G1 epitope in solution. Black color depicts G1 epitope in complex with HLA. **C.**

Root Mean Square Fluctuation (RMSF) plot from H1 epitope in solution, and in HLA complex at 310 K. X axis corresponds to residue number, residues 1 to 9 depict the H1 epitope MYNYPAMLG. Blue color depicts G1 epitope in solution. Black color depicts G1 epitope in complex with HLA. **D. Root Mean Square Fluctuation (RMSF) plot from H2 epitope in solution, and in HLA complex at 310 K**. X axis corresponds to residue number, residues 1 to 13 represents the H2 epitope LVRAYHAMSSTHE. Blue color represents G1 epitope in solution. Black color represents G1 epitope in complex with HLA.

S10 Fig. A. Clusterization plot from G1 epitope in solution at 310 K. X axis depict the cluster ID. Y axis depict the number of structures. **B. Clusterization plot from G2 epitope in solution at 310 K.** X axis represents the cluster ID. Y axis represents the number of structures. **C. Clusterization plot from H1 epitope in solution at 310 K.** X axis stand for the cluster ID. Y axis stand for the number of structures. **D. Clusterization plot from H2 epitope in solution at 310 K.** X axis represents the cluster ID. Y axis represents the number of structures. **E. Clusterization plot from G1 epitope in HLA at 310 K.** X axis depict the cluster ID. Y axis depict the number of structures. **F. Clusterization plot from G2 epitope in HLA at 310 K.** X axis stand for the cluster ID. Y axis stand for the number of structures. **G. Clusterization plot from H1 HLA in solution at 310 K.** X axis represents the cluster ID. Y axis represents the number of structures. **H. Clusterization plot from H2 epitope in HLA at 310 K.** X axis stand for the cluster ID. Y axis stand for the number of structures.

S11 Fig. A. Hydrogen bonds plot from G1 in solution at 310 K. Orange line depict total hydrogen bonds during 250 ns of simulation. **B. Hydrogen bonds plot from G2 in solution at 310 K.** Orange line depict total hydrogen bonds during 250 ns of simulation. **C. Hydrogen**

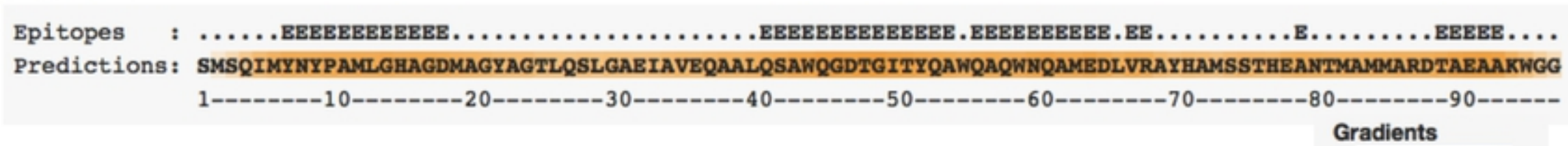
bonds plot from H1 in solution at 310 K. Orange line represents total hydrogen bonds during 250 ns of simulation. **D. Hydrogen bonds plot from H2 in solution at 310 K.** Orange line represents total hydrogen bonds during 250 ns of simulation. **E. Hydrogen bonds plot from G1 in HLA at 310 K.** Black line stand for total hydrogen bonds during 250 ns of simulation. **F. Hydrogen bonds plot from G2 in HLA at 310 K.** Black line stand for total hydrogen bonds during 250 ns of simulation. **G. Hydrogen bonds plot from H1 in HLA at 310 K.** Black line represents total hydrogen bonds during 250 ns of simulation. **H. Hydrogen bonds plot from H2 in HLA at 310 K.** Black line represents total hydrogen bonds during 250 ns of simulation.

V1 Fig. A. Hydrogen bonds plot from G1 in solution at 310 K. Orange line depict total hydrogen bonds during 250 ns of simulation.

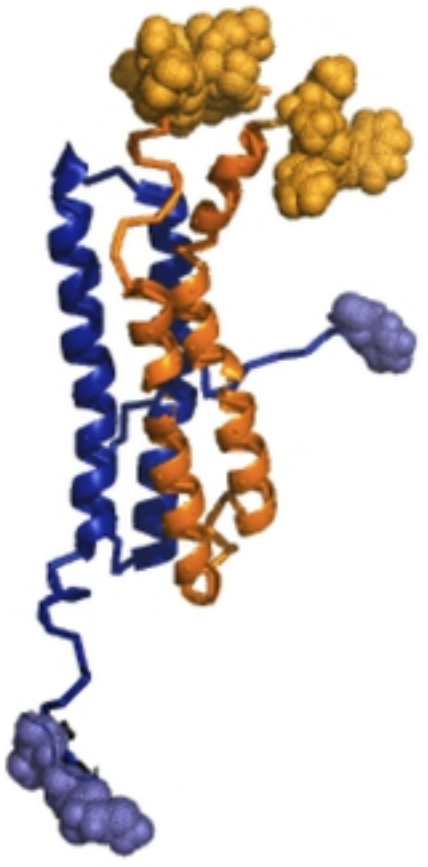
A. Linear B-cell epitope prediction

EsxG

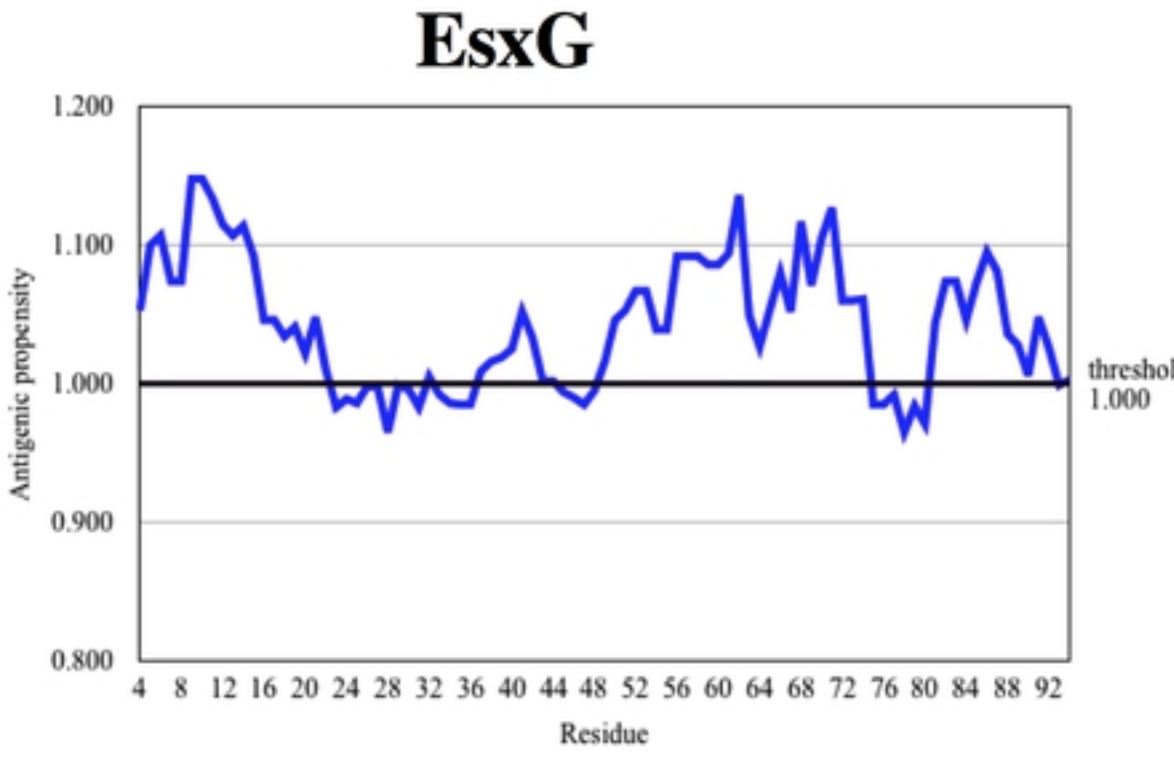
EsxH



B. Discontinuous B-cell epitope prediction



B. Discontinuous B-cell epitope prediction



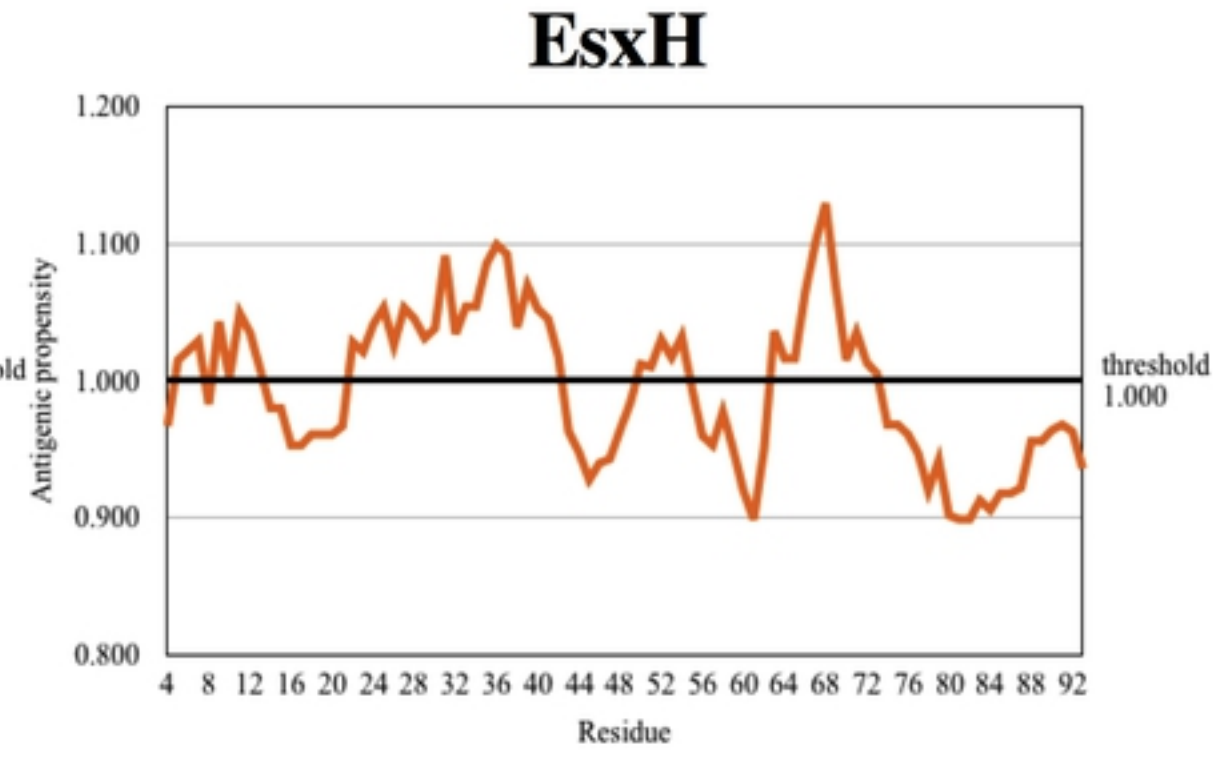
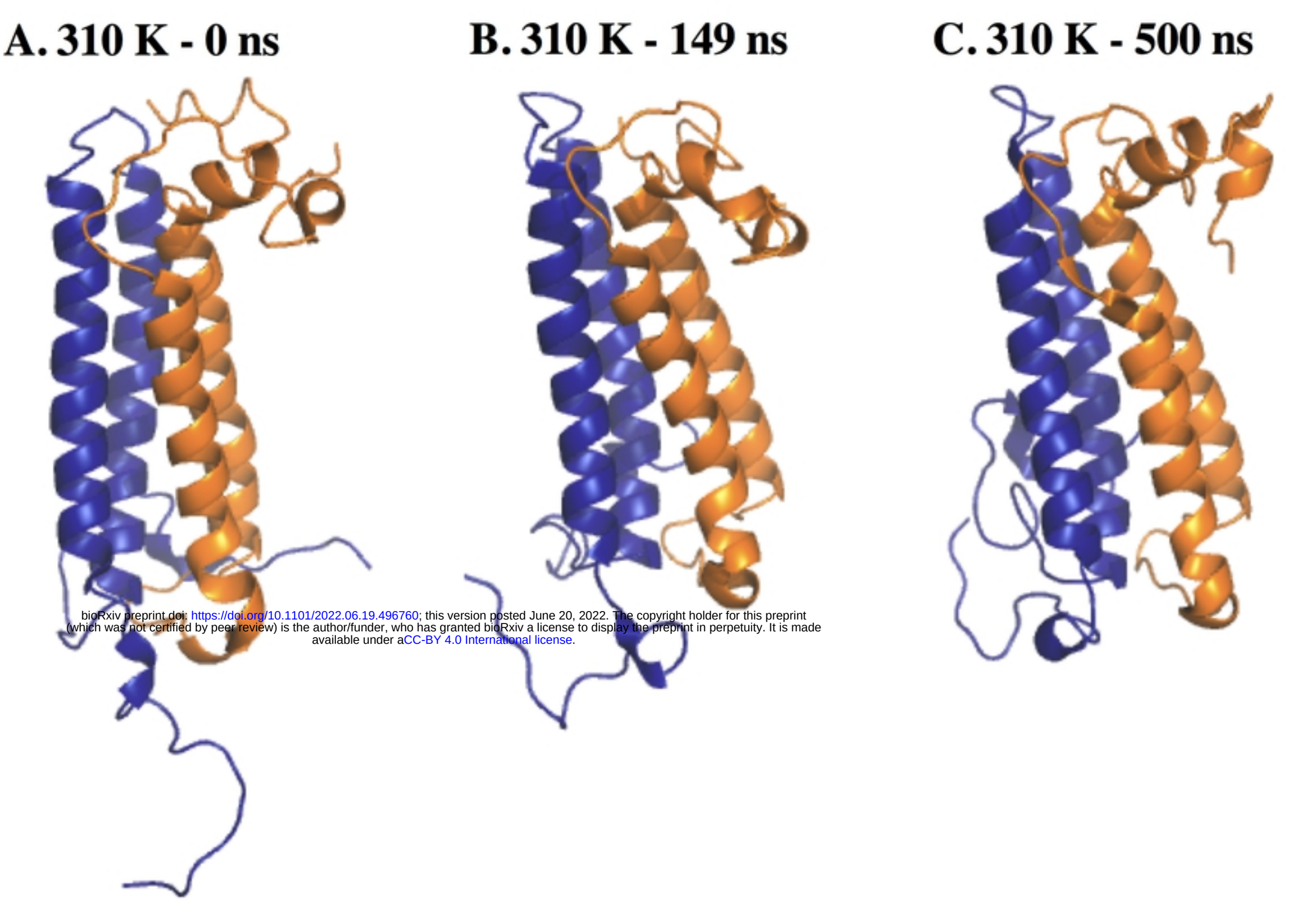
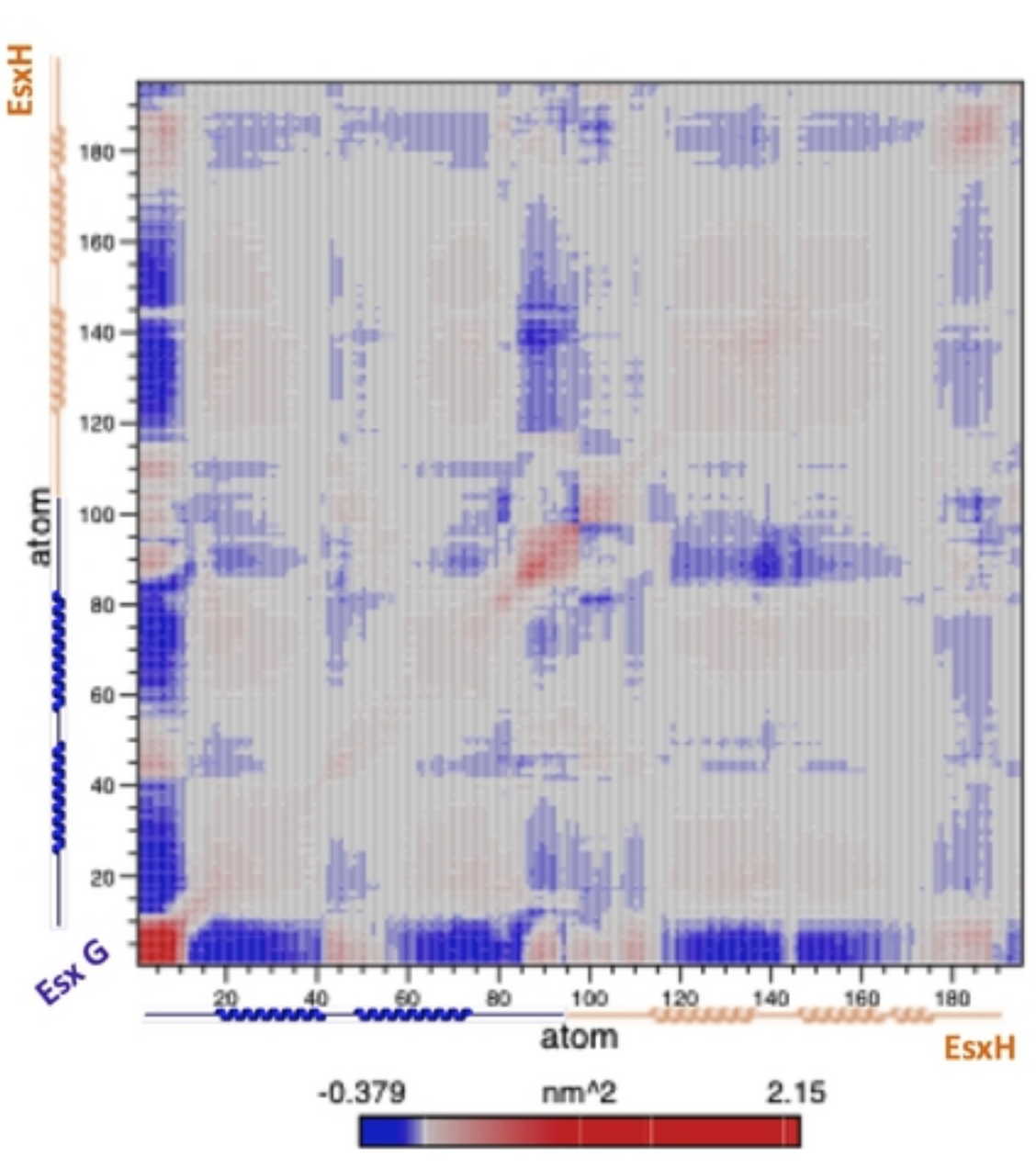


Figure1





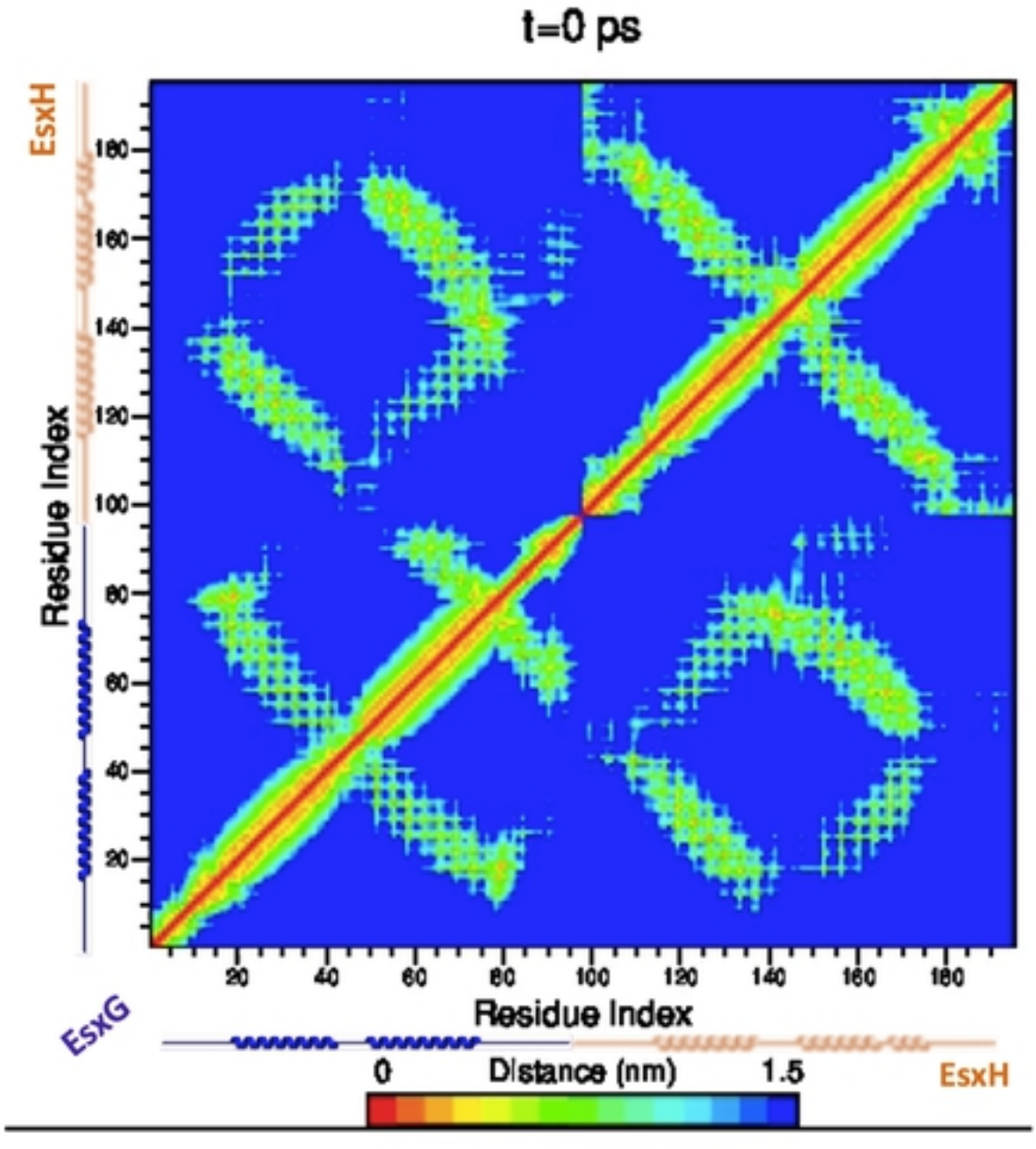


Figure2

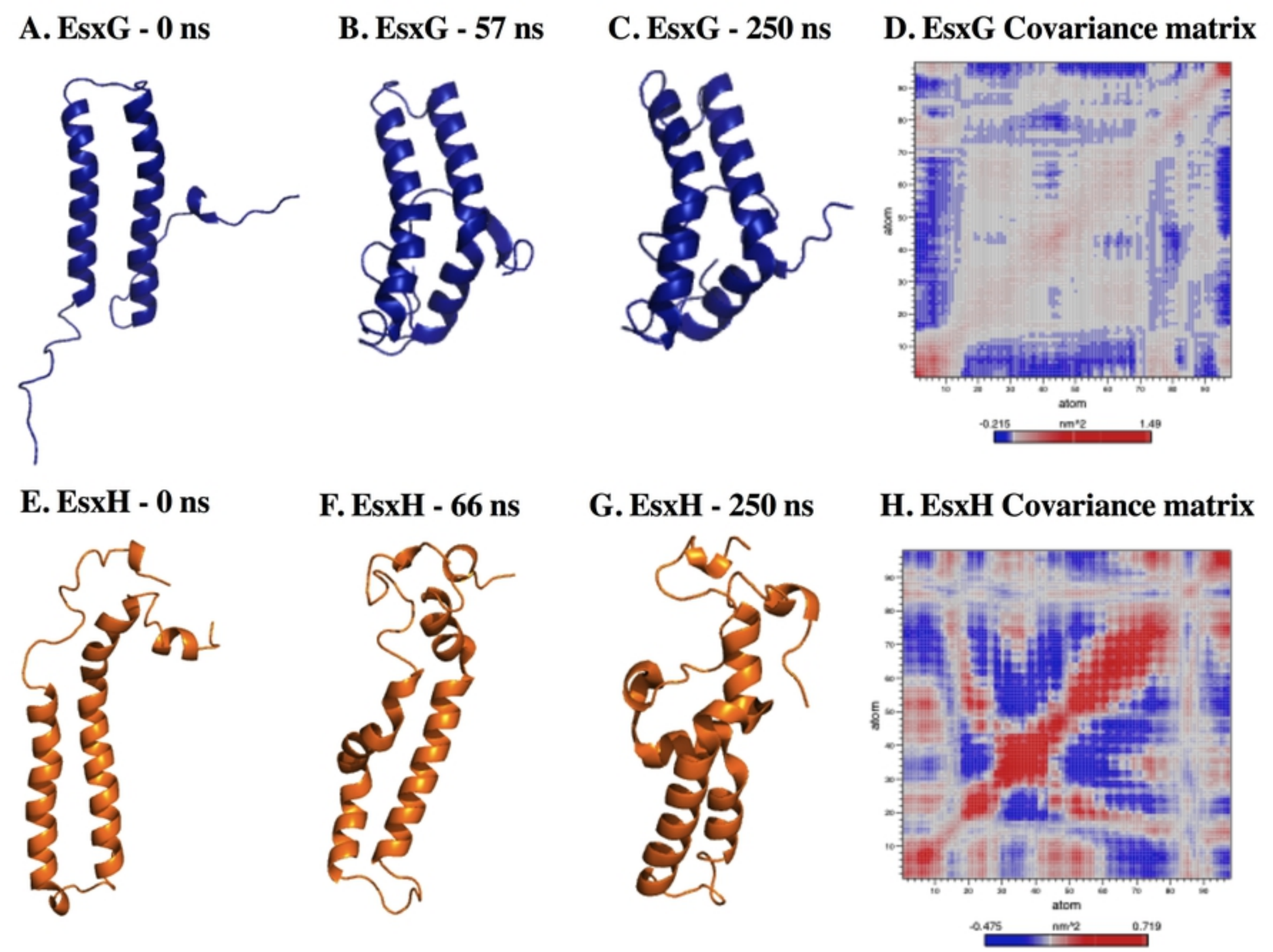
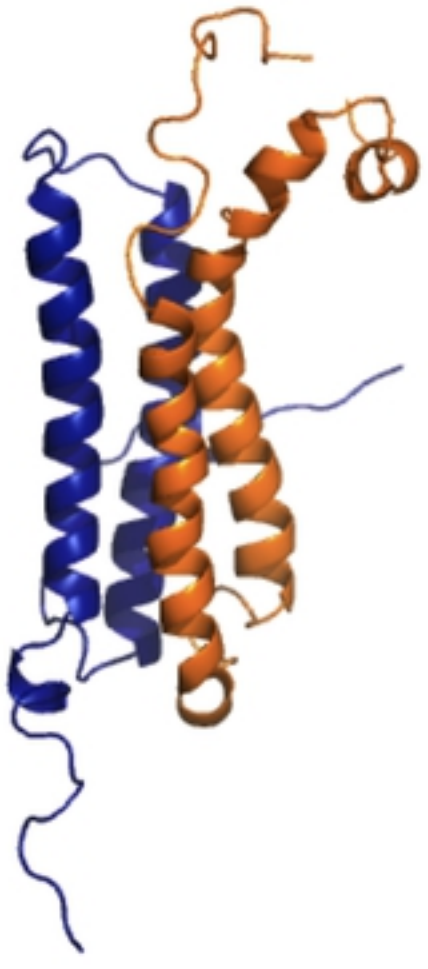

A. EsxG - 0 ns
B. EsxG - 57 ns
C. EsxG - 250 ns
D. EsxG Covariance matrix
atom
atom
-0.215
nm^2
1.49
E. EsxH - 0 ns
F. EsxH - 66 ns
G. EsxH - 250 ns
H. EsxH Covariance matrix
atom
atom
-0.475
nm^2
0.719

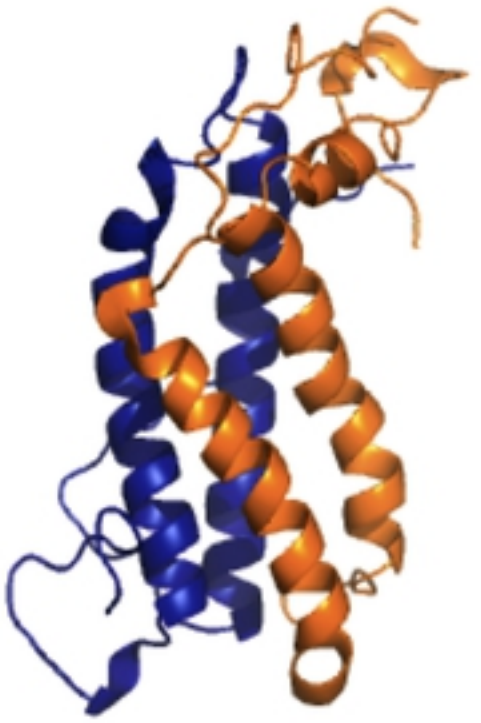
Figure3

A. 0 ns - 310 K

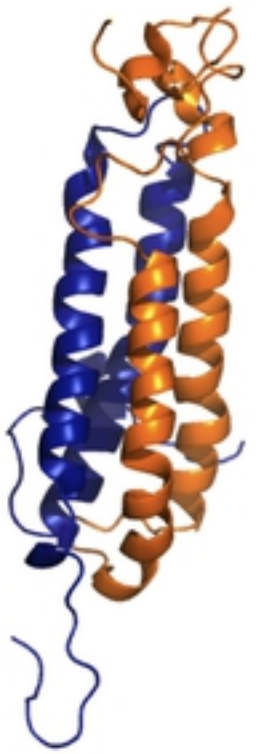


B. 5 ns - 350 K

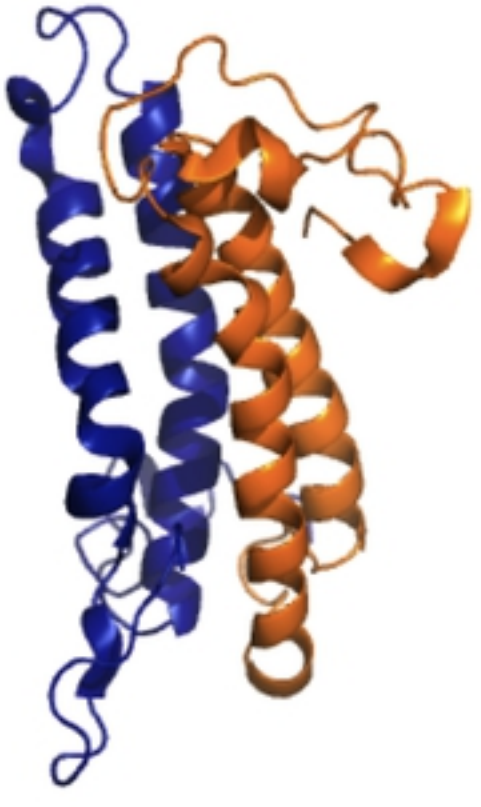
C. 50 ns - 350 K



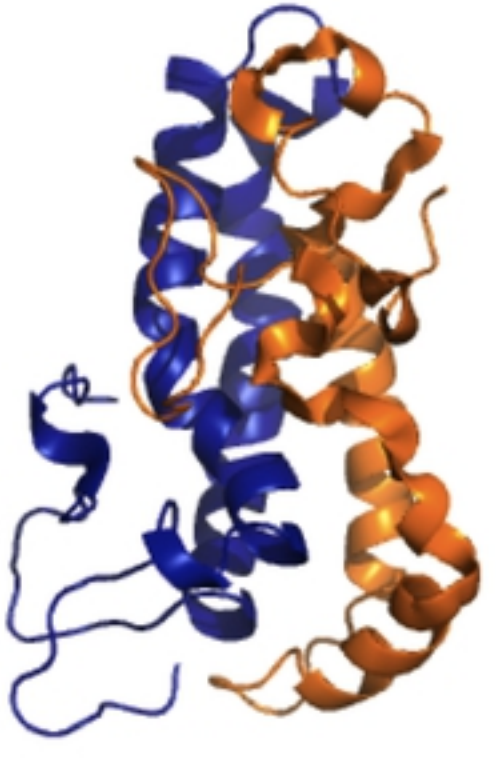
D. 5 ns - 400 K



E. 5 ns - 450 K



F. 5 ns - 500 K



G. 5 ns - 550 K

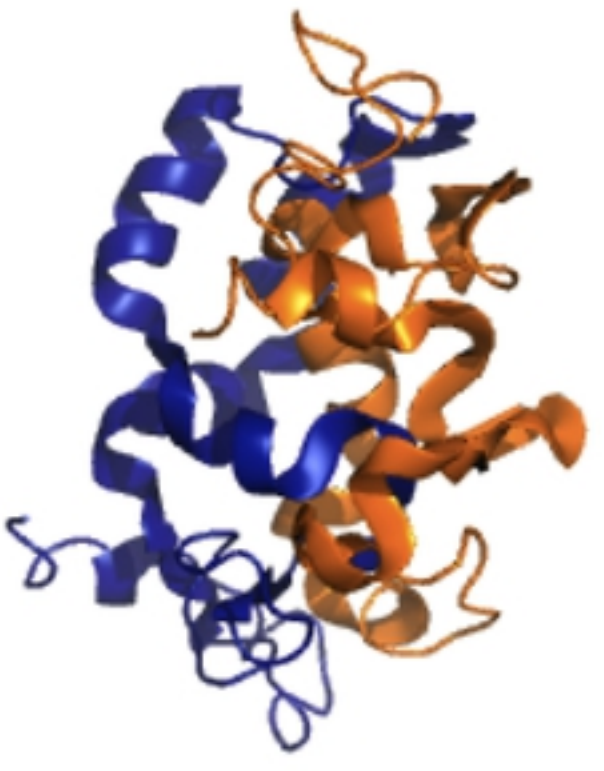
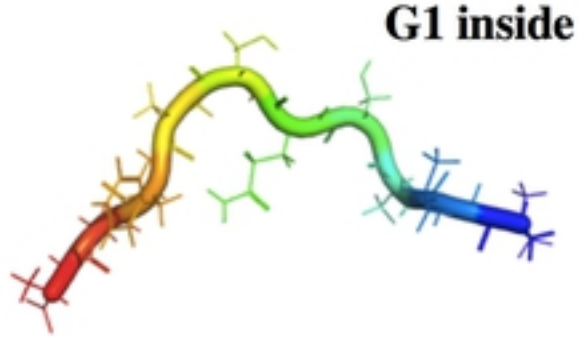
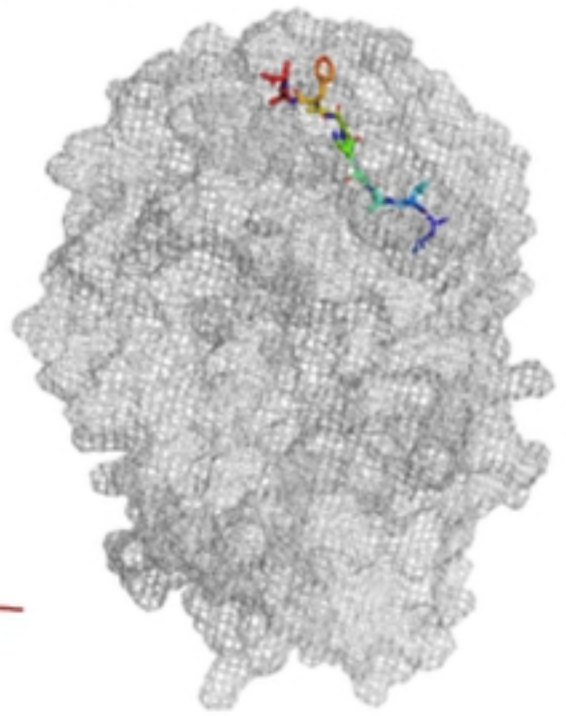


Figure4

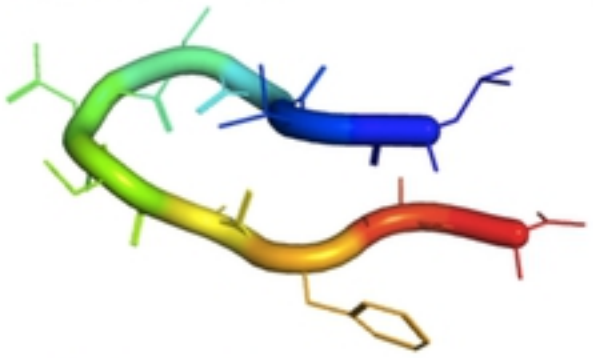
A. G1 epitope

G1 inside HLA-II



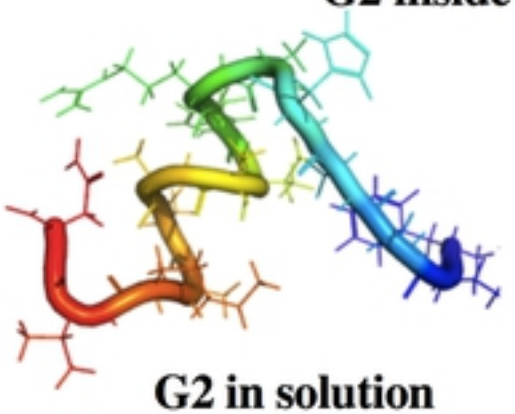


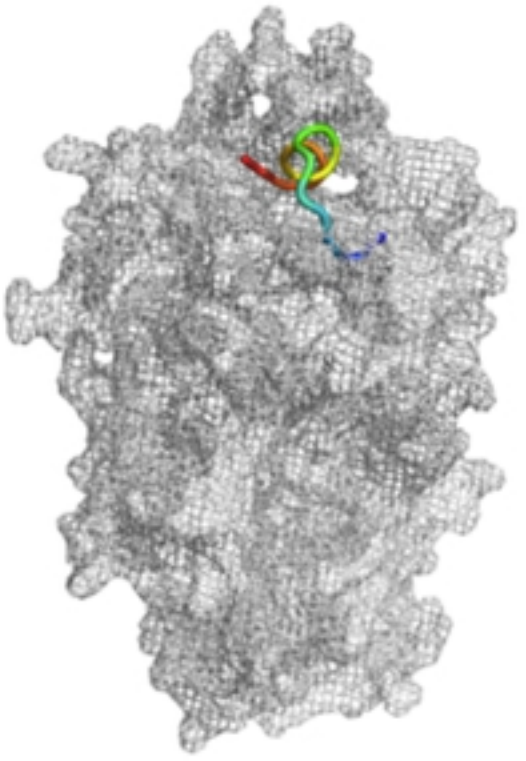
G1 in solution



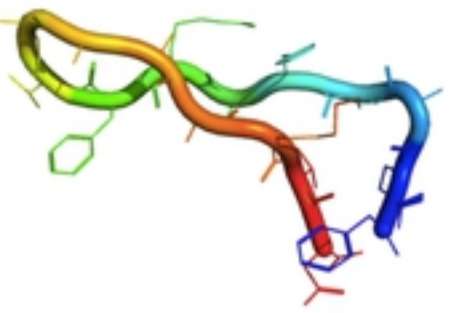
B. G2 epitope

G2 inside HLA-II



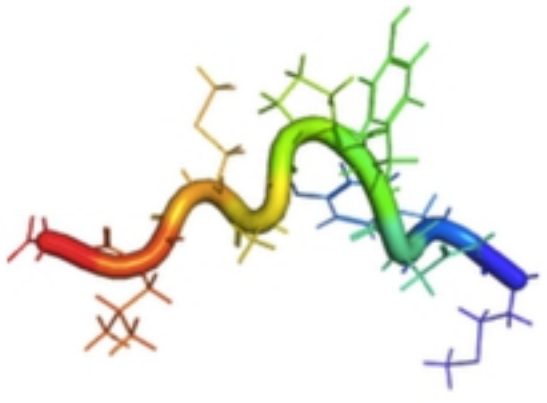


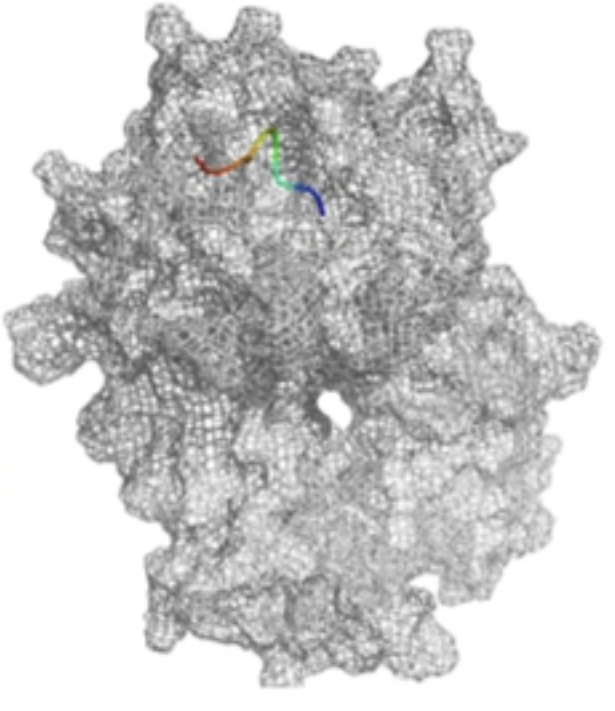
G2 in solution



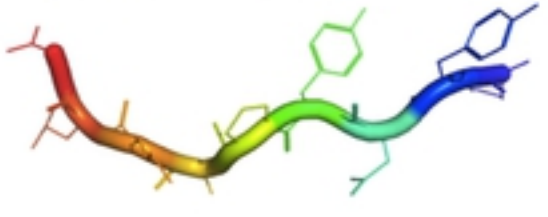
C. H1 epitope

H1 inside HLA-II



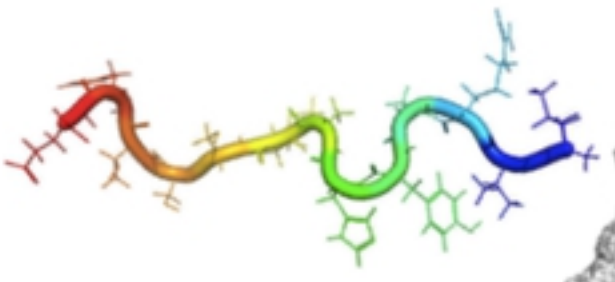


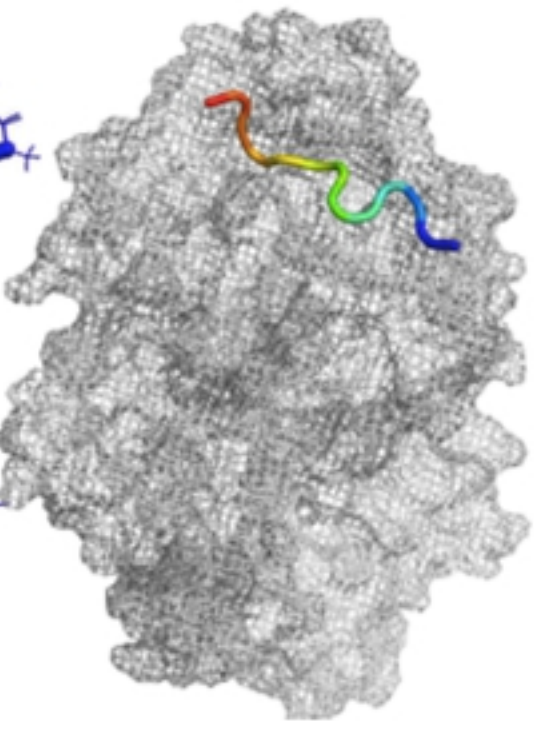
H1 in solution



D. H2 epitope

H2 inside HLA-II





H2 in solution

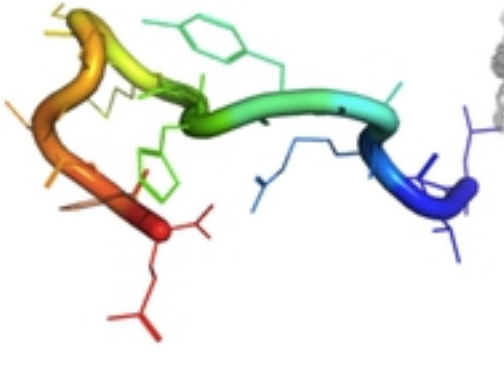


Figure5

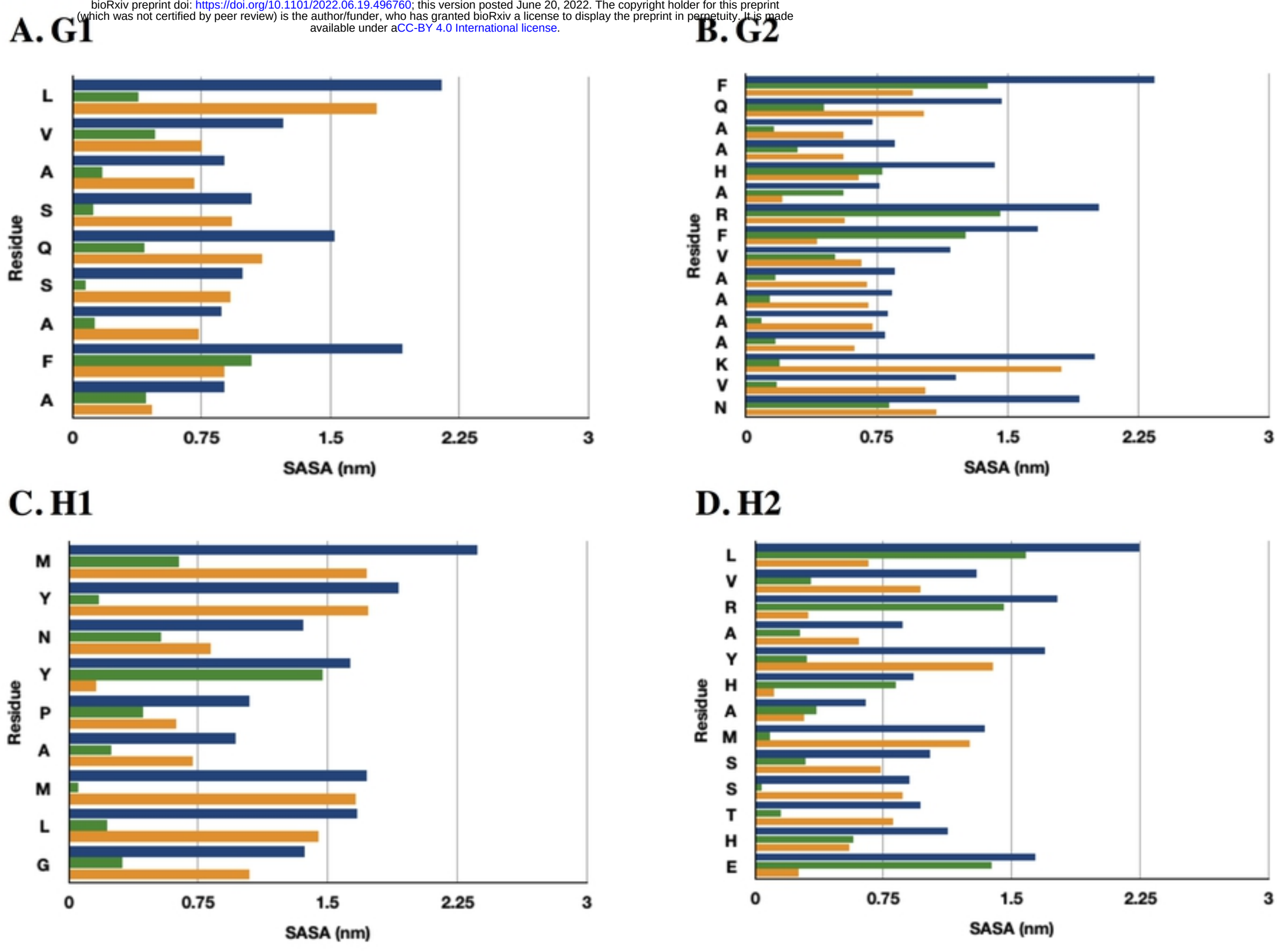


Figure6

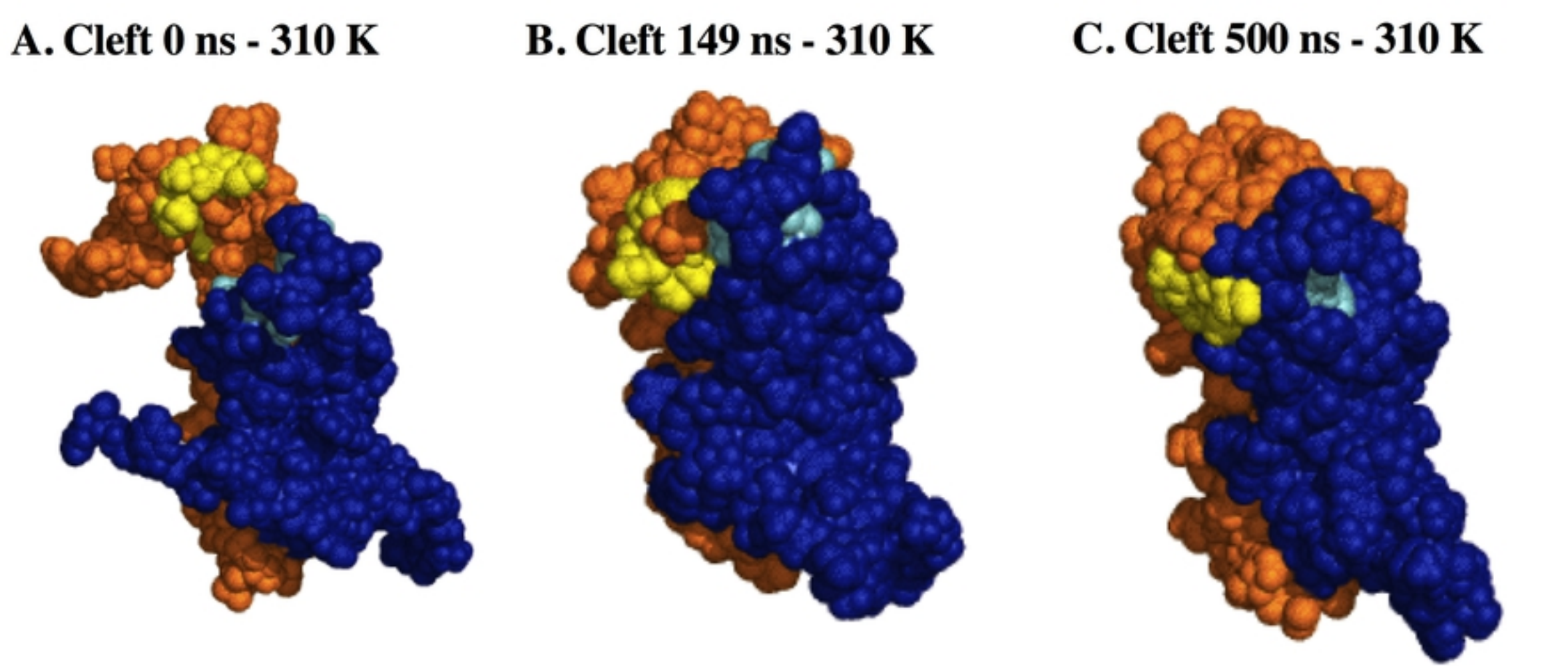


Figure7