PAPER

De novo design of immunoreactive conformation-specific HIV-1 epitopes based on Top7 scaffold†

Isabelle F. T. Viana,[abc] Thereza A. Soares,[a] Lucianna F. O. Lima,[c] Ernesto T. A. Marques,[b] Marco A. Krieger,[c] Rafael Dhalia*[d] and Roberto D. Lins*[a]

Epitope exposure strategies often fail due to high conformational instability, resulting in loss of native conformation and fast degradation. Protein scaffolding, where a structural element of an antigen protein is transplanted into a scaffold acceptor protein, has been employed as an alternative solution. However, while this approach aims to preserve the epitope structure, it frequently results in unstable chimeric scaffolds. To overcome this issue, we employed a novel computational approach to rationally engineer conformational antigens into a highly stable scaffold protein. This strategy is showcased to display conformational HIV-1 gp41-based epitopes in their native structure. From HIV-1 antigen sequence databases, we have identified short sequences with the most probable antibody-recognizable regions. These sequences were inserted into or replaced regions of the original Top7 protein with analogous secondary structure assignment. Molecular dynamics simulations were used to characterize the protein stability and structural dynamic of the chimeric proteins, leading to a selection of promising candidates whose immunogenic epitopes are suitably exposed for antibody recognition. These computer-designed recombinant proteins were produced in bacteria using codon optimized DNA sequences and their diagnostic performance was assessed by liquid microarray against a human cohort of 47 sera samples. Our results show that the Top7 protein is a suitable scaffold to provide the required structural stability to predetermined target shapes and sequences, allowing the potential use of the chimeric proteins as antigens for specific antibody recognition and/or stimulation of immune response.

Introduction

Detection of antigen and pathogen specific antibodies is among the most used approaches for clinical diagnostics. Antigen specific antibody detection is used in the diagnostic of cancers, autoimmune and infectious diseases and its application continues to expand.[1–4] In addition, antibodies are the most effective mechanism of protection provided by the adaptive immune system.[5,6] A critical, and many times limiting, aspect for eliciting highly specific antibodies for vaccines and for the detection of antigen specific antibodies is the structural transiency of certain conformational B-cell epitopes. In addition, important epitopes are often cryptic, frequently buried within the protein and only exposed when the virus binds to its receptor.[7] One solution to overcome these

limitations is the engineering of robust artificial structures to provide support for those unstable epitopes.[5,6,8–12] The availability of high-resolution crystal structures of biologically important epitopes and remarkable advances of computational techniques for protein design have enabled the rapid development of protein engineering technologies for generation of structurally stable protein chimeras capable of presenting selected epitopes and other structural determinants.[13] One thriving approach for the design of novel proteins is called protein scaffolding, where structural elements are transplanted from one protein antigen into a more robust scaffold acceptor protein.[8,14] It relies on the identification of scaffold proteins containing similar structures to the antigenic target, and then transplanting the desired structural motif into the scaffold protein. Therefore, the choice of an adequate scaffold is the critical step for the successful engineering of novel proteins, particularly for the design of conformation-specific immunogenic epitopes.[15,16]

Recently computational techniques have been employed to insert HIV-1 2F5 and 4E10 epitopes into heterologous proteins leading to novel epitope scaffold antigens able to elicit structure-specific immune responses with nanomolar affinity.[5,8] The X-ray structures obtained for these chimeric scaffolds attest to their ability to conformationally mimic the

aDepartment of Fundamental Chemistry, Federal University of Pernambuco, Recife, PE, 50740-560, Brazil. E-mail: roberto.lins@ufpe.br

bDepartment of Infectious Diseases and Microbiology, University of Pittsburg, 9022 BST3, Pittsburgh, PA 15261, USA

cCarlos Chagas Institute, FIOCRUZ, Curitiba, PR, 81350-010, Brazil

dAggeu Magalhães Research Center, Oswaldo Cruz Foundation, Recife, PE, 50670-420, Brazil. E-mail: rdhalia@cpqam.fiocruz.br

† Electronic supplementary information (ESI) available: List of HIV-1 envelope (gp160) amino acid sequences from Latin America viral isolates. See DOI: 10.1039/c3ra41562g

antigenic surface recognized by the respective target anti-bodies. As a test system, in this study, we expanded the small repertoire of conformation-specific epitope scaffold antigens directed to the HIV-1 virus using the synthetic protein Top7 as a structural platform.[17] Top7 is a small protein of 95 amino acids and remarkable chemical, thermal and mechanical stability.[17,18] We have previously shown that Top7 retains its overall structure at extreme conditions of temperature and/or pH due to the very stable anti-parallel β-sheet core.[19] These findings guided the design of a Top7-based scaffold that specifically bound the CD4 protein while remaining extremely resistant to thermal and chemical denaturation.[20] Our choice of Top7 as scaffold for conformation-specific epitopes relies on the assumption that less flexible structures are expected to better preserve the functional conformation of the grafted sequence. Hence, we have tested this assumption through a three-step protocol: i. selection of significantly exposed short sequences from well-characterized epitopes from Human Immunodeficiency Virus type 1 (HIV-1), ii. modification of the highly stable scaffold capable to present structure-based epitopes in their native conformation, and iii. evaluation of the immunological performance of the engineered imunogens.

The selected HIV-1 epitopes were inserted in three different sites of the Top7 protein for the display of linear and conformation-specific epitopes (Fig. 1). These sites corre-sponded to the loop region (residues 22 and 23) previously used to display an anti-CD4 peptide, and the two α-helices, since they were shown to play only a minor role in the stability of Top7.[19,20] The computationally devised chimeras were submitted to molecular dynamics simulations to determine its structure, and epitope and scaffold stability. All molecules had their corresponding DNA sequences optimized for expression in prokaryotic systems and were synthetically built using molecular biology methods. Computational predictions were assessed by immunological assays against a cohort of 21 HIV-1 negative and 26 HIV-1 positive human sera samples. Our findings indicate that computer designed protein chimeras containing the epitopes of HIV-1 gp41 protein grafted onto structurally compatible regions of Top7 that retained struc-tural features of the HIV-1 epitopes and maintained scaffold stability, can specifically bind to antibodies of HIV-1 positive individuals, discriminate between positive and negative HIV-1 human sera samples, and are potential candidates as vaccine and/or diagnostic antigens.

Results and discussion

Sequence selection

A bioinformatics approach was adopted to identify short-length putative HIV-1 gp41 epitopes. A consensus sequence for HIV-1 envelope (gp160) protein, which comprises the gp120 and gp41 proteins, was generated from sequence alignment of 64 primary viral isolates from Latin America. Following, a reverse search was performed to identify the primary viral isolate with the closest sequence match to the consensus
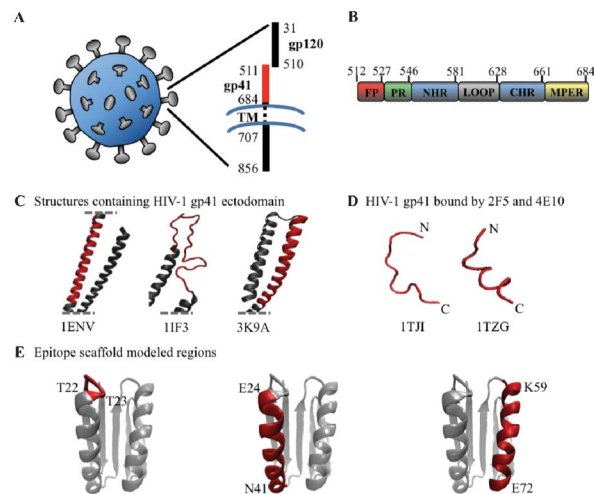

Fig. 1 Target epitopes and transplantation to Top7 acceptor scaffold. (A) gp41 ectodomain, depicted in red, which involves the residues 511 to 684 of gp160.[5] TM – transmembrane region. (B) Schematic overview of gp41 ectodomain. FP – fusion peptide; PR – polar region; NHR – N-terminal heptad repeat; connection loop; CHR – C-terminal heptad repeat; MPER – membrane proximal external region. (C) gp41 ectodomain structures previously determined. The regions highlighted in red represent, following from left to right, the NHR, Loop, CHR/MPER, which are identified by their correspondent PDB IDs.[21–23] (D) Crystallographic structures of 2F5 and 4E10-recognized conformations are identified by their PDB IDs.[24,25] (E) Schematic representation of modeled regions on the acceptor scaffold. From left to right: red highlighted regions correspond to the Top7 s loop (T22 and T23) and the two Top7 helices (E24 thru N41 and K59 thru E72), which are partly or fully replaced by the epitope aminoacid sequences. The scaffold is shown in cartoon representation, with the scaffold backbone colored in gray and the modeled regions in red. (Protein structures are represented as a cartoon model in C–E).

sequence. The gp41 ectodomain sequence was further identified, subjected to a Kyte–Doolittle hydropathy analysis[26] and the secondary structure of the selected hydrophilic epitopes assigned by their positions in the gp41 crystal structure. (Details can be found in Methods section). The putative epitopes were comprised of sequences from 8 to 18 residues long and are listed in Table 1.

Structural dynamics and stability

Molecular dynamics (MD) simulations were performed for all Top7-HIV-1 chimeras to ascertain that these variants kept both the native-like HIV-1 epitope conformations and the Top7 fold after insertion/substitution of the HIV-1 sequences in the scaffold. The conformational ensembles of the engineered structures were compared to the original Top7 after the equilibration period. The time-dependent behavior of physi-cochemical properties extracted from the molecular dynamics simulations was similar for most of the simulated systems, reaching convergence after 2–3 ns. As a criterion of structural stability, the atom positional root-mean-square deviation (RMSD) of all Cα atoms was calculated as a function of time over each 50 ns trajectory. The time evolution of the RMSD provides a measurement of the convergence of the structural dynamics of the proteins, and how the structural ensemble of sampled conformations differs from a given reference struc-

Table 1 Selected epitope sequences and their corresponding secondary structures

Epitope ID	Sequence	Position in gp41 HIV-1	Position in Top7	Motif	Replaced residues in Top7
E2F5[a]	ELDKWASLW	663–670	22–30	Loop	—
E4E10[a]	WNWFDITNW	671–679	22–30		
R5L	AVERYLKDQQ	583–592	22–31		
R2HA	GIVQQQNNL	548–556	33–41	Helix	ELMDYIKKQ
R2HB			60–68		KFAAILIKV
R3HA	AIEAQQHL	559–566	27–34		LQKVLNEL
R3HB			62–69		AAILIKVF
R8HA	RLIEESQNQQEKNEQELL	645–662	24–41		ESELQKVLNELMDYIKK
R8HB			55–72		KKQAQKFAAILIKVFAQL

[a] E2F5 and E4E10 sequences correspond to described epitopes 2F5 (HIV-1 HxB2 residues 662 to 671) and 4E10 (HIV-1 HxB2 residues 670 to 678), respectively.

ture. Each frame of the trajectory was superimposed onto the reference frame (energy optimized chimera structures) using the least-square fit method by removal of periodicity effects, rotational and translational motions were removed. The RMSD for the original Top7 system converged after 2 ns to *ca.* 0.18 nm (Fig. 2). The RMSD of the loop insert Top7_E4E10 presented a similar behavior to the original Top7 protein, while Top7_E2F5 and Top7_R5L showed a higher RMSD along with significant fluctuations throughout the trajectory reaching final average values above 0.24 nm (Fig. 2A). This is more pronounced for Top7_E2F5 suggesting that this system undergoes substantial conformational changes. The time evolution RMSD for this chimera shows a curve that shifts through two short-lived intermediate states (one around 0.2 nm and another around 0.25 nm) without a single well-defined conformational state. This dynamic polymorphism of E2F5 (also know as 2F5 epitope) is consistent with the experimentally observed conformational transition the epitope under-

goes between unstructured and an α-helical-like motif,[27] related to the role of the gp41 MPER in the virus cellular entry.

The RMSD for the proteins containing the HIV-1 sequence inserts in the helices Top7_R2HA, Top7_R2HB, Top7_R3HA and Top7_R3HB (8 and 9 residues for R2 and R3, respectively) converged to values around 0.2 nm at 2.5 ns (Fig. 2B). In contrast, the Top7_R8HA and Top7_R8HB systems deviated significantly from their respective reference structures compared to the other helical-modified chimeras. For these systems, the RMSD values fluctuated around 0.3 nm for the first 8 ns. In the remaining time span, the RMSD exhibited a rapid increase, rising to a plateau characterized by values near 0.4 nm. It should be stressed that helices in Top7_R8HA and Top7_R8HB have been completely replaced by HIV-1 sequence inserts of 18 residues each. All proteins presented a stable behavior after about 25 ns of simulation.

In summary, Top7_R5L, Top7_R8HA and Top7_R8HB presented the largest deviations, followed by Top7_R2HB, which exhibited also increased fluctuations, with a slightly but consistently higher RMSD value compared to the other designed proteins. Five out of the nine chimeras, namely Top7_E4E10, Top7_E2F5, Top7_R2HA, Top7_R3HA and Top7_R3HB were shown to be structurally stable throughout the simulations when compared to native Top7.

The residue averaged positional root-mean-square fluctuation (RMSF) was calculated for all systems along the 50 ns time length (Fig. 3). The three loop insert systems exhibit an overall profile analogous to the original Top7 (Fig. 3A). As expected, the regions of highest atomic fluctuations in all simulated systems correspond to the loop insert motifs. A sharp decrease in the RMSF flanked by the loop insert can be seen for these chimeras. Interestingly, all loop inserts seem to slightly destabilize the end of the first helix (Fig. 3A). The RMSF behavior for the helical insert systems Top7_R2HA, Top7_R2HB, Top7_R3HA and Top7_R3HB was very similar to the original Top7. In these systems, the highest atomic fluctuations occurred in the connecting loops whereas the lowest atomic fluctuations were localized in the secondary structure motifs, notably in the sequence-replaced helix (Fig. 3B). Conversely, the Top7_R8HA and Top7_R8HB systems exhibited a significant increase in atomic displacement along
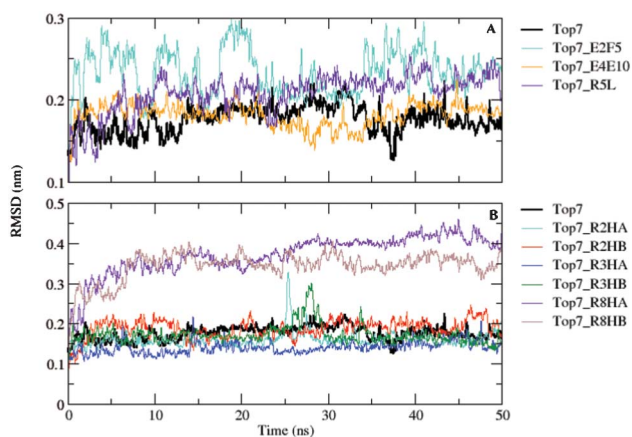

Fig. 2 Root mean square deviation (RMSD) for the backbone atoms of the original (black lines) and Top7 chimerical proteins (colored lines) as function of time. Rotational and translational fitting of pairs of structures was applied using all α-carbon atoms, except for residues corresponding to the loop/helices whose sequence was inserted in the respective chimeras. Sampling time window was 1 ps. (A) Loop-insert systems. (B) Helix-insert systems.
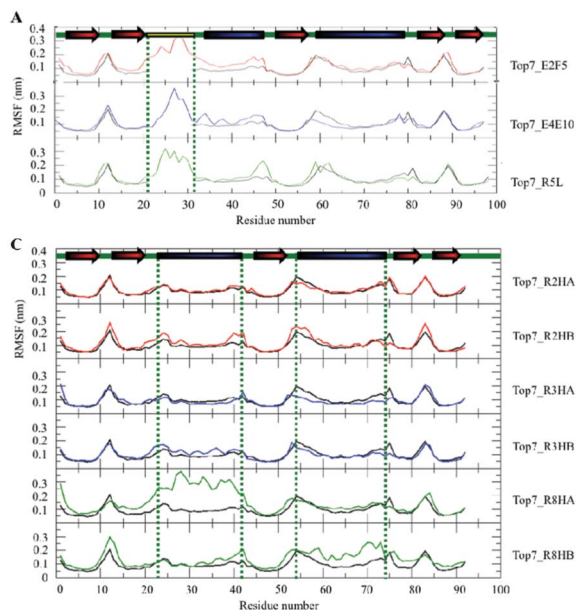
Fig. 3 Root mean square fluctuation (RMSF) time series of backbone atoms averaged for each residue of the original (black lines) and Top7 chimerical proteins (colored lines). Epitope inserts (addition or replacement) are delimited by a pair of green dashed lines (upper panel: loop insert systems; lower panel: helices insert systems).

regardless of the sequence inserts or substitutions. This finding is consistent with the previous observation that the β-sheet in Top7 comprises a stable core, which is responsible for its stability when it is subjected to changes in pH and temperature.[28,29] The average number of hydrogen bonds does not change significantly, except for Top7_R5L, Top7_R8A and Top7_R8B where there is a decrease of *ca.* 8% or more in these values.

The RMSD, RMSF and SASA analyses indicate that the global fold of the Top7 scaffold is not adversely affected by the HIV-1 sequence insertions, except for three inserts. The structural ensemble corresponding to inserts Top7_E2F5, Top_R8HA and Top_R8HB diverged from the initially modeled structures. Our hypothesis is that the loss of structural content and increased flexibility is expected to affect the scaffold and/or epitope conformational stability, limiting the antibody recognition aptness of these chimerical proteins.

To further characterize the conformational stability of the inserts and scaffold, time-dependent profiles of secondary structures for original and chimerical Top7 proteins along 50 ns of simulations were calculated with the DSSP software (Database of Secondary Structure in Proteins)[30] (Fig. 4). The secondary structure profiles revealed that the modeled loops cause a modest disturbance to the overall backbone structure in loop insert systems (Fig. 4B). Sequence inserts into the chimeras Top7_E2F5 and Top7_E4E10 affected the stability of the first helix of the protein (residues 28–43), while Top7_R5L noticeably reduces the stability of both helices. All three simulations exhibit secondary structure fluctuations in the region corresponding to the N-terminal residues of the fourth Top7 β-sheet (residues 78–84). Comparison of the six helix-replaced systems as function of time-dependent secondary structure changes shows that Top7 scaffold remains very robust for the systems R2H and R3H. These proteins maintain the Top7-original structure regardless of partial or total helix sequence substitutions. Although, the scaffold conformation was particularly conserved upon insertion of the epitope R2H and R3H, only the helical conformation of the modeled Top7_R2HA, Top7_R3HA and Top7_R3HB epitopes were entirely maintained (Fig. 4C and 5). The secondary structure time series for these systems are analogous to the original Top7. In contrast, a partial disruption of the helical structure

the helical replacement regions, indicative of the inability of these variants to maintain a stable helix bundle.

The averaged hydrophobic solvent-accessible surface area (SASA) did not show any significant change (Table 2). It suggests that all engineered proteins, either HIV-1 loop inserts or substitutions of both Top7 helices, do not exhibit large alteration of solubility compared to native Top7. The number of hydrogen bonds and residues in secondary structural motifs averaged over the last 25 ns is summarized in Table 2. While the overall average number of residues per secondary structure in the chimerical proteins remains reasonably close to the original Top7, a decrease in helical content can be observed for Top7_R5L, Top7_R8A and Top7_R8B. Interestingly, the total number of residues in the β-sheet is maintained

Table 2 Average[a] structural properties for Top7 and chimerical proteins obtained from molecular dynamics simulations

| Epitope ID | Average number of residues per secondary structure | | | SASA (nm²) | Average number of hydrogen bonds |
	Helix	Sheet	Turn + Coil		
Top7	34.7 ± 0.7	37.3 ± 1.8	19.9 ± 2.0	82.04	82.3 ± 4.0
E2F5	31.7 ± 0.6	36.5 ± 1.8	28.7 ± 1.9	86.85	78.9 ± 4.0
E4E10	30.0 ± 1.7	35.2 ± 1.9	32.0 ± 2.9	87.22	81.8 ± 5.6
R5L	29.1 ± 1.6	36.4 ± 2.0	32.5 ± 2.4	89.09	73.9 ± 3.8
R2HA	35.6 ± 0.7	36.8 ± 1.7	19.6 ± 1.8	80.13	82.0 ± 4.1
R2HB	33.8 ± 1.4	38.6 ± 2.0	20.6 ± 2.6	78.69	83.4 ± 4.2
R3HA	33.7 ± 0.6	34.8 ± 2.0	23.5 ± 2.1	79.41	81.2 ± 4.5
R3HB	35.5 ± 0.8	37.3 ± 1.6	19.1 ± 1.8	79.47	84.1 ± 4.0
R8HA	17.8 ± 0.7	35.4 ± 1.5	38.8 ± 1.6	83.77	76.9 ± 5.2
R8HB	26.9 ± 2.3	35.1 ± 1.8	30.0 ± 3.1	83.17	81.8 ± 4.9

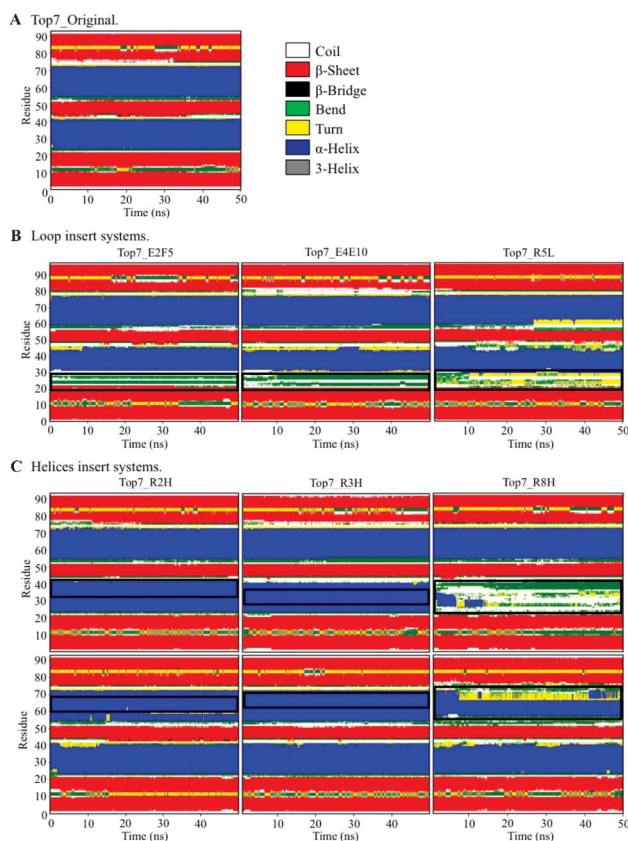[a] Values are averaged over the last 25 ns of the respective simulation.

Fig. 4 Time-dependent secondary structure content. The secondary structure is plotted *versus* simulation time for original (A) and chimerical Top7s (B–C). In figure C, the upper panels show the secondary structure time evolution for chimeras that feature residue substitution of their helix A, while the lower panels displays the corresponding quantity for helix B-modified chimeras. Color patterns are blue for α-helix, red for β-sheet, green for bend, yellow for turn, and white for coil. Secondary structure definition according to Kabsch and Sander.[31] The position of the inserted/replaced residues are approximately encompassed by a black square in each system.

of the modeled R2HB epitope takes place at about 26 ns as shown for the unfold of five epitope residues (Fig. 4C). A more dramatic change in secondary structure is observed when the Top7 helices were substituted by the HIV-1 R8H sequences (Top7_R8HA and Top7_R8HB). The helices motif unfolded within a few nanoseconds (Fig. 4C and 5).

Regardless of the modifications made to the Top7 structure due to HIV-1 sequence insertions, the time-dependent secondary structure profiles show that the hydrophobic central core maintains the β-sheet structure throughout the entire 50 ns simulation. These results are, once again, consistent with previous studies of a Top7-based scaffold for CD4 recognition, which showed remarkable structural stability at extreme conditions of pH and temperature due the maintenance of the β-sheet structure. It is also noticeable that the chimerical proteins retained most of the helical structural elements as the simulations proceeded (Table 2).

According to the outlined hypothesis and taking into account the calculated structural properties derived from the molecular dynamics simulations, Top7_E4E10, Top7_E2F5
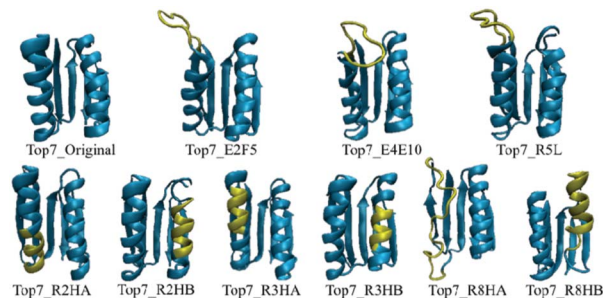

Fig. 5 Cartoon representation of structures of the Top7-HIV-1 chimeras after 50 ns of molecular dynamics simulation represented by their secondary structure content. Epitope regions are color coded in yellow and the original scaffold regions in cyan.

(loop inserts), Top7_R2HA, Top7_R3HA and Top7_R3HB (helix substitutions) constitute promising candidates to expose their epitopes in a conformationally stable and native-like fashion. Top7_R2HB has shown good folding stability, but the increased flexibility and partial unfolding of its native conformation may result in unspecific binding or inability of being recognized by HIV-1 antibodies. To test this hypothesis, all proteins were synthesized in bacteria. Genetic DNA sequences were engineered, cloned into plasmid vectors, expressed, purified and their ability to be recognized by antibodies in patient sera tested.

Design, production and diagnostic performance evaluation of the Top7-HIV-1 chimerical proteins

The DNA sequences coding for Top7-HIV-1 chimerical proteins were optimized for improved bacterial expression and commercially synthesized. The optimized genes were inserted in the expression vector and the resultant constructs were transformed into *E. coli* host cells. The expression procedure was optimized by testing different temperatures and induction periods (2 to 5 h). Protein accumulation increased up to 3 h and stayed constant thereafter. Purification of chimerical proteins was performed from cell lysates in non-denaturing conditions to keep the proteins in native conformation by unique affinity chromatography step. The produced proteins showed high expression levels (0.5 to 16.7 mg per liter of culture) and purity level of 95%. Except for Top7_R5L and Top7_R2HA, where working stocks were kept at near 0 °C (ice bath) in a cold room to avoid protein precipitation, all purified proteins presented remarkable stability during the experimental manipulations, which are probably due to maintenance of the native folding of the epitope–scaffold pair.

To assess the reactivity of the HIV-1 structures with anti-HIV antibodies the produced chimeric proteins were tested against 26 HIV-1 positive and 21 HIV-1 negative human serum samples in a cytometric bead array to capture anti-HIV-1 antibodies. The experimental conditions were set using different concentrations of each chimerical protein (25 to 100 μg mL^{-1}) against two pools composed by the positive or negative samples serially diluted, and ranging from 1 : 50 to 1 : 400. Simultaneously, serial dilutions (1 : 100 to 1 : 800) of the secondary antibody were also assayed (data not shown for

conciseness). The optimum antigen concentration, primary and secondary antibodies dilutions were determined according to the discrimination degree between negative and positive populations, and were established at 50 µg mL^{-1}, 1 : 100 and 1 : 200, respectively. These conditions were used in all subsequent experiments. These results indicate the potential of these chimeric proteins to be used as antigens in HIV diagnostic assays.

Following the experimental setup, the Top7_HIV-1 proteins were tested against the individual serum samples. Fig. 6 shows the median fluorescence intensity for all proteins and the results are presented as an average of individual measurements for each serum along with the corresponding standard deviation. Fig. 6A shows the immunological behavior of the chimeras considered conformationally stable and presenting remarkable maintenance of the epitope–scaffold structural properties (Top7_E2F5, Top7_E4E10, Top7_R2HA, Top7_R3HA and Top7_R3HB). Four of these chimerical proteins were shown to have good diagnostic performance, since they were strongly reactive against the tested serum samples and exhibited a clear and significant distinction between HIV-1 positive and negative populations (p < 0.05). This result corroborates the computational predictions, establishing a direct relation between maintenance of the overall structure of

the epitope–scaffold pair and maintenance of the devised immunological properties. Even though the Top7_E2F5 has demonstrated remarkable stability, this design showed a poor reactivity with serum of HIV-1 infected patients. The MFI values obtained for the HIV-1 positive patients were nearly identical to those obtained for non-infected patients (p = 0.7973), which is consistent with the low incidence of naturally generated antibodies against this epitope during natural HIV-1 infection.[16,32,33] The E2F5 epitope, recognized by pan-neutralizing 2F5 monoclonal antibody (mAb), does not normally elicit high titers of antibodies in infected individuals. However, since antibodies against this highly conserved epitope structure can neutralize a wide range of HIV-1 isolates it is considered one of the ideal targets for vaccine development. In order to ascertain the conformational stability predicted by the computational model, the Top7_E2F5 protein was tested against the monoclonal antibody specific for the 2F5 epitope by means of ELISA assay (Fig. 7). The original Top7 protein was probed to determine the background noise. The results provide strong support to the stability prediction by molecular dynamics simulations. In this procedure, Top7_E2F5 was specifically recognized by the respective monoclonal antibody and the titer for this chimerical protein was about 10^9 fold greater than the background signal detected for the original Top7 protein, indicating that this chimera expresses the structure recognized by the 2F5 mAb.

The immunological behavior of the chimeric proteins that did not present secondary structure stability in the computational analysis (Top7_R5L, Top7_R2HB, Top7_R8HA and Top7_R8HB) is shown on Fig. 6B. These proteins exhibited substantial fluorescence signal fluctuations. The difference between the MFI values obtained for HIV-1 positive and negative sera were only significant (p < 0.05) for Top7_R5L and Top7_R8HA. However, Top7_R8HA was not able to react with serum of HIV-1 positive populations and Top7_R5L presented an unacceptable false positive ratio. In addition, Top7_R5L and Top7_R2HB showed increased flexibility, which can lead to various new structural motifs and therefore to
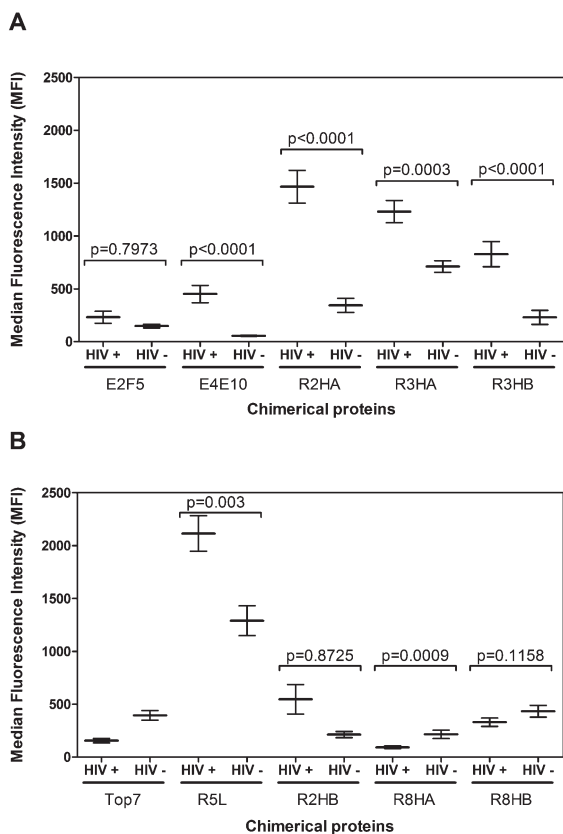

A

B

Fig. 6 Median fluorescence intensity for Top7-HIV-1 chimeras against HIV-1 positive and negative sera samples. The results are shown as mean values over all measurements and standard deviations. The p-values of two-tailed t-tests are indicated above each analysis. A) Conformationally stable chimeras. B) Conformationally unstable chimeras.
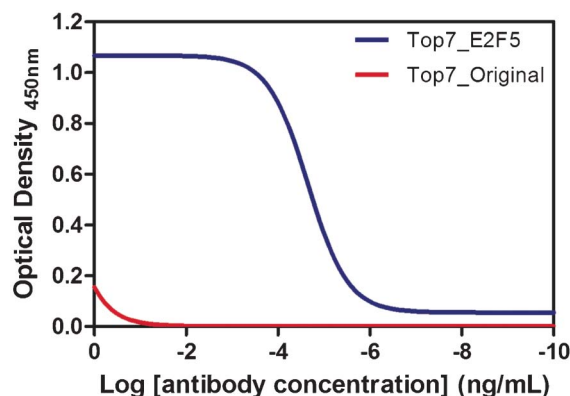

Fig. 7 Top7_E2F5 recognition by monoclonal antibody 2F5 through ELISA assay. The original Top7 protein was used to provide the unspecific binding measurement. Both proteins were tested against a series of 10-fold dilution of the 2F5 mAb starting at 1 ng mL^{-1}. Measured optical densities are shown vs. log of the 2F5 antibody concentration.

unspecific binding. On the other hand, Top7_R8HA and Top7_R8HB proteins reacted very weakly with the human serum samples and presented the poorest performance in the immunological assay, characterized by low levels of fluorescence intensity which are comparable with those obtained for Top7_Original. Such inability to distinguish HIV-1 infection is supported by the dramatic secondary structure loss predicted for both proteins in the molecular dynamics simulations.

Receiver operating characteristics (ROC) analysis was performed to assess the diagnostic performance of those proteins that are stable and correctly differentiate positive from negative HIV-1 serum samples. The fluorescence signal related to the original Top7 protein has been considered as reaction background and was subtracted from the median fluorescence intensities measured for the chimerical proteins. A ROC curve was generated for each protein by plotting the true-positive rate (sensitivity) against the false-positive rate (specificity) and provided the estimated values for accuracy of each test. The corresponding ROC curves are provided in Fig. 8. Area under the curve (AUC) was obtained in each case as an estimate of the probability that the protein being tested correctly classifies a HIV-1 positive or negative sample and, therefore, indexes the discriminating power of the protein. The AUC for all four curves was close to 1 (ranging from 0.89 to 0.96) indicating a high degree of cases correctly classified by the protein being tested. The AUC, sensitivity and specificity values are shown in Table 3.

Although Top7_R2HB presented significant variations, the overall scaffold structure was kept. This finding suggests that efficient discrimination between positive and negative HIV-1 sera requires not only maintenance of the scaffold structure, but that the displayed epitope is presented in a highly stable fashion. This hypothesis is supported by the relatively low average b-factor values for the α-carbons epitopes regions of the five available crystal structures containing the different epitopes (PDB IDs: 1ENV,[21] 3K9A,[23] 1IF3,[22] 1TJI,[24] 1TZG[25]), which range from 29.7 to 36.9 Å^2.
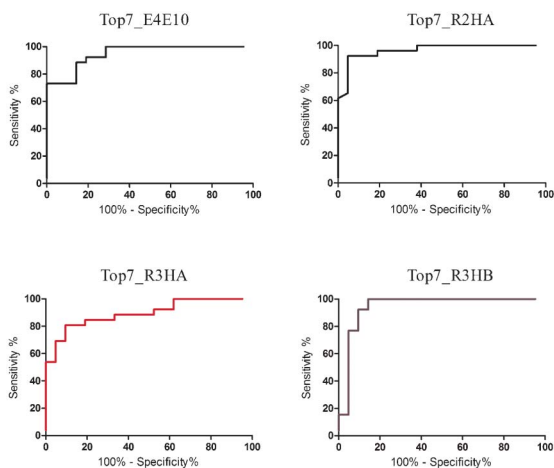

Fig. 8 ROC curve analysis of the proteins Top7_E4E10, Top7_R2HA, Top7_R3HA and Top7_R3HB. The paired results for sensitivity and specificity of each protein were plotted as points in a ROC space and the trade-off between these measures for different discrimination cutoffs are graphically showed.

Experiment

Sequence selection

64 amino acid sequences from Latin America viral isolates encoding the HIV-1 envelope protein (gp160) were identified in the National Center for Biotechnology Information – NCBI (Supplemental† Table 1). The sequences were aligned using the Basic Local Alignment Search Tool (BLAST) provided by NCBI[35,36] to generate the consensus sequence. A reverse search was then carried out to identify a primary viral isolate with the closest sequence match to the consensus sequence (access number: AADO3179). The correspondent region to gp41 protein ectodomain was identified (residues 512 to 684) and submitted to hydrophobic analysis by the Kyte–Doolittle scale[26] in order to identify the most hydrophilic regions, using a seven residues window. This process resulted in the final selection of six highly hydrophilic epitope sequences (including the sequences of the two well known 2F5 and 4E10 epitopes), ranging in length from 8 to 18 residues (the selected sequences are listed in the Table 1). Following, the Protein Data Bank (PDB) was searched for the gp41 ectodomain and the secondary structures identified (see Fig. 1C).

Computational modeling of epitope scaffolds

In the design stage, the six identified sequences were placed on the Top7 scaffold according to their secondary structure assignment observed in the PDB and solvent accessibility in Top7. Initial Top7 coordinates were taken from the crystallographic structure of PDB ID 1QYS with 2.5 Å resolution.[17] The residues from T22 to T23 of Top7 s loop were removed and the epitope sequences in loop conformations (sequences 1 to 3) were added into the scaffold. On the other hand, the two Top7 helices (E24 thru Q41 and K59 thru E72) were fully or partly replaced by the identified epitope sequences in helical motifs (sequences 4 to 6 in Table 1). On average, nine residues were transplanted to the Top7 s loop, while sequences containing 8, 9 or 18 residues were transplanted to the Top7 helix A and B (Fig. 1E). All the systems modeling were performed using the Swiss-PDB Viewer 4.0.1 software.[37]

Systems setup

The systems were embedded in cubic boxes, treated for periodic boundary conditions and solvated with explicit SPC model water molecules.[38] Counter ions were added, when necessary, to keep the overall systems neutral. A 1.0 nm short-range cutoff was used for all non-bonded interactions. Long-range electrostatic interactions were treated by the PME method,[39] with a grid size of 0.16 nm. No correction was applied for the long range van der Waals interactions. A description of the simulated systems is summarized in Table 4.

Molecular dynamics procedure

The simulation systems were equilibrated in a stepwise fashion in order to eliminate bad atomic contacts in a gradual way. The solvent and the inserted regions were energy-minimized, keeping the backbone atoms fixed, using 500 steps of steepest descent, followed by minimization of the whole system (solute and solvent) by an additional 500 steps

Table 3 ROC curve analysis data for chimerical Top7 proteins

Epitope ID	Sensitivity (%)	Specificity (%)	AUC[a]	Standard error[b]	95% confidence interval[c]
E4E10	100	72	0.95	0.02824	0.893 to 1.004
R2HA	96	81	0.96	0.02408	0.917 to 1.011
R3HA	89	67	0.89	0.04580	0.804 to 0.984
R3HB	100	86	0.95	0.04118	0.864 to 1.026

[a] Area under the curve, which represents the overall ability of the test to discriminate between positive and negative samples. [b] As calculated by DeLong et al.[34] [c] Binomial exact.

without solute constraints. Both equilibration steps were carried out by performing molecular dynamics (MD) for 50 ps at 300 K using a 2 fs time step. After the equilibration process, data acquisition was carried out for 50 ns at 300 K, and the configurations of the system were recorded as trajectory files at every 1.0 ps. All simulations were performed using the software package GROMACS v.4.0[40–43] in conjunction with the GROMOS force field parameter set 53A6[44] molecular dynamics simulations were carried out for the isothermal–isobaric (NPT) ensemble with a time step of 2 fs during the equilibration and production runs. The temperatures of solute and solvent were separately coupled to a Berendsen thermostat[45] with a relaxation time of 0.2 ps. The pressure was maintained at 1.025×10^5 Pa using weak coupling to a pressure bath[45] by means of isotropic coordinate scaling with a relaxation time $\tau = 0.4$ ps and compressibility of 4.5×10^{-5} $(kJ\ mol^{-1}\ nm^{-3})^{-1}$. Bonds involving hydrogen atoms were constrained by the LINCS algorithm[46] and a time step of 2 fs was used to integrate the equations of motion based on the leapfrog algorithm.[47]

Structural and electrostatic analysis

Calculated properties such as atomic positional root-mean-square deviation (RMSD) and fluctuation (RMSF), time-dependent secondary structure analysis, hydrogen bond content and solvent-accessible surface area (SASA) were computed using the implemented algorithms in the GROMACS package. The VMD 1.86 software was used for visual analyses of the molecular trajectories and electrostatic maps.[48]

Cloning strategy and genetic sequence optimization

The DNA sequences encoding the chimerical proteins were flanked by two multiple cloning sites (5′ Hind III and Nhe I and 3′ Nco I and Xho I) and optimized for enhanced mRNA transcription, RNA stability and translation in bacterial system. The criterion for optimization was balanced using the genetic algorithm from LETO 1.0 software (Entelechon®). The sequences were optimized for codon usage, reduced messenger RNA secondary structures, distribution of GC content and removal of repetitive sequences and motifs. The artificially optimized genes encoding the chimerical protein were synthesized by a commercial supplier (GeneScript®). Further, the synthetic optimized genes coding the Top7-HIV-1 chimerical proteins were inserted in the expression vector pRSETA (Invitrogen®) in-frame with the vector-encoded translation initiator codon and six-His tag encoding sequence (at N-terminus), under the control of the tightly regulated T7 promoter. The obtained proteins were subjected to restriction enzyme analysis in order to verify the sequence identity.

Recombinant protein expression and purification

Bacterial expression of the chimerical proteins was carried out in *Escherichia coli* BL21 Star™ (DE3)pLysS (Invitrogen®) strain transformed with the expression vectors. Transformants harboring the expression plasmids were expression screened to choose a clone that produced Top7–HIV-1chimeras maximally. Cells were grown in Luria-Bertani (LB) medium supplemented with 100 mg mL^{-1} of ampicillin (LB-AMP) to an OD_{600} of 0.5 (*ca.* 2×10^8 cells mL^{-1}) and induced with 1 mM isopropyl β-D-1-thiogalactopiranoside

Table 4 Description of the simulated systems

Epitope ID	System Code	# Atoms			Total # Atoms	Box Edge (nm)
		Solute	Water	Na$^+$ ions		
—	Top7_Original	997	49 831	3	50 831	8.0
E2F5	Top7_E2F5	1098	71 049	3	72 150	9.0
E4E10	Top7_E4E10	1121	98 013	3	99 137	10.0
R5L	Top7_R5L	1106	98 016	2	99 124	10.0
R2H	Top7_R2HA	1015	49 842	3	50 860	8.0
R2H	Top7_R2HB	1027	49 830	5	50 862	8.0
R3H	Top7_R3HA	1025	49 824	4	50 853	8.0
R3H	Top7_R3HB	1029	49 812	5	50 846	8.0
R8H	Top7_R8HA	1030	49 833	5	50 868	8.0
R8H	Top7_R8HB	1038	49 818	7	50 863	8.0

(IPTG) at 30 °C. After 4 h, expressing bacterial pellets were harvested by centrifugation at 5000 g at 4 °C/20 min, and resuspended in a suitable buffer (lysis buffer) containing: 50 mM Tris-HCl pH 7.4, 300 mM NaCl, 20 mM Imidazole, supplemented with protease inhibitor cocktail (Roche). Bacterial suspensions were sonicated, at 4 °C in a Vibracell VCX 750 Sonicator (amplitude setting of 40) using 6 pulses of 30 s each with 60 s off time between the pulses. The lysate was clarified by centrifugation at 8000 g/4 °C/10 min to remove cellular debris. The artificial proteins were purified trough affinity chromatography. Bacterial lysed extracts were applied to a nickel matrix column (Ni-NTA-QIAGEN), which was washed twice with washing buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 50 mM Imidazole) and the induced proteins eluted with the same buffer supplemented with 500 mM of imidazole (eluting buffer). The quantification of the purified recombinant proteins was measured using a Nanodrop 2000c/ 2000 spectrophotometer (Thermo Scientific).

Human sera samples

The serological cohort used in this work was obtained from the National Panel for Blood Screening Quality Control and is composed by 26 HIV-1 positive samples and 21 HIV-1 negative samples. All sera were elaborated and tested by the Technology Institute for Immunobiologicals (Bio-Manguinhos/Oswaldo Cruz Foundation) and the National Institute for Health Quality Control of the Oswaldo Cruz Foundation.

Cytometric bead array

Covalent coupling of chimerical proteins to microspheres. Paramagnetic carboxilated microspheres (Luminex Corp, Austin, USA) were coated with recombinant proteins through a two-step carbodiimide coupling, according to the manufacturer's instruction. Bead suspensions were vortexed vigorously and sonicated in a water bath sonicator (30 s each) aiming to disperse the microspheres as needed. Further, aliquots of 80 μL (2 × 10^6 microspheres per bead set) were washed twice with activation buffer (100 mM NaH_2PO_4, pH 6.2) and their carboxyl groups were activated with 80 μL of the same buffer supplemented with 500 μg of N-hydroxysulfosuccinimide (Pierce, Rockford-IL, USA) and 500 μg of 1-ethyl-3(3-dimethyl-laminopropyl)-carbodiimide hydrochloride (Pierce, Rockford-IL, USA). After incubation for 20 min at room temperature and under stirring (300 RPM), the activated microspheres were subsequently washed with coupling buffer [50 mM 2-(N-morpholino) ethanesulfonic acid – MES pH 5.5, 50 mM 2-(N-morpholino) ethanesulfonic acid – MES pH 6.5 or PBS 1X pH 7.0]. All incubations of beads were performed in the dark. Before coupling, the concentration of the recombinant proteins was adjusted to 50 μg mL^−1 in each respective coupling buffer. Each separate bead population was re-suspended in 100 μL of antigen solution and incubated for 2 h at 37 °C under stirring. The coated beads were then washed with storage buffer [PBS 1X, supplemented with 1% (w/v) bovine serum albumin – BSA, 0.05% Tween-20 and 0.02% sodium azide, suspended once again in 100 μL of the same buffer and the suspensions were allowed to block overnight at 4 °C protected from light. The protein coated microspheres concentration was determined using the Multisizer™ 3

Coulter Counter (Beckman Coulter) and the microspheres were stored at 4 °C in the dark until use.

Microsphere immunoassay standard protocol. Immunoreactions were performed in 96-well plates (SARSTEDT) and all incubations were performed at 37 °C on a microplate shaker (set at 300 rpm) and protected from light. For the wash steps, a Hydroflex plate washer with a magnetic support (TECAN, Durban, NC, USA) was used. Sera samples were diluted in buffer (PBS containing 1% BSA, 0.05% Tween-20) and the assays were conducted using standard procedures described by the manufacturer (Luminex Corp, Austin-TX, USA). Two thousand antigen-coated microspheres per bead set (a final volume of 50 μL) were added to each well of a 96-well plate and incubated with 50 μL of diluted serum for 30 min. Following two washes in 100 μL of assay buffer (PBS containing 1% BSA, 0.05% Tween-20, 100 mM Tris), the microspheres were incubated with goat anti-human IgG conjugated to phycoerythrin (MOSS) diluted 1 : 200 in the same buffer for 30 min. The microspheres were washed twice and re-suspended in 100 μL of assay buffer before Luminex measurement. The reporter fluorescence, expressed as median fluorescence intensity (MFI), was determined with a Luminex 200 using IS 2.3 software, counting 100 events for each analysis.

Statistical analysis

The diagnostic performance evaluation of each recombinant antigen was performed using the ROC curve analysis (GraphPad Prism 5). Determined AUC, sensibility and specificity values, for each immunological assay, are used as standard metric. Additionally, two-tailed unpaired t-tests were then used to compare the MFI values obtained for each protein against the HIV-1panels. Pearson's correlation was used to assess the co-variation among the groups analyzed. Differences were considered statistically significant when $p < 0.05$. All statistical tests were conducted at the 95% confidence level.

ELISA assay

The chimerical proteins Top7_E2F5 and Top7_Original at the concentration of 32 pMol per well were used to coat the wells of EIA/RIA flat bottom (Corning Incorporated®). The reactions were blocked with 5% milk in 1X PBS during 1 h/37 °C and washed 3 times with washing buffer (0.01% Tween-20 in 1X PBS). Following, the monoclonal antibody against the epitope 2F5 (NIH AIDS Research & Reference Reagent Program) was diluted from 1 ng mL^−1 to 1 × 10^−10 ng mL^−1 in assay diluent (1% milk in 1X PBS) and added to the wells during 2 h/37 °C. After rinsing off any unbound antibody, peroxidase-conjugated goat polyclonal anti-human IgG antibody (Jackson ImmuneResearch) was diluted to 1 : 2000 in the dilution buffer and added to the wells 1 h/37 °C. Microtiter plate was read in a PowerWave™ XS microplate reader (Biotek®) set to 450 nm. All samples were run in duplicate and results averaged.

Conclusions

Although the Top7 protein has been previously used as scaffold where a short sequence was inserted in one of its loops, this is the first time that modifications of its secondary elements by antigen sequences are proposed and shown to be viable. To translate structural information into antigen design, it is assumed that:

i. the sites of insertion exhibit a secondary structure similar to the selected segments of the HIV-1 gp41 ectodomain;

ii. the insert is exposed enough to allow binding to the target;

iii. the insert does not disrupt local secondary structure to the point where the native fold of the scaffold would be affected; and,

iv. the insert is conformationally stable to ensure antibody specificity, therefore avoiding unspecific binding.

Based on these principles, molecular dynamics simulations were used to single out 5 promising candidates out of 9 initially designed chimeras. These designed Top7-based chimeras retained roughly 80% to 90% of its original primary structure. Nevertheless, Top7 stability did not sustain the replacement of its helices by the R8H sequence, a change corresponding to 20% of the total Top7 residues. The R8H epitope unfolded regardless of the inserted position, compromising epitope conformation, and scaffold stability up to expression yield.

Four out of the five promising chimeric proteins were shown to have good immunogenic performance in differentiating HIV-1 positive from negative patient sera, consistent with the computational predictions. Even though Top7_E2F5, a loop insert system, did not discriminate positive from negative HIV-1 sera, its predicted structural stability is supported by a strong recognition signal from 2F5 monoclonal antibodies. The poor ability of this system in being recognized by HIV-1 antibodies from patient sera is attributed to the well-known low incidence of naturally generated antibodies against this epitope during natural HIV-1 infection. In addition, ROC curve analysis suggests that three chimeras, Top7_E4E10, Top7_R2HA and Top7_R3HB, may be potentially used in diagnostic kits. It is worth nothing that epitope sequences as presented here, were derived from Latin American viral isolates as well as the human patient serum samples. While out of the scope of this paper, further immunological tests should be carried out to assess the ability of these chimeras in differential diagnostics.

The results presented here demonstrate the correlation between the computational approach and experimental data to provide proof of concept for the feasibility of computationally engineering Top7-based proteins for supporting heterologous sequences. In addition, it shows for the first time that the helical regions of the Top7 scaffold may be engineered to hold sequences of a desired biological property, thereby increasing the range of options to engineer improved scaffold-based epitopes and vaccines.

References

1 S. Kobold, T. Luetkens, Y. Cao, C. Bokemeyer and D. Atanackovic, *Clin. Dev. Immunol.*, 2010, **2010**, 721531.

2 K. Yanamandra, M. A. Gruden, V. Casaite, R. Meskys, L. Forsgren and L. A. Morozova-Roche, *PLoS One*, 2011, **6**, e18513.

3 D. Lelli, A. Moreno, E. Brocchi, E. Sozzi, L. Capucci, E. Canelli, I. Barbieri, H. Zeller and P. Cordioli, *Virol. J.*, 2012, **9**, 81.

4 J. Cao, Q. Chen, H. Zhang, P. Qi, C. Liu, X. Yang, N. Wang, B. Qian, J. Wang, S. Jiang, H. Yang, S. Sun and W. Pan, *PLoS One*, 2011, **6**, e18477.

5 G. Ofek, F. J. Guenaga, W. R. Schief, J. Skinner, D. Baker, R. Wyatt and P. D. Kwong, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 17880–17887.

6 J. Guenaga, P. Dosenovic, G. Ofek, D. Baker, W. R. Schief, P. D. Kwong, G. B. Karlsson Hedestam and R. T. Wyatt, *PLoS One*, 2011, **6**, e16074.

7 A. Changela, X. Wu, Y. Yang, B. Zhang, J. Zhu, G. A. Nardone, S. O'Dell, M. Pancera, M. K. Gorny, S. Phogat, J. E. Robinson, L. Stamatatos, S. Zolla-Pazner, J. R. Mascola and P. D. Kwong, *J. Virol.*, 2010, **85**, 2524–2535.

8 M. L. Azoitei, B. E. Correia, Y. E. Ban, C. Carrico, O. Kalyuzhniy, L. Chen, A. Schroeter, P. S. Huang, J. S. McLellan, P. D. Kwong, D. Baker, R. K. Strong and W. R. Schief, *Science*, 2011, **334**, 373–376.

9 B. E. Correia, Y. E. Ban, D. J. Friend, K. Ellingson, H. Xu, E. Boni, T. Bradley-Hewitt, J. F. Bruhn-Johannsen, L. Stamatatos, R. K. Strong and W. R. Schief, *J. Mol. Biol.*, 2011, **405**, 284–297.

10 J. S. McLellan, B. E. Correia, M. Chen, Y. Yang, B. S. Graham, W. R. Schief and P. D. Kwong, *J. Mol. Biol.*, 2011, **409**, 853–866.

11 B. E. Correia, Y. E. Ban, M. A. Holmes, H. Xu, K. Ellingson, Z. Kraft, C. Carrico, E. Boni, D. N. Sather, C. Zenobia, K. Y. Burke, T. Bradley-Hewitt, J. F. Bruhn-Johannsen, O. Kalyuzhniy, D. Baker, R. K. Strong, L. Stamatatos and W. R. Schief, *Structure*, 2010, **18**, 1116–1126.

12 P. D. Kwong and L. Shapiro, *Sci. Transl. Med.*, 2011, **3**, 91ps29.

13 I. Samish, C. M. MacDermaid, J. M. Perez-Aguilar and J. G. Saven, *Annu. Rev. Phys. Chem.*, **62**, 129–149.

14 S. J. Fleishman, T. A. Whitehead, D. C. Ekiert, C. Dreyfus, J. E. Corn, E. M. Strauch, I. A. Wilson and D. Baker, *Science*, 2011, **332**, 816–821.

15 M. Totrov, X. Jiang, X. P. Kong, S. Cohen, C. Krachmarov, A. Salomon, C. Williams, M. S. Seaman, R. Abagyan, T. Cardozo, M. K. Gorny, S. Wang, S. Lu, A. Pinter and S. Zolla-Pazner, *Virology*, 2010, **405**, 513–523.

16 M. Lapelosa, E. Gallicchio, G. F. Arnold, E. Arnold and R. M. Levy, *J. Mol. Biol.*, 2009, **385**, 675–691.

17 B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard and D. Baker, *Science*, 2003, **302**, 1364–1368.

18 D. Sharma, O. Perisic, Q. Peng, Y. Cao, C. Lam, H. Lu and H. Li, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 9278–9283.

19 T. A. Soares, C. B. Boschek, D. Apiyo, C. Baird and T. P. Straatsma, *J. Mol. Graphics Modell.*, 2010, **28**, 755–765.

20 C. B. Boschek, D. O. Apiyo, T. A. Soares, H. E. Engelmann, N. B. Pefaur, T. P. Straatsma and C. L. Baird, *Protein Eng., Des. Sel.*, 2009, **22**, 325–332.

21 W. Weissenhorn, A. Dessen, S. C. Harrison, J. J. Skehel and D. C. Wiley, *Nature*, 1997, **387**, 426–430.

22 M. Caffrey, *Biochim. Biophys. Acta, Mol. Basis Dis.*, 2001, **1536**, 116–122.

23 W. Shi, J. Bohon, D. P. Han, H. Habte, Y. Qin, M. W. Cho and M. R. Chance, *J. Biol. Chem.*, 2010, **285**, 24290–24298.

24 G. Ofek, M. Tang, A. Sambor, H. Katinger, J. R. Mascola, R. Wyatt and P. D. Kwong, *J. Virol.*, 2004, **78**, 10724–10737.

25 R. M. Cardoso, M. B. Zwick, R. L. Stanfield, R. Kunert, J. M. Binley, H. Katinger, D. R. Burton and I. A. Wilson, *Immunity*, 2005, **22**, 163–173.

26 J. Kyte and R. F. Doolittle, *J. Mol. Biol.*, 1982, **157**, 105–132.

27 J. Guenaga and R. T. Wyatt, *PLoS Pathog.*, **8**, e1002806.

28 C. B. Boschek, D. O. Apiyo, T. A. Soares, H. E. Engelmann, N. B. Pefaur, T. P. Straatsma and C. L. Baird, *Protein Eng., Des. Sel.*, 2009, **22**, 325–332.

29 T. A. Soares, C. B. Boschek, D. Apiyo, C. Baird and T. P. Straatsma, *J. Mol. Graphics Modell.*, 2010, **28**, 755–765.

30 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.

31 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.

32 L. Stamatatos, L. Morris, D. R. Burton and J. R. Mascola, *Nat Med*, 2009, **15**, 866–870.

33 D. Kamdem Toukam, M. Tenbusch, A. Stang, V. Temchura, M. Storcksdieck Genannt Bonsmann, B. Grewe, S. Koch, A. Meyerhans, G. Nchinda, L. Kaptue and K. Uberla, *PLoS One*, 2012, **7**, e38068.

34 E. R. DeLong, D. M. DeLong and D. L. Clarke-Pearson, *Biometrics*, 1988, **44**, 837–845.

35 S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389–3402.

36 S. F. Altschul, J. C. Wootton, E. M. Gertz, R. Agarwala, A. Morgulis, A. A. Schäffer and Y.-K. Yu, *FEBS J.*, 2005, **272**, 5101–5109.

37 N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, **18**, 2714–2723.

38 H. J. C. Berendsen, J. R. Grigerra and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.

39 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

40 D. van der Spoel and H. J. C. Berendsen, *et al.*, *Comput. Phys. Commun.*, 1995, **91**, 43–56.

41 e. a. Lindahl, *J. Mol. Model.*, 2001, **7**, 306–317.

42 e. a. van der Spoel, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

43 e. a. Hess, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

44 C. Oostenbrink, A. Villa, A. E. Mark and W. F. van Gunsteren, *J. Comput. Chem.*, 2004, **25**, 1656–1676.

45 H. P. Berendsen, J. P. M. van Gunsteren, W. F. DiNola and A. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

46 B. B. Hess, H. Berendsen and H. J. C. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463.

47 R. W. Hockney, in *Methods in Computational Physics*, ed. A. Press, New York/London, Editon edn, 1970.

48 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33.