doi:10.1016/j.jmb.2008.10.089 J. Mol. Biol. (2009) 385, 675–691

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

In Silico Vaccine Design Based on Molecular Simulations of Rhinovirus Chimeras Presenting HIV-1 gp41 Epitopes

Mauro Lapelosa[1,2,3], Emilio Gallicchio[1,2*], Gail Ferstandig Arnold[2,3], Eddy Arnold[2,3] and Ronald M. Levy[1,2]

[1]*BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, NJ 08854, USA*

[2]*Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA*

[3]*Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ 08854, USA*

Received 30 June 2008; received in revised form 15 September 2008; accepted 31 October 2008
Available online 8 November 2008

A cluster of promising epitopes for the development of human immunodeficiency virus (HIV) vaccines is located in the membrane-proximal external region (MPER) of the gp41 subunit of the HIV envelope spike structure. The crystal structure of the peptide corresponding to the so-called ELDKWA epitope (HIV-1 HxB2 gp41 residues 662–668), in complex with the corresponding broadly neutralizing human monoclonal antibody 2F5, provides a target for structure-based vaccine design strategies aimed at finding macromolecular carriers that are able to present this MPER-derived epitope with optimal antigenic activity. To this end, a series of replica exchange molecular dynamics computer simulations was conducted to characterize the distributions of conformations of ELDKWA-based epitopes inserted into a rhinovirus carrier and to identify those with the highest fraction of conformations that are able to bind 2F5. The length, hydrophobic character, and precise site of insertion were found to be critical for achieving structural similarity to the target crystal structure. A construct with a high degree of complementarity to the corresponding determinant region of 2F5 was obtained. This construct was employed to build a high-resolution structural model of the complex between the 2F5 antibody and the chimeric human rhinovirus type 14:HIV-1 ELDKWA virus particle. Additional simulations, which were conducted to study the conformational propensities of the ELDKWA region in solution, confirm the hypothesis that the ELDKWA region of gp41 is highly flexible and capable of assuming helical conformations (as in the postfusion helical bundle structure) and β-turn conformations (as in the complex with the 2F5 antibody). These results also suggest that the ELDKWA epitope can be involved in intramolecular—and likely intermolecular—hydrophobic interactions. This tendency offers an explanation for the observation that mutations decreasing the hydrophobic character of the MPER in many cases result in conformational changes that increase the affinity of this region for the 2F5 antibody.

Keywords: vaccine design; monoclonal antibody 2F5; replica exchange method; chimeric virus

Edited by M. Levitt

Introduction

Due to the difficulty of providing universal anti-human immunodeficiency virus (HIV) drug therapies, the development of an effective AIDS vaccine remains the most promising long-term strategy to combat HIV/AIDS. There is a strong need for novel alternative approaches aimed at discovering an effective AIDS vaccine.[1] This work focuses on the development of vaccine components that are capable of eliciting broadly neutralizing immune responses.

*Corresponding author. BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, NJ 08854, USA. E-mail address: emilio@biomaps.rutgers.edu.

Abbreviations used: HIV, human immunodeficiency virus; MPER, membrane-proximal external region; nAb, neutralizing antibody; HRV14, human rhinovirus type 14; AGBNP, analytical generalized Born plus nonpolar; REMD, replica exchange molecular dynamics; MD, molecular dynamics; SASA, solvent-accessible surface area.

Optimized immunogens of this type could be combined with other B-cell immunogens, as well as with T-cell immunogens, for a multipronged attack on HIV, analogous to the multipronged attacks provided by drug cocktails. Binding of a broadly neutralizing antibody (nAb) to a pathogen containing large sequence diversity, such as HIV, is undoubtedly a prerequisite for the development of optimal B-cell immunogens. Nonetheless, the best immunogens to date remain insufficient for vaccination purposes.[2]

Antibodies can neutralize HIV-1 either by binding the virus surface and preventing it from interacting with cellular receptors or by binding after virion attachment and blocking the subsequent steps involved with virus entry.[3] HIV-1 fuses to host cells with its surface spikes, composed of trimeric complexes of envelope glycoproteins gp41 (transmembrane) and gp120 (surface). gp120 subunits, decorated with host cell carbohydrates, form much of the external surface of the spike, which is anchored to the viral membrane via noncovalent association with the gp41 subunits. Binding of gp120 to the CD4 receptor of the target cell is the first step in the viral entry process, followed by a large conformational reorganization of the spike, eventually leading to viral fusion with the host cell through juxtaposition of the viral and host cell membranes.[4] gp41 plays a major role in this process by forming a fusogenic six-helix bundle structure that is believed to drive the fusion process.[5] gp41 contains ~345 residues[6] and is divided into an ectodomain, a membrane-spanning segment, and a cytoplasmic tail.[7,8] The fusogenic six-helix bundle structure is formed when the N-terminal and C-terminal heptad repeats of the trimer of ectodomains of gp41 are able to associate in the absence of gp120.[9]

In this study, we model epitopes derived from the conserved membrane-proximal external region (MPER) of gp41. The MPER is composed of a tryptophan-rich segment, presumably promoting some degree of positioning at the interface between water and the lipid bilayer of the viral membrane.[10] Although the MPER is generally thought to be extensively cloaked in the context of the envelope spike, three of the broadest and most potent nAbs for HIV-1—2F5,[11] 4E10[12], and Z13e1[12,13]—have been shown to bind here in close proximity. The epitope sites in the MPER are thus considered as some of the most promising targets for AIDS vaccine development.[14,15] However, despite their promise, nAbs targeting the MPER, including 2F5, have been hypothesized to occur naturally only in a small minority of infected individuals. This phenomenon could be artifactual; this could be based on poor binding of serum antibodies to recombinant envelope mimics that may not have native-like conformations, and/or it could be that poor accessibility of the 2F5 epitope within the viral spike protein complex hinders recognition by the immune system or, alternatively, that this region could be exposed, but presented in a favorable immunogenic conformation only transiently.

The Arnold Laboratory has developed combinatorial libraries to display HIV-1 epitopes connected to surface loops of human rhinovirus type 14 (HRV14),[16,17] with the aim of enhancing the probability of identifying epitope presentations that are able to evoke protective immune responses. HRV14, like all picornaviruses, has an icosahedral capsid shell consisting of 60 copies of the so-called protomeric subunit, comprising one copy each of the viral coat proteins VP1–VP4. Genetic engineering of HRV has resulted in the presentation of foreign immunogens that are able to trigger immunological responses specifically directed to the corresponding pathogenic antigens, in some cases quite broadly.[16–18]

Our efforts to develop chimeric HRV displaying the 2F5 epitope of HIV-1 gp41 (called ELDKWA after its core sequence) have provided indications that it may be possible to elicit protective immune responses with this approach. We hypothesize that stronger immunogens in this class, which are more likely to present the ELDKWA epitope in a conformation compatible with effective binding with the complementarity-determining region of 2F5, can be produced. The primary aim of the work described here is to apply computational modeling to identify chimeric HRV constructs that display preferentially the epitope in the same β-turn conformation as seen in the crystal structure in complex with 2F5. Although structure-based[19] and bioinformatics[20] techniques have been used to find immunologically active sequences, the potential of computational modeling for vaccine design has only recently begun to be explored. Computational modeling approaches have been employed to design new peptides exhibiting molecular mimicry for known antigens;[21] to characterize chimeric viruses presenting a double epitope with immunologically stimulating activity;[22] and to compare free and bound crystal structures of monoclonal antibodies, illustrating the extent of induced-fit effects on antibodies binding to their targets.[23]

In this work, we seek to optimize the affinity of the 2F5 nAb for the ELDKWA epitope using a molecular simulation strategy that is similar in spirit to previous work on a bacterial epitope;[21] the goal is to design a virus construct that maximizes the propensity of the epitope to assume the conformation seen in the known structure of the complex with the 2F5 Fab.[24] This approach assumes that the binding affinity of the ELDKWA epitope for 2F5 is increased when preferentially presenting this epitope in a conformation predisposed for binding. Conformational reorganization is an important factor for protein recognition and binding. Both binding partners incur a free-energy penalty for reorganizing the ensembles of conformations present in their unbound forms to those compatible with complexation.[25] Binding affinity is enhanced when the binding partners already assume binding-competent conformations in their unbound forms. Minimizing the binding reorganization free energy is particularly important in vaccine design applications

because, in general, epitope composition cannot be freely varied to increase the binding affinity for target antibodies.

Achieving this goal requires high-quality atomic-resolution models. We employ in this study an energy function based on the optimized potential for liquid simulations all-atom force field[26,27] combined with the analytical generalized Born plus nonpolar (AGBNP) implicit solvent model,[28] which has been shown to be suitable for high-resolution modeling of peptides and proteins.[29–31] This free-energy model has been recently reported to perform well for structure prediction of protein surface loops,[32] a problem intimately related to the vaccine design strategy outlined here.

Another important requirement for this study is the use of a canonical conformational sampling method that is capable of predicting the thermodynamic stability of conformational states at physiological temperature. Here we employ the temperature replica exchange molecular dynamics (REMD) conformational sampling algorithm,[33] an advanced technique that consists of simulating many instances of the system in parallel over a series of temperatures. This canonical algorithm enhances conformational sampling efficiency, alleviating some of the ergodicity problems associated with standard molecular dynamics (MD) algorithms, but requires more extensive computational resources for its implementation.

We have carried out REMD simulations of various sequences derived from the MPER of gp41 in different structural contexts—as free peptides in solution and as covalently attached peptides within a solvent-accessible loop on the surface of HRV14. The peptide simulations provided information on the intrinsic conformational propensities of the epitopes and suggested sequences that are more likely to be presented on HRV14 in the appropriate conformation for antibody binding. In addition, the peptide simulations provided insights into the native structure of the MPER within the context of the gp41 subunit. The simulations confirmed the hypothesis that the ELDKWA region of the MPER is highly flexible and capable of assuming helical conformations (as seen in the crystal structure of the post-fusion helical bundle that extends to the EL motif of the ELDKWA epitope)[34,35] and β-turn conformations (as seen in the crystal structure of an ELDKWA peptide complexed with the 2F5 Fab).[24] The results also provide an interesting perspective on the role of hydrophobic interactions among several tryptophan residues present in this region in influencing the structure of the MPER of gp41.

The simulations of the chimeric HRV carrying ELDKWA-derived peptides show that the foreign insert could fold in a variety of conformations, most of which are unsuitable for binding to the 2F5 nAb. We identified several energetic and topological factors that affect the population of binding-competent conformations of the inserted peptide and, using this knowledge, we identified a particular construct that is capable of displaying the ELDKWA epitope in a well-exposed conformation that is able to fit into the 2F5 nAb cleft. The quality of agreement between some of the modeled conformations of the inserted peptide with the crystal structure of the 2F5 Fab complexed with the ELDKWA epitope has allowed us to produce a high-resolution structural model of the complex between the 2F5 nAb and the optimized chimeric HRV.

Results and Discussion

Limited exposure of the critical DKW residues of the ELDKWA epitope in the reference chimeric virus construct (cHRV1)

We have recently characterized an HRV14:HIV-1 chimeric rhinovirus displaying the gp41 ELDKWA epitope named cHRV1 (Table 1). When used to immunize guinea pigs, this chimeric virus was unable

Table 1. Peptides and chimeric viruses investigated in this work

Peptide	Sequence[a]
7-mer peptide 1	662-ELDKWAS-668
7-mer peptide 2	662-ALDKWAS-668
7-mer peptide 3	662-EADKWAS-668
7-mer peptide 4	662-ELAKWAS-668
7-mer peptide 5	662-ELDAWAS-668
7-mer peptide 6	662-ELDKAAS-668
7-mer peptide 7	662-ELDKWAA-668
21-mer peptide	660-LLELDKWANLWNWFDISNWLW-680
21-mer mutated peptide	660-LLELDKWANLANAFDISNALA-680

Chimeric virus	HRV	N-linker	2F5 epitope	C-linker	HRV
cHRV1 (14-C40-1)	…DLS	PCG	ALDKWAS	SPDCS	VVG…
cHRV2	…DLS	PCG	LLELDKWANLANAFDISNALA		VVG…
cHRV3	…DLSSAN		ELDKWAS		EVVG…

[a] Mutations from the reference sequence are underlined. For the chimeric construct, the sequence of the inserted 2F5 epitope is given together with the linker sequences and flanking HRV14 residues.

to elicit neutralizing anti-HIV-1 antibodies by itself; however, when the virus was boosted with a keyhole-limpet-hemocyanin-conjugated 14-mer peptide corresponding to the ELDKWA region (also unable to elicit neutralizing anti-HIV-1 antibodies by itself), antibodies that were capable of neutralizing diverse isolates of HIV-1 were elicited (data not shown). We speculate that the presentation of the ELDKWA epitope on the chimeric virus was not sufficient to elicit a broadly neutralizing response by itself, but that the peptide boosts stimulated maturation of the secondary antibodies, enabling some of them to recognize and neutralize diverse isolates. Presumably, such a response reflects the fact that the chimeric virus is an imperfect immunogen, but not extremely so, as other virus immunogens with peptide boosts remained unable to elicit favorable immune responses.

To better understand the nature of this immunogen and to guide improved designs of immunogen constructs, we performed an REMD simulation of cHRV1 to assess the likely exposure and conformational propensities of the ELDKWA epitope on the surface of this rhinovirus. The simulation indicated that the ELDKWA epitope was highly mobile on the surface of the chimeric rhinovirus; however, the exposure of the inserted sequence was found to be rather limited on average (Fig. 1a) and, thus, not optimal for binding to the 2F5 nAb. The ensemble of conformations from the simulation was analyzed by hierarchical clustering based on pairwise C^{α} root mean square deviation (RMSD) values. In the most representative conformation, the side chains of the critical DKW residues point away from each other rather than being aligned as in the structure of the complex with 2F5,[24] and they fold over the VP2 subunit (Fig. 1a). In the dominant conformational cluster, only 17% of the PCG-ALDKWAS-SPDCS insert is solvent-exposed. This conformation has numerous clashes when superimposed with the corresponding paratope of 2F5. Nevertheless, in a minority of conformations, the critical DKW residues recognized by 2F5 are more solvent-exposed, consistent with the observation[19] that nAb 2F5 can bind and neutralize this virus. In Fig. 2a, the distribution of conformations in the simulated ensemble is shown; C^{α} RMSD values of the ALDKWAS motif of each conformation are plotted with respect to that of the solution NMR structure of an ELDKWA-derived 13-mer peptide (1LCX)[36] and of the crystal structure of an MPER-derived 7-mer peptide in complex with 2F5 (1TJG).[24] The majority of conformations from the simulated ensemble of cHRV1 structures lie in the upper-right quadrant of Fig. 2a, indicating that they do not generally resemble the structural target represented by the crystal structure of the corresponding peptide complexed with the 2F5 Fab[25] or the solution NMR structure.[36]

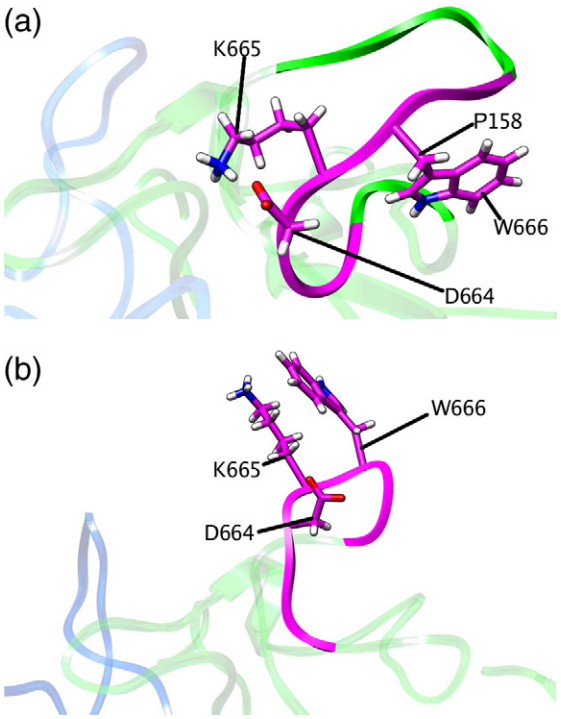

Fig. 1. View of the representative modeled conformation of the neutralizing immunogenic site II region of the cHRV1 HRV14:HIV-1 chimeric construct (a) and that of cHRV3 (b). The VP2 subunit and the inserted sequence are shown in green and magenta, respectively, including the side chains of the core ELDKWA epitope residues (DKW) and P158 (cHRV1 only) of the N-terminal linker. The VP1 subunit of HRV14 is shown in blue.

The fraction of conformers of the inserted ELDKWA epitope dissimilar to the corresponding peptide in the crystal structure is a measure of the unfavorable reorganization free energy that must be overcome in order to achieve successful binding to the 2F5 nAb. As seen in Fig. 2a, the REMD trajectory at 310 K shows that this fraction is relatively large (approximately 95%, as measured by clustering analysis), leading to the prediction that cHRV1 has a relatively low affinity for 2F5 due to the unfavorable reorganization free energy required for binding.

Higher-affinity binding can be achieved by maximizing the fraction of binding-competent conformations. To achieve this goal, we set out, as described next, to analyze the intrinsic conformational propensities of gp41-derived peptides, as well as the influence of the framework provided by the HRV coat protein.

Dominance of α-helical and β-turn conformers of the ELDKWAS peptide sequence

To discriminate the effect of the intrinsic conformational propensity of the sequence from the influence of the framework in which it is embedded, we performed REMD simulations of the 7-residue ELDKWAS peptide in solution. Studies conducted by two laboratories[36,37] on related peptides did not provide a consensus interpretation of the solution NMR data. Unlike the same sequence integrated within gp41,[38] the free 7-residue peptide binds weakly to 2F5.[39] It has been further shown that longer 13- to 15-residue peptides containing the

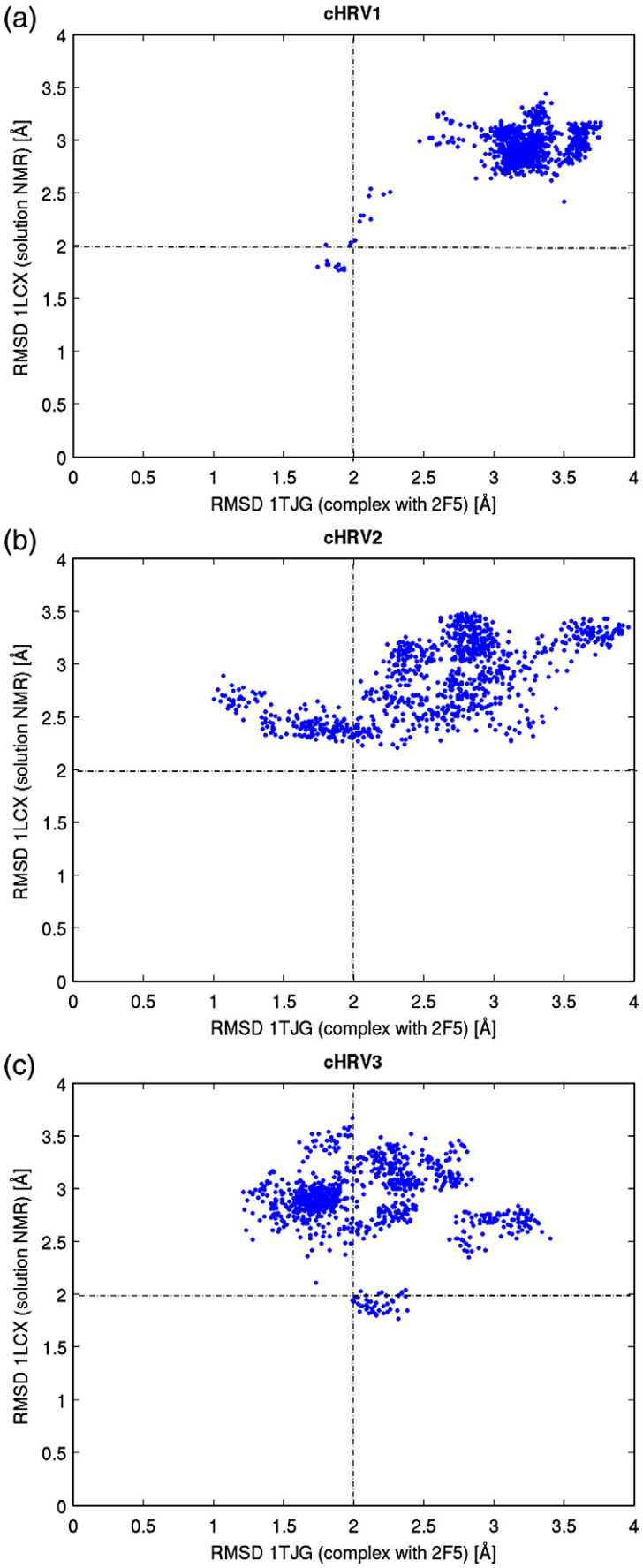

Fig. 2. Scatter plot of the C^{α} RMSD values of the conformations of the ELDKWAS motif from the REMD ensembles, at 313 K, of the chimeric constructs cHRV1 (a), cHRV2 (b), and cHRV3 (c), with respect to the corresponding motifs of the peptide in the 1TJG crystal structure (*x*-axis) and in the 1LCX NMR structure (*y*-axis).

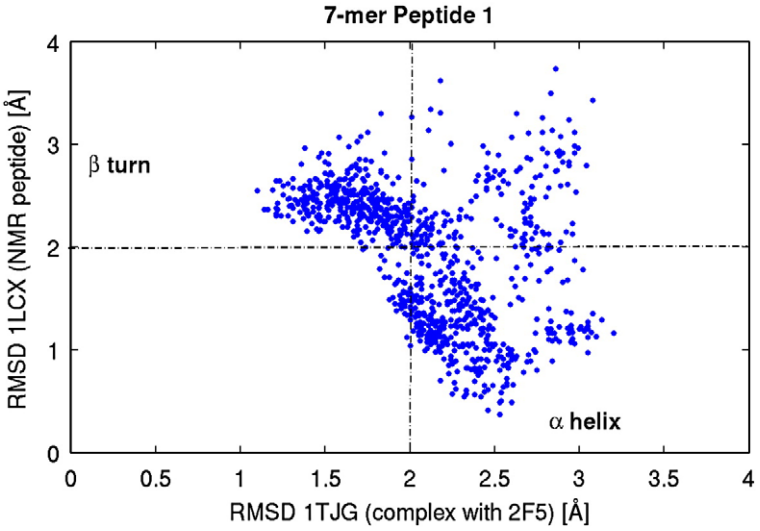

Fig. 3. Scatter plot of the C^{α} RMSD values of the conformations of the ELDKWAS peptide in solution from the REMD ensembles at 313 K, with respect to the corresponding sequence motifs of the peptide in the 1TJG crystal structure (x-axis) and in the 1LCX NMR structure (y-axis). The conformations in the lower-right panel are mostly α-helical, whereas the upper-right panel holds mainly β-turn conformations.

minimal ELDKWA epitope display an increased binding affinity for 2F5.[37] These findings illustrate how the conformational propensity of this sequence can be modified by altering the surrounding structural context. They suggest that, in longer sequences, the fraction of the epitope core sequence (DKW) compatible with 2F5 binding can be increased relative to that of the ELDKWAS minimal peptide, and also that some of the surrounding residues (such as the adjacent leucine residues in LELDKWASL) may make favorable contacts to stabilize binding with 2F5.

REMD simulations of the minimal 7-residue peptide ELDKWAS indicate that this peptide is characterized by significant structural heterogeneity. Figure 3 presents the computed distribution at 313 K of the C^{α} RMSD values of the ELDKWAS motif with respect to the 1LCX NMR structure[36] and the 1TJG crystal structure,[24] as described above for the cHRV1 chimera. The 1LCX solution NMR structure is the 3_{10}-helical structure reported by Biron *et al.*[36] of a 13-residue peptide containing the ELDKWAS motif (Barbato *et al.*[37] reported a less homogeneous NMR ensemble for a closely related peptide), while the 1TJG crystallographic structure represents the β-turn conformation of the ELDKWAS peptide complexed with the 2F5 Fab reported by Ofek *et al.*[24]

Hierarchical clustering analysis based on pairwise C^{α} RMSD values of the simulated ensemble identifies two highly populated clusters corresponding to helical (Cluster 1) and type I β-turn (Cluster 2) conformations representative of the solution NMR structure and the crystal structure, respectively. The "helical" cluster (Cluster 1), which corresponds approximately to samples in the lower-right quadrant of Fig. 3, has a fractional population of 47%. The structural correspondence between the representative structure of Cluster 1 and the NMR solution structure is illustrated in Fig. 4a. In contrast, the structure representative of Cluster 2 is more similar to the crystal structure (1.5 Å RMSD) than to the NMR structure (2.4 Å RMSD). This "β-turn" cluster (Cluster 2), which corresponds approximately to the samples in the upper-left quadrant of Fig. 3, has a fractional population of 44%. The structural similarity between the representative structure of Cluster 2 and the crystal structure from the 2F5 complex is illustrated in Fig. 4b. Here both the crystal structure and the modeled structure adopt a type I β-turn conformation. The difference between the conformations in Cluster 2 and the crystal structure is mainly limited to the orientation of the W666 indole ring, which is often rotated toward the interior of the peptide in the simulated conformations rather than facing outward as in the crystal structure (Fig. 4b).

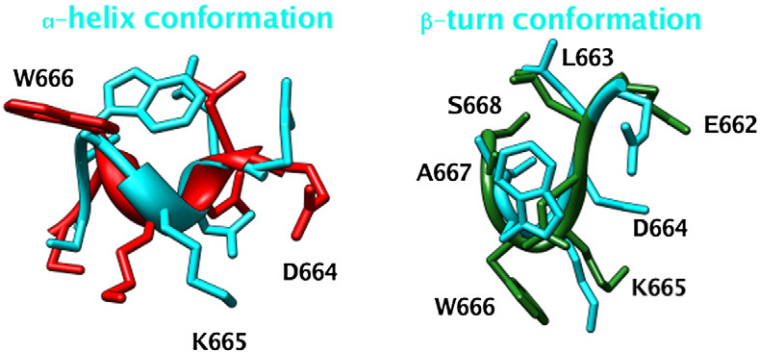

Fig. 4. Structural comparison between the representative conformation of the α-helical conformational cluster of the ELDKWAS peptide (7-mer peptide 1; cyan) and the ELDKWAS region of the 3_{10}-α-helical NMR structure (1LCX model 1; red) (a; left). Structural comparison between the representative conformation of the β-turn conformational cluster of the ELDKWAS peptide (7-mer peptide 1; cyan) and the ELDKWAS motif of the crystal structure of the peptide complexed with the 2F5 Fab (1TJG; green) (b; right). The structures are superimposed with respect to the C^{α} atoms of residues 662–668.

This bias observed in the computed structural ensemble was confirmed by performing an MD simulation starting from the conformation of the peptide in the crystal structure in complex with 2F5. In this simulation, the side chain of the W666 rotated after 8 ns from the outward orientation to the inward orientation, as observed in most of the conformations from Cluster 2.

These results indicate that the structure of the ELDKWAS peptide interconverts in solution between an α-helical conformation and a type I β-turn conformation. This structural heterogeneity is presumably the origin of the discordant conclusions of NMR studies in which this peptide has been described as assuming a 3_{10}-helix conformation[36] or multiple diverse conformations.[37] Both α-helical and β-turn conformations, corresponding to different biological functions for this region, have been observed: the immunogen has been seen to interact with 2F5 in a β-turn conformation[24] and to interact (at least in its N-terminal portion) with the corresponding oligomeric MPER gp41 segments in the fusion-driving six-helix bundle structure.[34,35] Although it is not known whether the six-helix bundle or the triple helix that is thought to precede it actually extends all the way into the ELDKWA epitope, it is considered likely to do so, especially given that: (1) ELDKWA-based peptides can form α-helical conformations in aqueous solution;[36] (2) the epitope for 4E10 (distal to the ELDKWA epitope with respect to the six-helix bundle) is known to form an α-helix in the presence of the antibody; and (3) there are no major helix-breaking residues within the ELDKWA epitope. In addition, (4) related helical conformations were obtained in an NMR study of a 19-residue peptide incorporating the ELDKWAS sequence into dodecyl phosphocholine micelles, mimicking the viral membrane environment.[40] In this study, the peptide was shown to be positioned at the interfacial region between the membrane and the solvent, with the tryptophan rings preferentially inserted into the membrane bilayer.

It has been shown that alanine substitutions in the gp41 MPER of an HIV pseudovirus (HIV-1_{JR2}) can alter antigenic recognition by the 2F5 nAb, as well as by the 4E10 nAb, whose epitope is proximal to the 2F5 epitope.[38] In particular, it was observed that alanine single mutations of the DKW core of the 2F5 epitope (D664A, K665A, and W666A) substantially lowered the affinity of 2F5 for the pseudovirus. In contrast, the replacement by alanine of many hydrophobic residues nearby, including those in the 4E10 epitope (W670A, W672A, and W678A), increased the binding affinity for the 2F5 nAb, even though these residues have not been shown to have direct binding interactions with the 2F5 nAb. Similarly, alanine substitution for W666 in the center of the ELDKWA epitope was found to increase the binding affinity of 4E10 for its epitope, also in the absence of any evidence for direct binding interactions. Thus, the affinity of the 2F5 and 4E10 nAbs for the MPER of gp41 is influenced by the presence of hydrophobic residues, even though these may not be directly involved in binding and are relatively distant in sequence from the characterized epitopes. These observations suggest that the composition of the hydrophobic flanking regions indirectly influences the binding affinity by modulating the conformation of the epitopes at a distance. One possibility is that reducing the hydrophobicity of these flanking regions could favor conformations of epitope regions that have better shape complementarity with the corresponding paratope on antibody molecules.

These observations led us to explore the possibility that the affinity of the ELDKWAS sequence for 2F5 could be determined by the conformational propensity of the ELDKWAS sequence to assume a β-turn conformation. To test this possibility, we performed *in silico* alanine scanning REMD simulations aimed at predicting the conformational propensities of a series of peptides derived from the native ELDKWA epitope (ELDKWAS; Peptide 1) by mutating each residue, in turn, to alanine (peptides 2–7). We found that the average helicity of the overall ensemble (35.1%, 37.2%, and 38%, respectively) of peptides with D664A, K665A, and W666A mutations was significantly greater than that of both natural sequences (20% for ELDKWAS and 15% for ALDKWAS). These are the same mutations that were shown to most markedly decrease the affinity of 2F5 for the envelope spikes of the corresponding pseudoviruses.[38] The present computational predictions suggest that alanine replacements of the DKW core epitope, in addition to the obvious loss of critical binding interactions, may further disfavor 2F5 binding by stabilizing helical conformations at the expense of the β-turn conformations observed in the complex with 2F5.[24] This possibility is supported by the observation that constructs in which the ELDKWAS sequence was incorporated into helical frameworks have no detectable 2F5 reactivity.[41] It should be noted however that the dependence of 2F5 affinity on the α-helical propensity of the ELDKWAS motif is unclear as enhanced binding to the 2F5 antibody upon increased α-helical content has also been reported.[42]

Reduction of hydrophobic residue content enhances epitope exposure in a 21-mer MPER-derived peptide

To further investigate the role of alanine substitution in the binding affinities of 2F5 and 4E10 (the antibody against HIV known to be the most broadly neutralizing, which binds between ELDKWA and the viral membrane),[38] we conducted solution REMD simulations of a 21-amino-acid MPER-derived peptide containing the epitopes of both 2F5 and 4E10 (sequence LLELDKWANLWNWFDISNWLW from HIV-1 HxB2 gp41 residues 660–680, with the epitopes underlined). The results of the RMSD analysis of the 313 K simulated ensemble based on the ELDKWAN core motif, with respect to the crystal and NMR reference structures used above (Fig. 5a), show grouping similar to that seen for the ELDKWAS peptide (Fig. 3). The helical and β-turn conformations

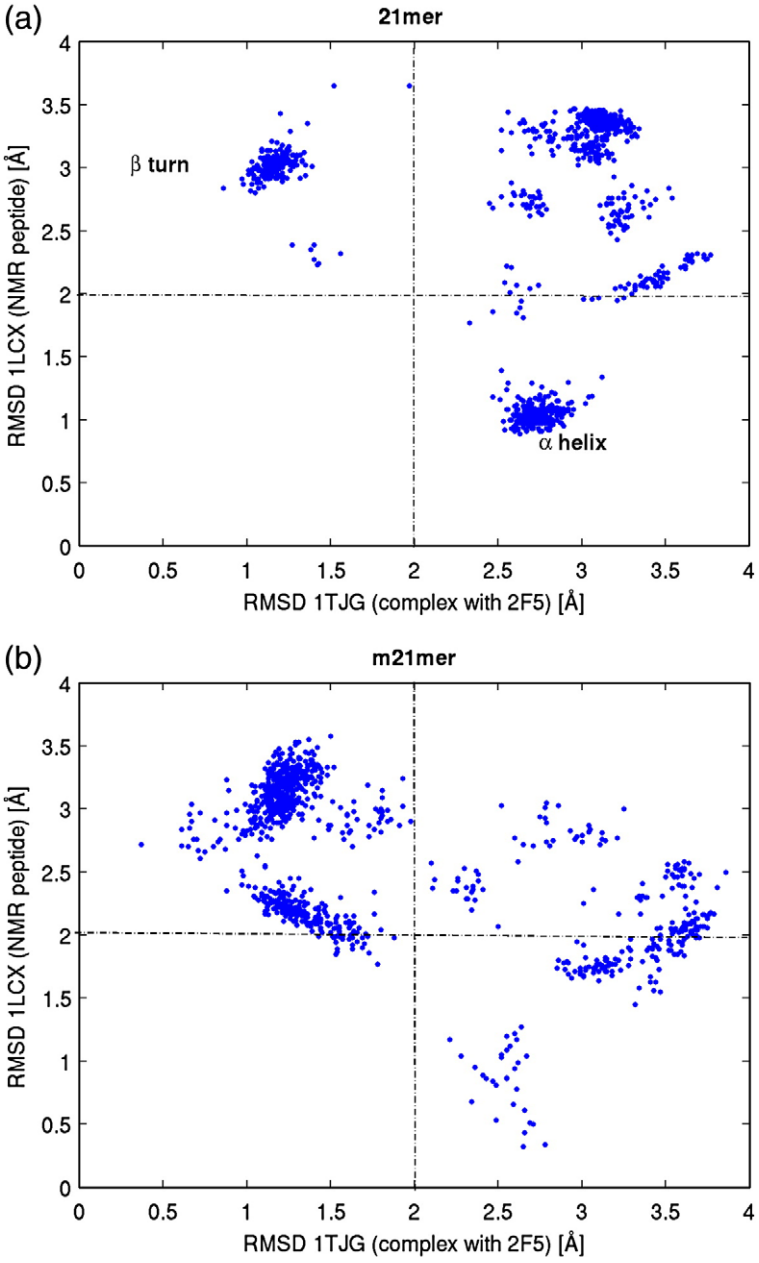

Fig. 5. Scatter plot of the of the C^{α} RMSD values of the ELDKWAS motif of the 21-mer LLELDKWANLWNWFDISNWLW peptide (a) and of the 21-mer LLELDKWANLANAFDISNALA mutated peptide (b), with respect to the corresponding motifs of the peptide in the 1TJG crystal structure (*x*-axis) and in the 1LCX NMR structure (*y*-axis).

correspond to 14% and 25% of the structures, respectively. In addition, we find a significant number (61%) of conformations (upper right in Fig. 5a) in which the ELDKWAN motif does not assume either helical or β-turn structures. Some of these conformers are likely to correspond to a series of dominant conformers identified by clustering analysis that all display extensive long-range hydrophobic interactions, as in the representative structure shown in Fig. 6.

The large hydrophobic nucleus observed in these structures is mainly stabilized by Trp-Trp π-stacking interactions. The critical W666 and W672 residues of the 2F5[12,24] and 4E10[12,43] epitopes, respectively, appear to play central roles in the formation of the hydrophobic nucleus in nearly all of the structures at 313K. Thus, if such intramolecular hydrophobic interactions occur *in vivo* or if alternative intermolecular hydrophobic interactions occur *in vivo* instead among the conserved tryptophans, then these tryptophans would be unavailable for intermolecular binding interactions with their cognate antibodies (or possibly other biological partners). These results indicate that the binding affinity of the gp41 MPER is modulated by a competition between conformations with extensive intramolecular hydrophobic interactions in which W666 and W672 are likely to be sequestered, and conformations in which these residues are available for binding. The greater antigenic affinity observed for alanine-substituted mutants is rationalized by the tendency of these mutations to disrupt the hydrophobic interactions, thereby shifting the equilibrium toward conformers in which the W666 and W672 side chains become more available for binding to their respective antibodies. We confirmed this hypothesis by carrying out an REMD simulation of the quadruple mutant peptide LLELDKWANLANAFDISNALA,

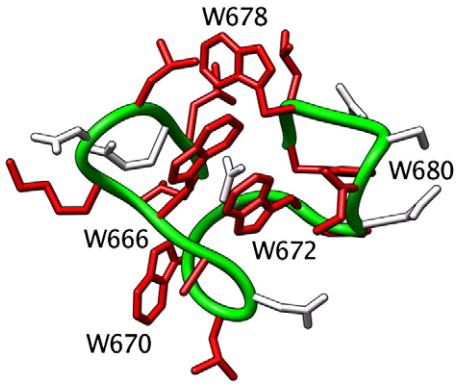

Fig. 6. Structural representation of a highly populated conformational state of the LLELDKWANLWNWFDISNWLW peptide in solution from the REMD ensemble at 310 K illustrating the hydrophobic amino acids (red) involved in the formation of a tightly packed hydrophobic cluster.

in which four tryptophan residues (W670, W672, W678, and W680) of the natural MPER 21-residue peptide were simultaneously replaced with alanine residues. We found that in the conformational ensemble of this mutated sequence, the tendency toward hydrophobic nucleation was considerably diminished relative to the tryptophan-rich natural sequence. Concomitantly, there was greater occurrence in this ensemble relative to the ensemble of the natural sequence of conformations in which the backbone of the DKW sequence structurally matched that of the same residues in the crystal structure in which a peptide is complexed with the 2F5 nAb [35% for the quadruple alanine mutant (Fig. 5b) *versus* 19% for the natural sequence (Fig. 5a)]. The clustering analysis showed that conformations structurally similar to the crystal structure (1TJG) became more populated, and that the population of the cluster of helical conformations decreased (Fig. 4a and b). The REMD results are consistent with a small free-energy difference between two functionally relevant conformations. Furthermore, in all of the conformations of the alanine-substituted peptide, the critical W666 residue did not form hydrophobic contacts and was fully available for intermolecular interactions (e.g., with 2F5). This is in contrast to the natural peptide ensemble, where W666 was sequestered and not available for binding even for those of conformations (19%) in which the DKW motif matched the structure of the peptide complexed with 2F5. This is illustrated in Table 2, which reports a quantitative measure of the degree of pairwise hydrophobic packing for the five key tryptophan positions (W666, W670, W672, W678, and W680) in the natural 21-mer sequence among conformations in which the structure of the DKW motif matched that of the corresponding residues in complex with 2F5. In this table, values near 1 indicate a high degree of packing, whereas values near 0 correspond to little or no packing. The data show that, in the natural peptide, the critical W666 residue forms extensive hydrophobic contacts, particularly with W672 and W670 (with packing degrees of 0.77 and 0.72, respectively) and also interacts with W678 and W680 despite their relatively distant sequence position. Similar trends toward the formation of extensive hydrophobic contacts are observed for the W672 residue, which is part of the 4E10 epitope. In contrast, the same analysis conducted on the conformational ensemble corresponding to the alanine-mutated sequence resulted in values of the degree of pairwise packing close to 0 (not shown), indicating that the occurrence of intramolecular hydrophobic interactions is greatly reduced in this peptide.

Although the simulation performed here suggests that intramolecular hydrophobic aggregation hinders exposure of the 2F5 and 4E10 epitopes, hydrophobic associations within the gp41/gp120 trimeric complex and/or with the viral membrane or other biological partners could similarly make them less available for interacting with their cognate antibodies. In particular, the tryptophan residues, whose substitution by alanine promoted improved binding by the antibody to the adjacent MPER epitope, may normally insert themselves into the viral membrane. In possible support of this hypothesis, the complementarity-determining region H3 fingers of 2F5 and 4E10 could interact with more than the MPER epitopes thus far characterized, including possibly the viral membrane.[24,43] Nevertheless, the increase in the exposure of the 2F5 and 4E10 epitopes observed in the simulations upon reduction of the hydrophobic character of the 21-mer MPER-derived peptide is likely to occur regardless of which specific hydrophobic sequestering mechanism might be operating.

The simulation indicates that the natural 21-mer MPER-derived peptide is highly flexible, consistent with the hypothesized large conformational reorganization that occurs in the MPER during the viral fusion process. One of the most populated conformers adopts an α-helical conformation, consistent with the conformation of the crystal structure of the late-fusion state of gp41 and hypothesized to play a major role in the viral entry process.[37,44,45] During the fusion process, it is believed that the outer layer of the six-helix bundle, corresponding to the MPER of gp41, makes extensive contacts with

Table 2. Degree of pairwise hydrophobic packing between pairs of tryptophan residues measured in the REMD simulation of the LLELDKWANLWNWFDISNWLW peptide among conformations in which the structure of the DKW motif matched that of the corresponding residues in complex with 2F5

	W670	W672	W678	W680
W666	0.72	0.77	0.22	0.46
W670		0.96	0.56	0.40
W672			0.64	0.29
W678				0.88

Values closer to 1 indicate a higher occurrence of hydrophobic interactions. Sequence numbering follows the gp41 nomenclature.

the viral membrane, mediated by the numerous tryptophan residues present in this region. We indeed observed that, in the helical conformers (Cluster 1) obtained from the simulation of the 21-mer natural peptide, the side chains of the aromatic residues (especially W666, W670, and W672) are often located on the same side of the helix, in an orientation consistent with the hypothesized propensity of these residues to be localized at the membrane–protein interface.[46]

Insertion of the quadruple alanine-substituted 21-mer MPER-derived peptide into HRV (cHRV2) results in poor exposure of the epitope

Longer MPER sequences are more likely to reproduce the natural conformation of the MPER as part of the HIV spike complex. Simulations of the structures of the quadruple alanine-substituted 21-mer peptide (LLELDKWANLANAFDISNALA) in solution resulted in an increase in the exposure of the critical W666 side chain and an increase in the fraction of β-turn conformations of the DKW region of the epitope (characteristic of the crystal structure of the complex with 2F5). In an attempt to exploit both of these potential benefits in the design of the chimeric virus, we decided to model a chimeric HRV (cHRV2) in which the 21-mer alanine-mutated peptide was inserted into the VP2 puff of the parent HRV14 with removal of the VP2 residues S158-A159-N160-E161, as was performed with cHRV1. The rationale behind the alanine substitutions was to avoid the formation of intramolecular hydrophobic clusters involving W666. Further analysis of the REMD trajectory, however, showed that the majority of the conformations of the engineered loop were not compatible with 2F5 recognition (Fig. 2b). The other shortcoming that we observed was the relatively limited exposure of the region of the engineered loop that was critical for 2F5 binding (Fig. 7). This is likely due to the tendency of the loop to fold over onto the HRV capsid, presumably due to the length of this hydrophobic insert. This result indicates that, although the quadruple mutant peptide often adopts a favorable conformation for antibody binding in solution, it adopts a less favorable conformation when linked to HRV.

The structure of the 7-mer ELDKWAS peptide inserted into HRV (cHRV3) matches that of the peptide in the 2F5/antibody complex

In an attempt to restructure the DKW core epitope and to increase the likelihood of best exposing this critical region at the insert/capsid interface, we reduced the length of the insert. To this end, we modeled a series of shorter engineered loops, the most promising of which resulted in a third chimeric construct (cHRV3). In this construct, seven residues of the 2F5 epitope (ELDKWAS) were inserted into the VP2 puff of HRV14. In contrast to cHRV1, this chimeric construct was allowed to retain the four HRV14 VP2 residues that were previously removed (S158, A159, N160, and E161), and the insert was placed between N160 and E161.

The REMD simulation of cHRV3 showed that the engineered loop often adopts a structure likely to be recognized by 2F5 (Fig. 2c). As much as 41% of the modeled conformations in the ELDKWAS core epitope are within 2 Å C^{α} RMSD of the crystal structure of the corresponding peptide complexed with the 2F5 Fab. In these conformations, the ELDKWAS insert protrudes from the body of the virus capsid, extending approximately 19 Å from the base to the tip (Fig. 1b). The loop contains 37 residues from VP2 (residues L156-H193) and includes the 7-residue 2F5 core epitope sequence (ELDKWAS) at its center, corresponding to the tip of the loop. The engineered loop can be subdivided into three structural regions: the N-terminal base, a

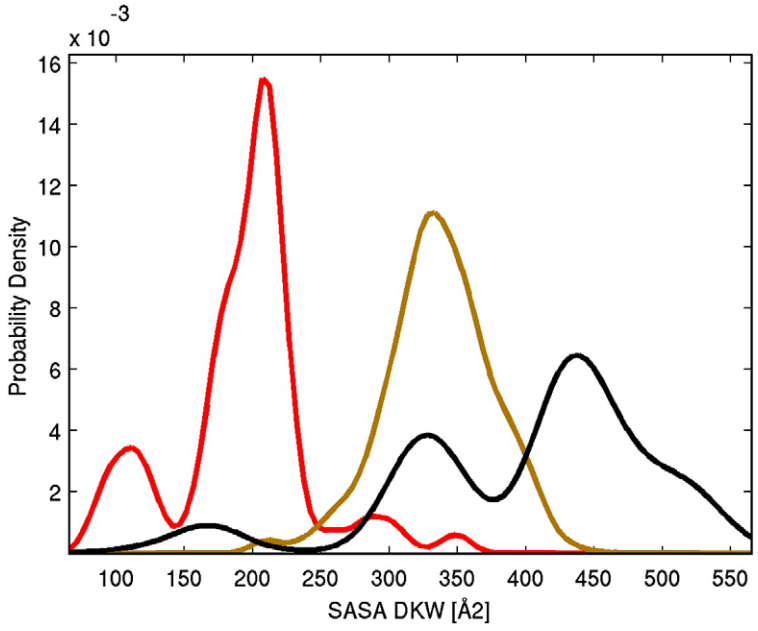

Fig. 7. Distribution of the SASA of the DKW core motif for the three chimeric constructs: cHRV1 (red), cHRV2 (brown), and cHRV3 (black). In cHRV3, the core of the epitope is significantly more solvent-exposed than those in the other two constructs.

flexible body extended away from the base, and the C-terminal edge, which is folded within the base of the VP2 subunit. The REMD model suggests that a network of hydrogen-bonding interactions at the two bases of the loop may account for the favorable positioning of the ELDKWAS motif at the tip of the loop (Fig. 8). The carbonyl oxygen of E662 (for the inserted peptide, the residue numbers follow HIV-1 gp41 numbering, and the residues of VP2 conserve the wild-type numbering) hydrogen bonds with the hydroxyl group of S668 of the same loop (Fig. 8), and VP2 H135 forms hydrogen bonds at the N-terminal and C-terminal ends of the inserted peptide with the carbonyl oxygen of N160 and the amide nitrogen of G164, respectively. Also, K167 forms hydrogen bonds with the carbonyl oxygen of A159 and with Oε1 of E161 at the two ends of the epitope.

Similar stabilizing interactions are not seen in the cHRV1 and cHRV2 constructs, which, however, are characterized by alternative interactions that prevent correct positioning of the core epitope. For example, in cHRV1 (Fig. 1a), the backbone of the linker residue P158 of the N-terminal PCG linker region (numbered according to the VP2 wild-type numbering) upstream of the 7-residue ELDKWAS insertion is often hydrogen-bonded to W666, causing it to fold inward. This proline residue also appears to induce an unfavorable backbone conformation of the engineered loop. Similarly, the amino hydrogen and carbonyl oxygen of C159 of the same linker form hydrogen bonds with Oε1 of S668 and with Hδ1 of H135, respectively, limiting the exposure of the inserted epitope. This pattern of interactions causes lengthening of the distance between the N-terminal and C-terminal ends of the engineered loop in cHRV1 relative to cHRV3 (approximately 9 Å in cHRV1 *versus* 6 Å in cHRV3), thereby favoring more extended and less exposed conformations of the epitope in cHRV1.

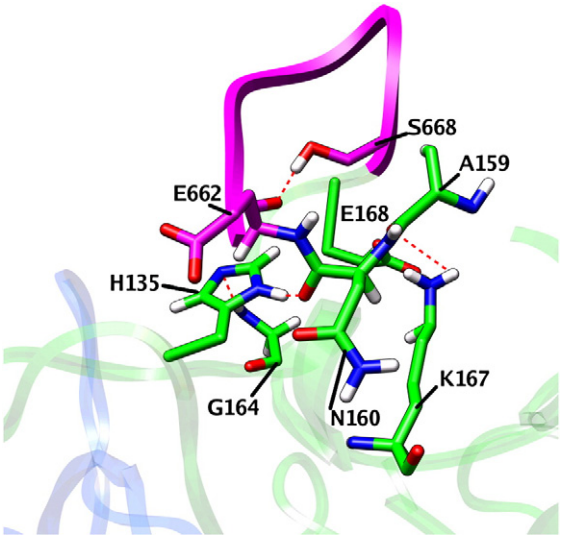

Fig. 8. Representation of the hydrogen-bonding network in the insertion region of the cHRV3 model. Hydrogen bonds are indicated by red dashed lines. The inserted peptide and the VP2 are drawn in magenta and green, respectively. The residues involved in the hydrogen-bonding network are shown in stick representation.

In the simulation with cHRV3, we observe little global deviation of the system from the crystal structure of wild-type HRV14. Superimposition based on C^{α} atoms reveals a relative RMSD of 0.86 Å of the VP2 subunit between the representative simulated conformation and the crystal structure. Most of the displacement occurs at the insertion site, between VP2 residues N160 and E161 of the VP2 puff. It appears that the HRV carrier is stable, owing to the plasticity of the insertion region that accommodates the insertion without major disruptions of native interactions.

The DKW region of the epitope at the tip of the loop was found primarily to adopt a β-turn conformation (Fig. 1b). Previous work on a cyclic peptide incorporating a lactam bridge between the two proximal amino acids flanking the DKW segment[47] suggests that, even though the type I β-turn structural motif is the basic feature for recognition by 2F5, a high binding affinity requires a specific cation–π stacking interaction between K665 and W666, which is also observed in our simulation. We have also analyzed the distribution of solvent-accessible surface areas (SASAs) of the core DKW segment from the ensemble of conformations generated by the simulation. Comparison of the SASA distribution of cHRV3 with those of cHRV1 and cHRV2 (Fig. 7) shows that, in cHRV3, the foreign epitope has significantly greater exposure. This suggests that the epitope would have a higher probability of protruding deeply into the binding cleft of the concave paratope of 2F5. This is in contrast to the W666 side chain of the 7-mer ELDKWAS peptide in solution, which assumed a conformation that was rotated inward, with the side chain sequestered in an intramolecular hydrophobic core without significant reorganization of the predominantly β-turn backbone structure or the α-helical backbone structure (Fig. 4a and b). The occurrence of these differing structures, combined with the observed inward rotation of W666 in the 21-mer peptide (Fig. 5), suggests that there may be a common hinge mechanism occurring during antibody recognition and during some of the other intermolecular interactions characterizing this dynamic region of gp41.

Structural model of the antibody/cHRV3 complex

Noting that the structure of the DKW segment of cHRV3 matched that of the corresponding segment in the crystal structure of the 2F5 complex in a significant fraction of conformations, we have constructed a structural model of the complex of cHRV3 with the 2F5 nAb by superimposing the ELDKWAS residues of a representative conformation of cHRV3 to the corresponding residues of the 1TJG crystal structure of the 2F5 complex (Fig. 9). The structure so obtained exhibits a remarkable shape complementarity between the modeled epitope and the 2F5 paratope from the crystal structure. The side chains

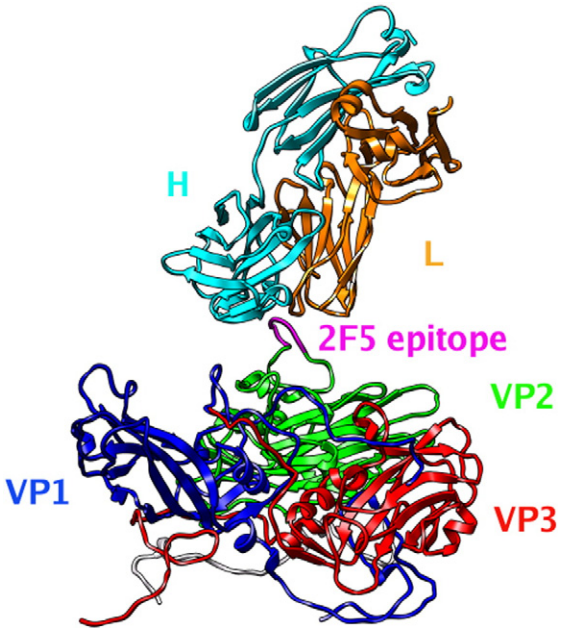

Fig. 9. Representative structure of part of the cHRV3 chimeric virus from the REMD simulation. The ELDKWAS insert (magenta) is connected to the VP2 protomeric subunit of cHRV3 (green). The VP1, VP3, and VP4 protomeric units are shown in blue, red, and gray, respectively. The Fab fragment of the 2F5 antibody is shown nearby to illustrate how it might bind to the epitope in the context of the virus. The heavy chain of 2F5 is shown in cyan, and the light chain of 2F5 is shown in orange.

of the DKW segment project inside the 2F5 hydrophobic crevice, as they do in the crystal structure. The structure also indicates a high degree of charge complementarity between the paratope and the epitope, and the buried surface is mainly composed of hydrophobic amino acids. The epitope occupies a cavity that is approximately 10 Å deep and 14 Å wide at the interface between the light chains and the heavy chains of the antibody hypervariable region. The intermolecular contacts between cHRV3 and 2F5 include salt bridges and hydrophobic interactions. The lysine side chain of K665 points to a hydrophobic pocket and is proximal (4.3 Å) to D54 of the heavy chain. The D664 side chain forms interactions (3.9 Å) with R95 of the heavy chain. Packing interactions can also be observed between the rings of W666 and F32 of the heavy chain. The side-chain positions and relative orientations of K665 and W666, which are both critical for binding, are reproduced remarkably well in the modeled complex. The same is true for the D664 side chain.

Conclusions

We have used the REMD conformational sampling method to study the conformational distributions of peptides and chimeric constructs derived from the immunogenic MPER of the gp41 protein of the HIV-1 envelope spike. The main aim of the work was to find chimeric constructs based on a rhinovirus carrier (HRV14) that is able to present the HIV-1 ELDKWA epitope with a high fraction of antibody-binding-competent conformations. Such constructs are predicted to bind with high affinity to the 2F5 antibody and thus to have greater likelihood to elicit, in turn, the production of antibodies that are broadly neutralizing and with similar characteristics. We have identified a promising HRV14:HIV-1 chimeric construct (cHRV3) that was obtained by inserting the ELDKWAS sequence into the VP2 puff of HRV14. The predicted structure of the ELDKWA epitope of this construct matched that of the crystal structure of the corresponding peptide complexed with 2F5 with high frequency. The close similarity of many conformations generated by REMD to the crystal structure allowed us to create a structural model of the complex of the chimeric HRV14 with the 2F5 antibody. The binding-competent conformations of the epitope, when inserted into HRV14, are stabilized by a network of specific hydrogen bonds, as well as by a number of specific hydrophobic interactions and by the absence of competing electrostatic and hydrophobic interactions that unfavorably affect the other constructs that we have investigated.

The results of this work complement published work and ongoing biochemical and spectroscopic research aimed at structure-based vaccine design through the investigation of structural propensities of MPER-derived peptides in different environments,[37,40,48] including insertion onto protein carriers.[18,49,50] The relevance of the present modeling study extends also to research aimed at uncovering the structural properties of the gp41 subunit as part of the HIV-1 envelope spike and the mechanisms by which the relevant structures elicit broadly neutralizing immunological activity. Our results confirm the hypothesis that the MPER of gp41 is highly flexible and capable of assuming a variety of conformations consistent with its biological function, and suggest that epitope exposure in gp41 is countered by intramolecular hydrophobic interactions. The latter observation offers an explanation for the finding[38] that mutations of hydrophobic residues that are distant in sequence and not directly involved in antibody binding, by weakening these hydrophobic interactions, can substantially increase the affinity of the 2F5 and 4E10 nAbs for gp41. The conformational heterogeneity of this region of gp41 and its interactions with components of the envelope spike and with the viral membrane likely play important roles in the processes of antibody recognition and antibody maturation.

The computational methodology introduced in this work is generally applicable to the *de novo* design of recombinant vaccines when, such as in the present case, structural information on a target is available. By identifying the factors that influence the presentation of the epitope toward the target, modeling work can guide the selection of vaccine candidates. The physical insights concerning the conformational propensities of epitopes inserted

into HRV14 provided by this work will facilitate the designs of focused combinatorial libraries of HRV14-based chimeras. The main design principles that emerged from this study are as follows: limit the size of the insert, optimize the central positioning of the critical DKW motif within the insert, and minimize the hydrophobic character and structural constraints determined by the flanking sequence. These focused libraries will be assayed in terms of their affinity for the 2F5 nAb and their ability to elicit strong and broadly neutralizing immune responses. This work is currently under way in our laboratory.

Materials and Methods

Replica exchange molecular dynamics

The temperature REMD method,[29,33,51] as implemented in the IMPACT computational package,[52] was used in this work to explore the conformational space of several peptides derived from the MPER of HIV gp41 in solution, as well as inserted into the HRV14 viral capsid. REMD is an advanced canonical conformational sampling algorithm that was designed to help overcome the sampling problem encountered in biomolecular simulations. It consists of running a series of MD simulations of the molecular system in parallel over multiple processors at different temperatures. A Monte Carlo move, aimed at swapping atomic coordinates of one replica with those of another, is attempted periodically. The acceptance probability expression for the swap move is derived based on the detailed balance condition. The lowest-temperature replica corresponds typically to the temperature of interest, whereas the high-temperature replicas provide a means for rapidly exploring the extent of conformational space. In the temperature replica exchange method, efficient interconversion between low-energy conformations occurs, whereby the temperature switches from low to high, allowing the conformation to overcome potential energy barriers and then transition to another conformation. A new low-energy conformation can then be established when this conformation transfers back to low-temperature replicas.

Force field and implicit solvation model

All simulations employed the optimized potential for liquid simulations all-atom fixed charge force field[26,27] and the AGBNP[28] implicit solvent model to mimic the water environment. Simulations using implicit solvent models are not as computationally intensive as those using explicit solvent models; this is particularly true for REMD simulations because fewer replicas are needed to achieve reasonable replica exchange rates due to the smaller system size and broader overlap of the energy distributions among replicas at different temperatures. AGBNP is an implicit solvent model based on a novel implementation of the pairwise descreening scheme of the generalized Born model[53,54] for the electrostatic component and a novel nonpolar hydration-free energy estimator. The nonpolar term consists of an estimator for the solute–solvent van der Waals dispersion energy designed to mimic explicit solvent–solute–solvent van der Waals interaction energies, in addition to a surface area term corresponding to the work required for cavity formation. It has been used to study protein conformational equilibrium and protein allostery.[31]

Peptide simulations

The REMD simulations in the solution of gp41-derived peptides were started from the fully extended conformation. Before the performance of REMD simulations, the system was equilibrated at each temperature. The temperature ladder was adjusted to obtain sufficient overlap of the potential energy distributions at neighboring temperatures to ensure frequent acceptance of temperature exchange moves. The REMD simulations of the ELDKWAS peptide and its alanine-substituted mutants were carried out using eight replicas for 20 ns (total production run, 160 ns) at temperatures of 313, 328, 345, 363, 381, 400, 421, and 442 K. Convergence progress was analyzed by monitoring the total potential energy and key interatomic distances as functions of time. Helical conformations were determined, using the program STRIDE,[55] for each residue in the trajectory of the 313-K replica. Average helicity is obtained by averaging the fractional helical populations of the residues in the sequence. REMD simulations of the LLELDKWANLWNWFDISNWLW peptide and its mutants were similarly performed using replicas at 313, 328, 347, 367, 388, 411, 434, 460, 486, and 515 K.

Model preparation

The three HRV14:HIV chimeric constructs investigated in this work were obtained by inserting gp41-derived epitopes into the βE–βF region of the so-called VP2 puff of the VP2 subunit of the HRV14 protomeric unit, the largest of three loops forming the neutralizing immunogenic site II of HRV14.[56,57] The first chimera (designated as cHRV1 in this work and as 14-C40-1 in Arnold *et al.*, submitted) was obtained by removing four residues of the VP2 puff (S158, A159, D160, and N161) and replacing them with the "ELDKWA" epitope sequence (ALDKWAS), flanked by PCG on the amino-terminal side of the epitope and by SPDCS on the carboxyl-terminal side of the epitope (Table 1). The second chimeric construct (cHRV2) was obtained by inserting the sequence LLELDKWANLANAFDISNALA, a quadruple alanine-substituted MPER peptide corresponding to residues 660–680, into the same HRV site as for cHRV1, except for the absence of the flanking linkers PCG and SPDCS (Table 1). The third chimeric construct (cHRV3) containing the sequence ELDKWAS was inserted between N160 and E161 of VP2, without the removal of any HRV14 residues or the addition of any flanking residues. Starting structures for the cHRV1, cHRV2, and cHRV3 chimeras have been obtained by homology modeling using the program Prime (Schrodinger, Inc.) and the structure of wild-type HRV14 (Protein Data Bank ID 4RHV) as structural template (97% sequence identity). These models can be built with a high degree of confidence, since the target and template structures are expected to differ only in the proximity of the inserted epitope—an HRV region that is only 12 residues long. Therefore, the estimated confidence factor for the homology models was high. The energy-minimized models have been assessed for quality using ProCheck, indicating 84.6%, 14.6%, 0.7%, and 0.1% in the most favored regions, allowed regions, generously allowed regions, and disallowed regions of the Ramachandran plot,[58] respectively.

Nine peptides were used for these studies (Table 1). All have NH^{3+} groups at their amino termini and COO^{-}

groups at their carboxy termini. There were seven natural and singly alanine-substituted (corresponding to HxB2 gp41 residues 662–668) 7-mer peptides, and two natural and quadruply alanine-substituted (corresponding to HxB2 residues 660–680) 21-mer peptides.

REMD simulations of the chimeric constructs

REMD simulations of the cHRV1, cHRV2, and cHRV3 viral constructs were conducted to estimate the conformational propensities of the inserts. The production REMD runs consisting of 30 replicas, with temperatures in the range 310–495 K, were 2 ns/replica in length, for a total of 60 ns for each of the three systems. The production runs were preceded by minimization and equilibration calculations starting from the comparative models described above. To reduce computational complexity, the residues within all atoms farther than 20 Å from any atom of the foreign insert were not included in the model. Because of the different sizes of the inserts, the numbers of residues that were completely free were 227, 276, and 176 in cHRV1, cHRV2, and cHR3, respectively. Portions of the VP1 and VP3 protomeric units within 20 Å of any atom of the insert of the viral capsid were included as well. The positions of C^{α} atoms of the residues at the ends of dangling protein chains were harmonically restrained using an isotropic force constant of 0.3 kcal/mol/Å^2. This setup prevents unfolding of the simulated protein region at high temperature, while providing a reasonably accurate description of the protein environment surrounding the inserted sequence. This methodology was tested by applying it to the chimeric HRV displaying part of the HIV-1 V3 loop studied by Smith *et al.*[17] We confirmed that the main cluster of conformations of the loop generated by REMD structurally matched that of the crystal structure.

Hydrophobic packing measure

An algorithm aimed at quantitative characterization of the degree of packing interactions among residues has been designed. The measure is based on the ratio:

$$x_{ij} = \frac{R_g(i,j)}{\sqrt{R_g(i)R_g(j)}}$$

between the radius of gyration $R_g(i,j)$ of a residue pair and the geometric mean of the radii of gyration $R_g(i)$ and $R_g(j)$ of the individual residues. A small value of x_{ij} approaching 1 indicates that the two residues are in contact. Conversely, a value of x_{ij} much larger than 1 indicates that the two residues are distant from each other relative to their sizes. A normalized degree of packing, $f(x_{ij})$, ranging between 0 (residues are not in contact) and 1 (residues are in contact) is obtained by applying the following linear interpolation function:

$$f(x) = \begin{cases} 1 & x \le x_{\min} \\ ax + b & x_{\min} < x < x_{\max} \\ 0 & \text{otherwise} \end{cases}$$

where

$$a = \frac{-1}{x_{\max} - x_{\min}}$$

and

$$b = \frac{x_{\max}}{x_{\max} - x_{\min}}.$$

The function above returns to 1 for x_{ij} smaller than $x_{\min}$, to 0 for x_{ij} larger than $x_{\max}$, and to intermediate values for x_{ij} between these two limits. Here we set $x_{\min} = 2.5$ and $x_{\max} = 3.5$.

The degree of packing $\langle f(x_{ij}) \rangle$ between residues i and j, averaged over an REMD trajectory, measures the fractional propensity of the two residues to be in contact. The closer is this quantity to 1, the higher is the occurrence of packing between a pair of residues. We used this measure to analyze the occurrence of intramolecular hydrophobic contacts in the simulation of peptides derived from the MPER of gp41. In that context, we applied the packing measure only to those conformations in which the most essential ELDKWA residues, DKW, assumed a conformation similar to that obtained from the crystal structure 1TJG based on a 30° threshold maximum difference between the backbone dihedral angles.

Solvent-accessible surface calculation

The SASA[59] was determined to quantify the conformational exposure of the engineered loop insert on the chimeric HRV. The SASA has been calculated for each atom using the program SURFACE based on the algorithm of Sridharan *et al.*, with a water probe radius of 1.4 Å.[60] Hydrogen atoms have been ignored in the calculation of the SASA. The distribution of the SASA for the ELDKWA epitope residues DKW has been determined for the structures present in the replica exchange trajectories of the chimeric constructs cHRV1, cHRV2, and cHRV3.

Hierarchical clustering

The computed conformational ensembles have been organized into conformational clusters using the XCluster program (Schrodinger, Inc.). XCluster implements a hierarchical clustering algorithm[61] based on the minimal RMSD between user-defined sets of atoms of two conformations. The clustering procedure starts by computing the all-to-all distance matrix between the selected sets of conformations. A series of clustering arrangements, or clustering levels, is then produced. At the first level, each of the N conformations forms a separate cluster. The following level with $N-1$ clusters is obtained by joining into a single cluster the pair of conformations closest to each other. This process continues by joining at each step the closest pair of clusters until, at the Nth level, only one cluster encompassing the entire ensemble remains. XCluster employs the single-linkage convention to compute the distance between two clusters, defined as the smallest distance between any two conformations of the two clusters. Each clustering level i corresponds to $N-i$ clusters. The choice of the clustering level to pick is aided by clustering quality measures computed by XCluster such as the minimum separation ratio, which is the smallest ratio between the shortest intracluster and intercluster distances among all clusters. Most of the results obtained here have been obtained at the clustering level corresponding to approximately 10 clusters.

Acknowledgements

This work has been supported, in part, by National Institutes of Health grants GM30580 (to R.M.L.) and 1AI071874 (to G.F.A.). The computer

simulations for this work have been performed at the BioMaPS High Performance Computing Center at Rutgers University funded in part by the National Institute of Health shared instrumentation grant no. 1 S10 RR022375.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.10.089

References

1. Klausner, R. D., Fauci, A. S., Corey, L., Nabel, G. J., Gayle, H., Berkley, S. *et al.* (2003). Medicine. The need for a global HIV vaccine enterprise. *Science*, **300**, 2036–2039.
2. Montero, M., van Houten, N. E., Wang, X. & Scott, J. K. (2008). The membrane-proximal external region of the human immunodeficiency virus type 1 envelope: dominant site of antibody neutralization and target for vaccine design. *Microbiol. Mol. Biol. Rev.* **72**, 54–84.
3. Burton, D. R., Desrosiers, R. C., Doms, R. W., Koff, W. C., Kwong, P. D., Moore, J. P. *et al.* (2004). HIV vaccine design and the neutralizing antibody problem. *Nat. Immunol.* **5**, 233–236.
4. Eckert, D. M. & Kim, P. S. (2001). Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc. Natl Acad. Sci. USA*, **98**, 11187–11192.
5. Weng, Y., Yang, Z. & Weiss, C. D. (2000). Structure–function studies of the self-assembly domain of the human immunodeficiency virus type 1 transmembrane protein gp41. *J. Virol.* **74**, 5368–5372.
6. Weissenhorn, W., Calder, L. J., Dessen, A., Laue, T., Skehel, J. J. & Wiley, D. C. (1997). Assembly of a rod-shaped chimera of a trimeric GCN4 zipper and the HIV-1 gp41 ectodomain expressed in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **94**, 6065–6069.
7. Gabuzda, D. H., Lever, A., Terwilliger, E. & Sodroski, J. (1992). Effects of deletions in the cytoplasmic domain on biological functions of human immunodeficiency virus type 1 envelope glycoproteins. *J. Virol.* **66**, 3306–3315.
8. Roux, K. H. & Taylor, K. A. (2007). AIDS virus envelope spike structure. *Curr. Opin. Struct. Biol.* **17**, 244–252.
9. Lu, M., Blacklow, S. C. & Kim, P. S. (1995). A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nat. Struct. Biol.* **2**, 1075–1082.
10. de Planque, M. R., Kruijtzer, J. A., Liskamp, R. M., Marsh, D., Greathouse, D. V., Koeppe, R. E., II *et al.* (1999). Different membrane anchoring positions of tryptophan and lysine in synthetic transmembrane alpha-helical peptides. *J. Biol. Chem.* **274**, 20839–20846.
11. Muster, T., Steindl, F., Purtscher, M., Trkola, A., Klima, A., Himmler, G. *et al.* (1993). A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J. Virol.* **67**, 6642–6647.
12. Zwick, M. B., Labrijn, A. F., Wang, M., Spenlehauer, C., Saphire, E. O., Binley, J. M. *et al.* (2001). Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. *J. Virol.* **75**, 10892–10905.
13. Zwick, M. B. & Burton, D. R. (2007). HIV-1 neutralization: mechanisms and relevance to vaccine design. *Curr. HIV Res.* **5**, 608–624.
14. Dimitrov, A. S., Rawat, S. S., Jiang, S. & Blumenthal, R. (2003). Role of the fusion peptide and membrane-proximal domain in HIV-1 envelope glycoprotein-mediated membrane fusion. *Biochemistry*, **42**, 14150–14158.
15. Munoz-Barroso, I., Salzwedel, K., Hunter, E. & Blumenthal, R. (1999). Role of the membrane-proximal domain in the initial stages of human immunodeficiency virus type 1 envelope glycoprotein-mediated membrane fusion. *J. Virol.* **73**, 6089–6092.
16. Resnick, D. A., Smith, A. D., Gesiler, S. C., Zhang, A., Arnold, E. & Arnold, G. F. (1995). Chimeras from a human rhinovirus 14-human immunodeficiency virus type 1 (HIV-1) V3 loop seroprevalence library induce neutralizing responses against HIV-1. *J. Virol.* **69**, 2406–2411.
17. Smith, A. D., Geisler, S. C., Chen, A. A., Resnick, D. A., Roy, B. M., Lewi, P. J. *et al.* (1998). Human rhinovirus type 14:human immunodeficiency virus type 1 (HIV-1) V3 loop chimeras from a combinatorial library induce potent neutralizing antibody responses against HIV-1. *J. Virol.* **72**, 651–659.
18. Ding, J., Smith, A. D., Geisler, S. C., Ma, X., Arnold, G. F. & Arnold, E. (2002). Crystal structure of a human rhinovirus that displays part of the HIV-1 V3 loop and induces neutralizing antibodies against HIV-1. *Structure*, **10**, 999–1011.
19. Ghiara, J. B., Ferguson, D. C., Satterthwait, A. C., Dyson, H. J. & Wilson, I. A. (1997). Structure-based design of a constrained peptide mimic of the HIV-1 V3 loop neutralization site. *J. Mol. Biol.* **266**, 31–39.
20. Hagmann, M. (2000). Computers aid vaccine design. *Science*, **290**, 80–82.
21. Oomen, C. J., Hoogerhout, P., Bonvin, A. M., Kuipers, B., Brugghe, H., Timmermans, H. *et al.* (2003). Immunogenicity of peptide-vaccine candidates predicted by molecular dynamics simulations. *J. Mol. Biol.* **328**, 1083–1089.
22. Nuzzaci, M., Piazzolla, G., Vitti, A., Lapelosa, M., Tortorella, C., Stella, I. *et al.* (2007). Cucumber mosaic virus as a presentation system for a double hepatitis C virus-derived epitope. *Arch. Virol.* **152**, 915–928.
23. Nair, D. T., Singh, K., Siddiqui, Z., Nayak, B. P., Rao, K. V. & Salunke, D. M. (2002). Epitope recognition by diverse antibodies suggests conformational convergence in an antibody response. *J. Immunol.* **168**, 2371–2382.
24. Ofek, G., Tang, M., Sambor, A., Katinger, H., Mascola, J. R., Wyatt, R. & Kwong, P. D. (2004). Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. *J. Virol.* **78**, 10724–10737.
25. Ma, B., Shatsky, M., Wolfson, H. J. & Nussinov, R. (2002). Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci.* **11**, 184–197.
26. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. (2001). Evaluation and reparameterization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, **105**, 6474–6487.
27. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236.
28. Gallicchio, E. & Levy, R. M. (2004). AGBNP: an analytic implicit solvent model suitable for molecular

dynamics simulations and high-resolution modeling. *J. Comput. Chem.* **25**, 479–499.
29. Felts, A. K., Harano, Y., Gallicchio, E. & Levy, R. M. (2004). Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. *Proteins*, **56**, 310–321.
30. Ravindranathan, K. P., Gallicchio, E., Friesner, R. A., McDermott, A. E. & Levy, R. M. (2006). Conformational equilibrium of cytochrome P450 BM-3 complexed with *N*-palmitoylglycine: a replica exchange molecular dynamics study. *J. Am. Chem. Soc.* **128**, 5786–5791.
31. Ravindranathan, K. P., Gallicchio, E. & Levy, R. M. (2005). Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein. *J. Mol. Biol.* **353**, 196–210.
32. Felts, A. K., Paris, K., Gallicchio, E., Friesner, R. A. & Levy, R. M. (2008). Predicting long loops with the AGBNP implicit solvent model using hierarchical torsion angle sampling and protein local optimization. *J. Chem. Theor. Comp.* **4**, 855–858.
33. Sugita, Y. & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151.
34. Tan, K., Liu, J., Wang, J., Shen, S. & Lu, M. (1997). Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl Acad. Sci. USA*, **94**, 12303–12308.
35. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell*, **89**, 263–273.
36. Biron, Z., Khare, S., Samson, A. O., Hayek, Y., Naider, F. & Anglister, J. (2002). A monomeric 3(10)-helix is formed in water by a 13-residue peptide representing the neutralizing determinant of HIV-1 on gp41. *Biochemistry*, **41**, 12687–12696.
37. Barbato, G., Bianchi, E., Ingallinella, P., Hurni, W. H., Miller, M. D., Ciliberto, G. *et al.* (2003). Structural analysis of the epitope of the anti-HIV antibody 2F5 sheds light into its mechanism of neutralization and HIV fusion. *J. Mol. Biol.* **330**, 1101–1115.
38. Zwick, M. B., Jensen, R., Church, S., Wang, M., Stiegler, G., Kunert, R. *et al.* (2005). Anti-human immunodeficiency virus type 1 (HIV-1) antibodies 2F5 and 4E10 require surprisingly few crucial residues in the membrane-proximal external region of glycoprotein gp41 to neutralize HIV-1. *J. Virol.* **79**, 1252–1261.
39. Ho, J., Uger, R. A., Zwick, M. B., Luscher, M. A., Barber, B. H. & MacDonald, K. S. (2005). Conformational constraints imposed on a pan-neutralizing HIV-1 antibody epitope result in increased antigenicity but not neutralizing response. *Vaccine*, **23**, 1559–1573.
40. Schibli, D. J., Montelaro, R. C. & Vogel, H. J. (2001). The membrane-proximal tryptophan-rich region of the HIV glycoprotein, gp41, forms a well-defined helix in dodecylphosphocholine micelles. *Biochemistry*, **40**, 9570–9578.
41. Ho, J., MacDonald, K. S. & Barber, B. H. (2002). Construction of recombinant targeting immunogens incorporating an HIV-1 neutralizing epitope into sites of differing conformational constraint. *Vaccine*, **20**, 1169–1180.
42. Joyce, J. G., Hurni, W. M., Bogusky, M. J., Garsky, V. M., Liang, X., Citron, M. P. *et al.* (2002). Enhancement of alpha-helicity in the HIV-1 inhibitory peptide DP178 leads to an increased affinity for human monoclonal antibody 2F5 but does not elicit neutralizing responses in vitro. Implications for vaccine design. *J. Biol. Chem.* **277**, 45811–45820.
43. Cardoso, R. M., Zwick, M. B., Stanfield, R. L., Kunert, R., Binley, J. M., Katinger, H. *et al.* (2005). Broadly neutralizing anti-HIV antibody 4E10 recognizes a helical conformation of a highly conserved fusion-associated motif in gp41. *Immunity*, **22**, 163–173.
44. Salzwedel, K., West, J. T. & Hunter, E. (1999). A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. *J. Virol.* **73**, 2469–2480.
45. Suarez, T., Gallaher, W. R., Agirre, A., Goni, F. M. & Nieva, J. L. (2000). Membrane interface-interacting sequences within the ectodomain of the human immunodeficiency virus type 1 envelope glycoprotein: putative role during viral fusion. *J. Virol.* **74**, 8038–8047.
46. Reithmeier, R. A. (1995). Characterization and modeling of membrane proteins using sequence analysis. *Curr. Opin. Struct. Biol.* **5**, 491–500.
47. McGaughey, G. B., Citron, M., Danzeisen, R. C., Freidinger, R. M., Garsky, V. M., Hurni, W. M. *et al.* (2003). HIV-1 vaccine development: constrained peptide immunogens show improved binding to the anti-HIV-1 gp41 mAb. *Biochemistry*, **42**, 3214–3223.
48. Opalka, D., Pessi, A., Bianchi, E., Ciliberto, G., Schleif, W., McElhaugh, M. *et al.* (2004). Analysis of the HIV-1 gp41 specific immune response using a multiplexed antibody detection assay. *J. Immunol. Methods*, **287**, 49–65.
49. Coeffier, E., Clement, J. M., Cussac, V., Khodaei-Boorane, N., Jehanno, M., Rojas, M. *et al.* (2000). Antigenicity and immunogenicity of the HIV-1 gp41 epitope ELDKWA inserted into permissive sites of the MalE protein. *Vaccine*, **19**, 684–693.
50. Kim, M., Qiao, Z., Yu, J., Montefiori, D. & Reinherz, E. L. (2007). Immunogenicity of recombinant human immunodeficiency virus type 1-like particles expressing gp41 derivatives in a pre-fusion state. *Vaccine*, **25**, 5102–5114.
51. Gallicchio, E., Levy, R. M. & Parashar, M. (2007). Asynchronous replica exchange for molecular simulations. *J. Comput. Chem.* **29**, 788–794.
52. Banks, J. L., Beard, H. S., Cao, Y., Cho, A. E., Damm, W., Farid, R. *et al.* (2005). Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **26**, 1752–1780.
53. Qiu, D., Shenkin, P. S., Hollinger, F. P. & Still, W. C. (1997). The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A*, **101**, 3005–3014.
54. Bashford, D. & Case, D. A. (2000). Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**, 129–152.
55. Dmitrij Frishman, P. A. (1995). Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Genet.* **23**, 566–579.
56. Arnold, E. & Rossmann, M. G. (1990). Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 A. *J. Mol. Biol.* **211**, 763–801.
57. Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J. *et al.* (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature*, **317**, 145–153.
58. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to

check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
59. Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **221**, 709–713.
60. Sridharan, S. A., Nicholls, A. & Honig, B. (1992). A new vertex algorithm to calculate solvent accessible surface areas. *Biophys. J.* **61**, 995a.
61. Peter, S. & Shenkin, D. Q. M. (1994). Cluster analysis of molecular conformations. *J. Comput. Chem.* **15**, 899–916.