Folding Molecular Dynamics Simulation of a gp41-Derived Peptide Reconcile Divergent Structure Determinations
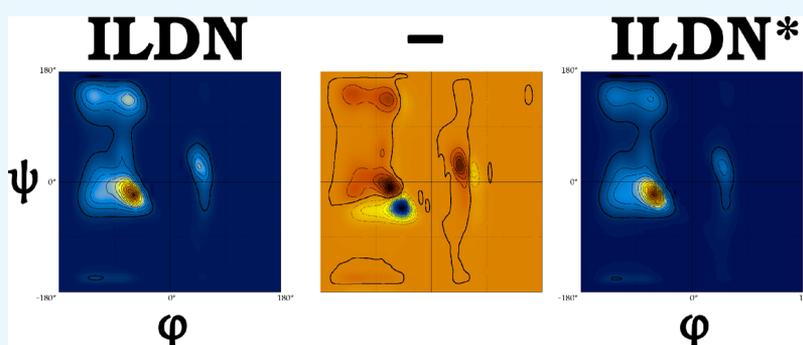
Panagiota S. Georgoulia*[ID] and Nicholas M. Glykos[ID]

Department of Molecular Biology and Genetics, Democritus University of Thrace, University Campus, Alexandroupolis 68100, Greece

S Supporting Information

ABSTRACT: T-20 peptide is the first FDA-approved fusion inhibitor against AIDS/HIV-1 gp41 protein. Part of it, the gp41[659−671] peptide, that contains the complete epitope for the neutralizing 2F5 monoclonal antibody, has been found experimentally in a number of divergent structures. Herein, we attempt to reconcile them by using unbiased large-scale all-atom molecular dynamics folding simulations. We show that our approach can successfully capture the peptide's heterogeneity and reach each and every experimentally determined conformation in sub-angstrom accuracy, whilst preserving the peptide's disordered nature. Our analysis also unveils that the minor refinements within the AMBER99SB family of force fields can lead to appreciable differences in the predicted conformational stability arising from subtle differences in the helical- and β-region of the Ramachandran plot. Our work underlines the contribution of molecular dynamics simulation in structurally characterizing pharmacologically important peptides of ambiguous structure.

1. INTRODUCTION

Deletions or mutations in the tryptophan-rich C-terminal region (residues 660−683) of the AIDS/HIV-1 gp41 protein can prevent fusion with the host cell membrane.[1,2] This region of gp41 encompasses the synthetic peptide T-20 that corresponds to residues 638−673, which has been approved by the FDA (under the names Fuzeon or Enfuvirtide) as the first fusion inhibitor targeting gp41 with an EC_{50} of 1 ng/mL denoting a new class of anti-HIV drugs that prevent infection of new cells.[3−7] Part of T-20 is the gp41[659−671] peptide which contains the complete epitope (sequence ELDKWA) for the broad-spectrum neutralizing 2F5 monoclonal antibody.[8−11] Notably, the affinity of the 2F5 antibody is reduced after the binding between gp120 and CD4, suggesting that the gp41[659−671] epitope is solvent exposed in the prefusogenic form but becomes less accessible or restructured after fusion.[12−14] It is for these reasons that the solution structure (in prefusogenic form) of the gp41[659−671] has been long recognised as being important for the development of new peptide antigens targeting AIDS/HIV infection.

The structure and dynamics of the gp41[659−671] peptide in solution are a source of controversy in the literature. The NMR/CD study by Anglister's group[15] showed that the peptide takes-up a mostly 3_{10}-helical conformation in solution and based on a sufficient number of NOE restraints they deposited their structure (PDB entry 1LCX), lending support to previously published CD data.[16] Later UV resonance Raman scattering studies revealed instead a variety of motifs, with half the peptide bonds in helical conformation (mainly 3_{10} and π-helix, less α-helix) and the rest in extended and unfolded conformations (mainly β-turn and polyproline II).[17] Follow-up studies by Pessi's group using a combination of experimental and computational approaches suggested that the initially proposed 3_{10}-helical structure by Anglister's group may be an error: their NMR (PDB entry 1MZI) and CD data suggested that the peptide is mostly disordered, spanning an ensemble of conformers, including helices and turns.[18] On top of that, Crain's group argued as well for a conformational plasticity using a combination of CD and MD data.[19,20] However, there was inconsistency even between the experimental findings with their MD data. The group chose for the simulations the CHARMM force field, which has repeatedly been demon-
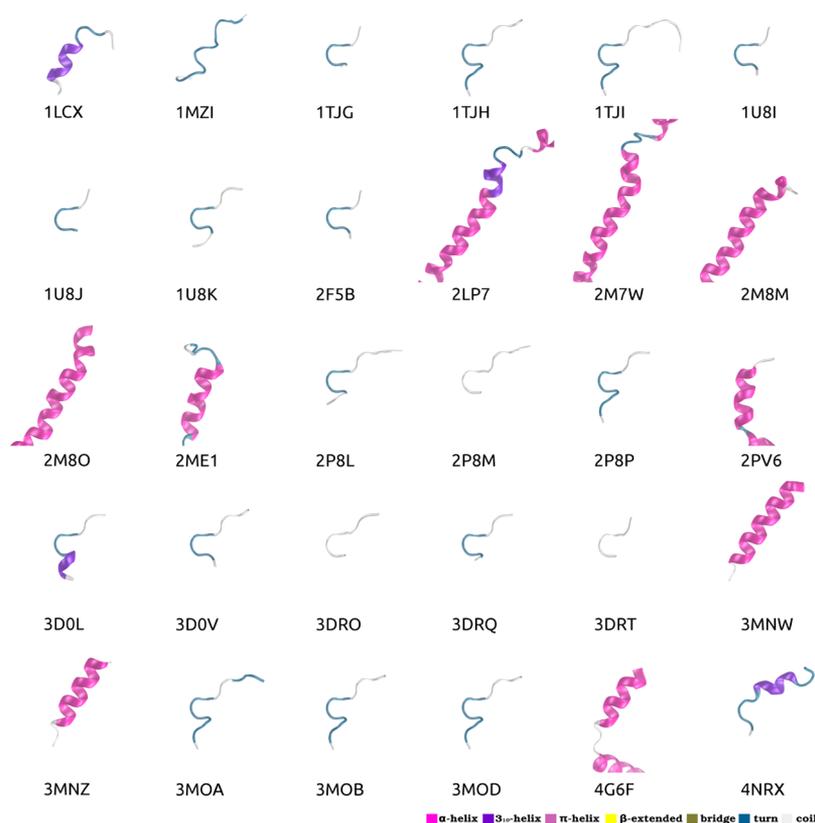
Figure 1. Structural diversity of the T20 peptide. Schematic diagram of a collection of peptide structures deposited in the PDB database that entail the sequence of the T20 peptide, corresponding to residues [659−671] of gp41. All structures are superimposed on the central ELDKWA epitope. Secondary structural elements are highlighted using cartoon representation and color coded using STRIDE-derived secondary structure assignments (as shown in the color-box). More information on the sequence context and the structure determination details are provided in Table S1.

strated to suffer from an α-helical bias.[21−27] The result was that only at elevated temperatures their MD data were in qualitative agreement with the experiment, whereas the β-turn motif suggested by Pessi's group was found unstable.[19] Furthermore in a later study, the peptide was found to exhibit autonomous helical folding in the presence of membrane mimicking sodium dodecyl sulfate micelles, which was tunable by pH variation, a behaviour that was not observed in aqueous solutions.[20]

Given the indications for the highly dynamic nature of the gp41[659−671] peptide in solution a direct experimental attack on elucidating the peptide's structure has been problematic. Herein, we aim to study the folding behaviour of this malleable peptide using unbiased large-scale all-atom molecular dynamics simulation. Using a previously validated force field[25−33] coupled with an adaptive tempering protocol and a sufficient simulation length to guarantee the convergence of the peptide's derived atomic structures we attempt to thoroughly characterise its structure and dynamics. We show that simulations can bridge the gap among divergent structure determinations by reaching in sub-angstrom accuracy every experimentally available peptide structure, whilst preserving its disordered nature. Detailed analysis of the simulations interestingly showed that even the best performing force fields still suffer from imbalance between the different secondary structural elements. This diversity seems to impact inadvertently the accurate structural characterisation of highly dynamic peptides such as gp41[659−671]. Our simulations contribute to a better characterisation of the structure and dynamics of this pharmacologically and therapeutically

important peptide and to the unceasing refinement of current generation force fields.

2. RESULTS AND DISCUSSION

2.1. Structural Heterogeneity of the T-20 Derived gp41[659−671] Peptide. Even a cursory examination of the collection of structures deposited in the PDB database (Figure 1 and Table S1) reveals that the structures of the T-20 derived peptide, from the 7-mer to the 17-mer, free (PDB entry 1LCX,[15] PDB entry 1MZI[18]) or in complex with the 2F5 antibody (PDB entries 1TJG, 1TJH, 1TJI[34]) can vary from helical to turn to almost fully extended.

Turn structures seem to be the dominant population when the peptide 7-mer, 9-mer, 11-mer or 13-mer is in complex with the 2F5 monoclonal antibody (PDB entries 1U8I, 1U8J, 1U8K, 2F5B, 2P8L, 2P8M, 2P8P, 3D0V, 3DRO, 3DRQ, 3DRT, 3MOA, 3MOB, 3MOD[35−37]). It is worth mentioning however the inconsistency among the studies regarding the resolution of the electron density data for the C-terminal residues [668]SLWN[671]. This was attributed to either the end-capping of the peptide through amidation that deprives the extra end-charge or to the extension of the end by just a couple of residues ([670]WN[671]). In both cases, the formation of a key hydrogen bond of the peptide's end with the heavy chain of the antibody stabilises the C-terminal part of the peptide.[36] This conformational stabilisation was found particularly persistent under high ionic strength conditions (PDB entry 3D0L[36]) mostly due to an additional intrachain hydrogen bond between the carbonyl oxygen of [666]Trp with the backbone amide of
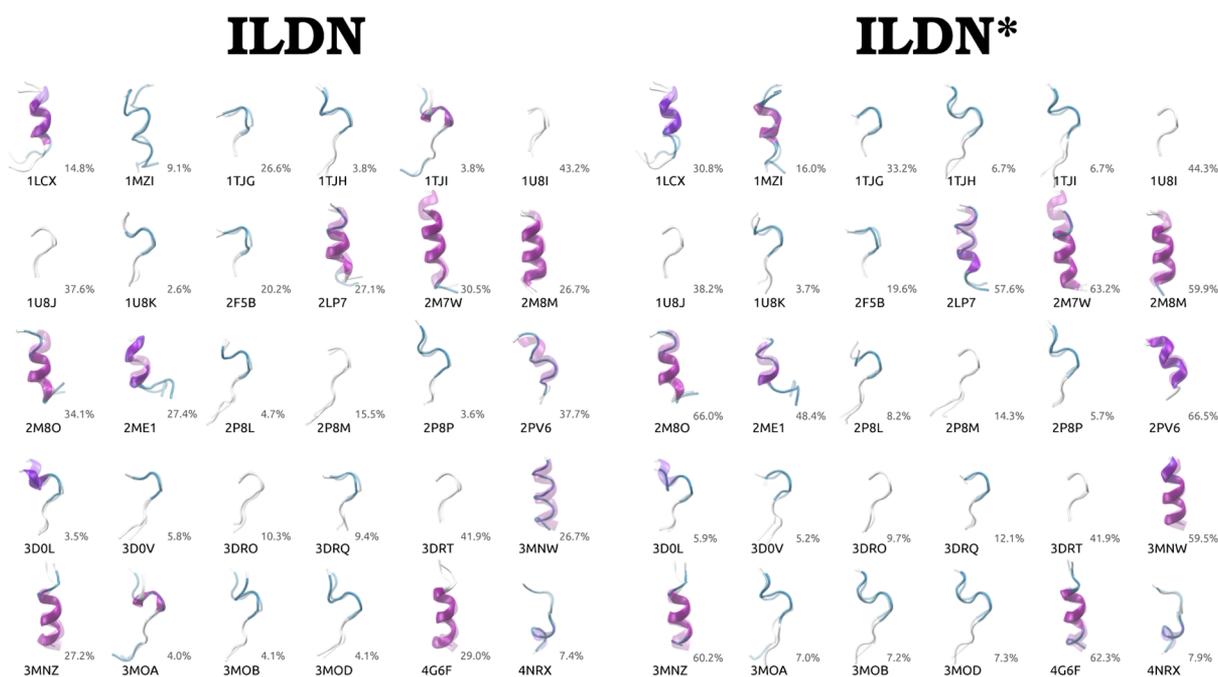
Figure 2. Similarity to experimental structures. Superposition of the closest simulation-derived structure for the ILDN (left) and ILDN* (right) force fields to each one of the experimentally known structures (see also Figure 1 and Table S1). Secondary structural elements are highlighted using cartoon representation and color coded using STRIDE-derived secondary structure assignments as in Figure 1. The experimental structures are overlayed in transparency for clarity. The occupancy of each ensemble of structures is noted next to each PDB code. The occupancies were calculated using an rmsd cut-off of 2.0 Å as determined from the rmsd histogram distributions in Figure S1, over the stable conformations of the trajectory ($T < 340$ K).

[669]Leu (i, $i + 4$), denoting the start of an α-helix extending through the C-terminus. This helix is also observed in the longer peptides (PDB entries 2LP7, 2M7W, 2M8M, 2M8O, 2ME1, 2PV6, 3MNW, 3MNZ, 4G6F, 4NRX[38−43]).

Indeed, helical structures seem to be favoured in the larger context like in the 59-mer, that is extended towards the N-terminus this time to include the membrane-proximal external region (MPER), especially in the lipid environment of the viral membrane, with a bend around the epitope to expose the amino acids involved in the antibody binding (PDB entries 2LP7, 2M7W, 4G6F, 4NRX[38,42,43]). A significant helical tendency for the gp41[657−669] peptide fragment in water was also found in later studies by the Anglister group in the longer gp41[636−677] peptide, with the preceding N-terminal part unstructured.[44]

Both findings regarding turn and helical preferences were confirmed by later studies in non-aqueous environments revealing the folding of the 28-mer gp41[656−683] to a continuous kinked helix (PDB entries 2M8M, 2M8O[39]) whilst shorter or longer peptides showed non-continuous partial structuring to 3_{10} and α-helix. In particular, the kink comprised residues [665]KW[666] preceded by 3_{10}-helical conformation for residues [661]LEL[663], consistent with the type I β-turn for the [664]DKW[666] core epitope sequence. This kink or tilt in the N-terminal α-helix (664−672, MPER region) that connects via a short hinge to the flat C-terminal helical segment (675−683) was observed by yet another group (PDB entries 2ME1, 2PV6[40,41]) and it was also observed when the peptide was in complex with another broad neutralising antibody, 10E8 (PDB entry 4G6F[42]). In this case, the N-terminal α-helix extends from residues 657−667, tightens to a 3_{10}-helix for [668]SL[669] followed by a turn for [670]WN[671] to continue to the C-terminal α-helix for residues 672−683. This longer peptide displayed

also an independent capability to interact with a membrane surface, with different segmental secondary structures being adopted during the fusion process. The bias towards helical folding seems to be pH tuned, with no defined fold type at basic pH and helical near physiological and acidic pH environment.[20]

The common characteristic amongst all helical structures is the exposure of five key residues [660]Leu, [663]Leu, [664]Asp, [665]Lys, [666]Trp, with the last three involved directly in antibody binding (pdb entry 4NRX[43]). It seems that antibody recognition requires a continuous helical structure interrupted by a flexible kink/turn/loop conformation for residues [664]DKW[666] that redirects the gp41 backbone at the pre-transmembrane region, thus orienting the [666]Trp and [669]Leu side-chains in parallel and the [664]Asp side-chain negative charge projecting from the main chain axis in different direction.[39,43] This crucial involvement of side-chain interactions gives the epitope specificity. In total, the conformation of residues [651]LELDKWAS[668] when bound to neutralizing antibodies is conserved in most observed structures, whereas the conformation of residues [669]LW[670] varies significantly, like in structures 2P8P, 1TJH, and 3D0L.[36] One of the most distinct structures is that of a 3_{10}-helix supported by the seminal study of Anglister's group.[15] This type of helix was attributed mainly to the strong presence of an i, $i + 3$ hydrogen bond between the carbonyl oxygen of [664]Asp and the amide proton of [667]Ala together with $^3J_{HN\alpha}$ coupling constants and chemical shifts supporting helical conformations and only minor populations of random coil, arguing that the presence of this hydrogen bond could be key to deter antibody interaction. Indeed it seems that the most potent short constrained peptides that inhibit HIV-1 entry are not the ones with high helical propensity.[45] Both proper [664]DKW[666] β-turn conformation and side-chain positioning are
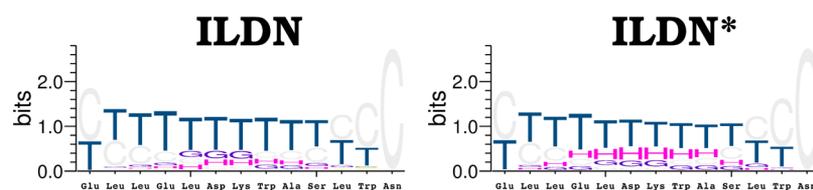
Figure 3. Secondary structure preferences of the gp41[659−671] peptide. Weblogo-like representation of the per residue STRIDE-derived secondary structure assignments for ILDN (left) and ILDN* (right) force fields, calculated for all the trajectory. The symbols used correspond to coil (C, grey), turn (T, blue), 3_{10}-helix (G, purple), α-helix (H, magenta), extended β (E, yellow) and bridge (B, tan).

crucial for designing immunogens capable of eliciting neutralizing antibodies.

2.2. Excellent Agreement between Experimental and Simulation-Derived Structures. The performance of the force fields against folding a peptide with such divergent structures is examined in Figure 2. By using a straightforward root-mean-square deviation (rmsd) calculation, we can identify simulation-derived structures that match each one of the experimental structures to sub-angstrom accuracy. Some of the structures are scarcely visited, as expected due to lesser secondary structural content and consequently increased dynamics. These structures are expected to be further stabilised in the presence of the interacting antibody. On the other hand, helical populations seem to be preferred for this isolated peptide in solution, even though the length of the peptide is smaller compared to the experimental ones' for which helical conformations were found (see Table S1 for respective peptide lengths per structure). In general, we observe a high agreement on the relative percentage preference for each experimental structure among the two force fields. This means that the delicate balance between helical, beta, turn and coil structures is preserved, as expected for two so closely related force field versions. However, it cannot be overlooked that the ILDN* force field gives constantly higher representation of all the structures. This phenomenically better performance cannot be fully evaluated since its highly disordered nature prevents an accurate description of the peptide's folding landscape. Nonetheless, when it comes to describe the stable peptide conformations, the ILDN* force field presents itself as a better choice.

To elaborate further, the two earliest solution NMR structures of the 13-mer peptide, 1LCX[15] and 1MZI,[18] although they are the closest to our simulation system, they seem to present only a fraction of the peptide's available conformations, occupying 10−30% of the stable conformers (see also Figure S1). The 3_{10}-helix in particular, was proposed after using the ^{664}Asp−^{667}Ala hydrogen bond as a restrain, as mentioned in the previous sections. This hydrogen bond even though is observed, is not persistent (see also Figure S2). In the simulation with the ILDN force field, that hydrogen bond is present in 57% of the low-temperature associated frames (22% of the total simulation time) and in the simulation with the ILDN* force field, in 88% of the low-temperature associated frames (33% of the total simulation time). This implies that the presence of that hydrogen bond is not enough to restrain the conformation to a 3_{10}-helix, as the persistence of the hydrogen bond is 3−4 times higher than the occupancy of this set of structures. The imposition of this restrain could thus affect the accurate determination of a representative NMR ensemble.

Likewise, turn structures that are observed in the context of the short 7-mer, 9-mer and 11-mer in complex with the 2F5

antibody are restrained by 3 intrapeptide hydrogen bonds (Figure S2). These hydrogen bonds constrain the peptide to a straight conformation with side-chain arrangement that favors antibody recognition.[34,35] Apparently one of them, between ^{664}Asp−^{667}Ala, is the same one that contributes to the 3_{10}-helix formation discussed above. The second hydrogen bond between ^{666}Trp−^{669}Leu, is quite persistent as well, being present in 65% of the low-temperature associated frames (26% of the total simulation time) for the ILDN force field and in 86% of the low-temperature associated frames (32% of the total simulation time) for the ILDN* force field. This hydrogen bond, as discussed in the previous section, was often associated with the initiation of an α-helix extending through the C-terminus of the longer peptides. On the other hand, the third hydrogen bond between ^{664}Asp−^{666}Trp is quite scarce with only 2.2% presence in the low-temperature associated frames (0.9% of the total simulation time) for the ILDN force field and 1.8% presence in the low-temperature associated frames (0.7% of the total simulation time) for the ILDN* force field. This last hydrogen bonds seems to "lock" the ^{664}DKW^{666} β-turn conformation that is crucial for antibody recognition. Thereby, these short turn structures comprising the central peptide epitope residues are visited more often, with 35−45% occupancy, whilst peptide structures with mixed turn/coil conformations along the whole peptide length are still visited but with occupancies of around 5%.

Helical conformations persist longer, with occupancies of no less than 25% that can reach up to 66% of simulation time (accounting for the low-temperature associated frames). This seems contradictory to the fact that the peptide is found experimentally in helical conformations only in a larger context (21-mer to 59-mer) and either in complex with an antibody or in membrane-mimicking environment. It seems though that there is an intrinsic helical preference for this isolated 13-mer peptide in aqueous solutions as well. If this is the case, then such a computational approach to characterise the full ensemble of available structures for a highly disordered peptide as this, is extremely valuable as it presents all the peptide's accessible conformations with their relative stabilities and preferences.

Closing this section of successfully reproducing all available experimental structures, it is worth mentioning that no other significant population (with occupancy of more than 1% of the low-temperature associated frames) of stable structures was found during our simulations, that differed by more than 2 Å from at least one of the experimental structures included in the collection of structures in Figure 1.

2.3. Differential Secondary Structure Preferences between ILDN and ILDN* Force Fields. From a general perspective, the ILDN and ILDN* simulations show an overall agreement on the preferred peptide structures, in the sense that they give similar relative representation of the helical
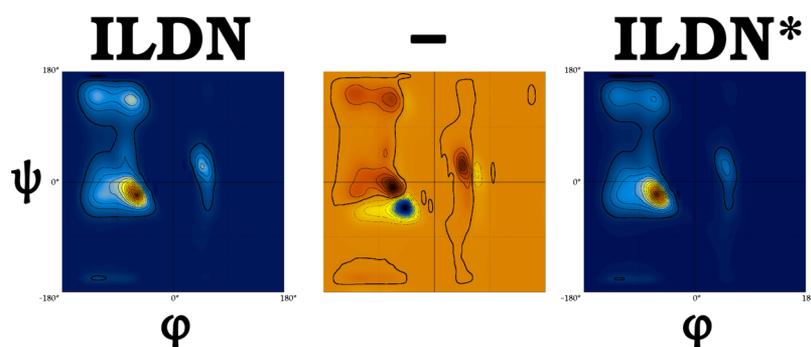
Figure 4. Ramachandran plots. Density distributions of the dihedral angles of the peptide calculated over the whole trajectory for the ILDN (left) and ILDN* (right) force field, with yellow-red colours representing high densities. The middle graph is the difference plot between the two force fields where blue colors correspond to higher preferences for ILDN* (right) and yellow-red colors correspond to higher preference for ILDN (left) force field.
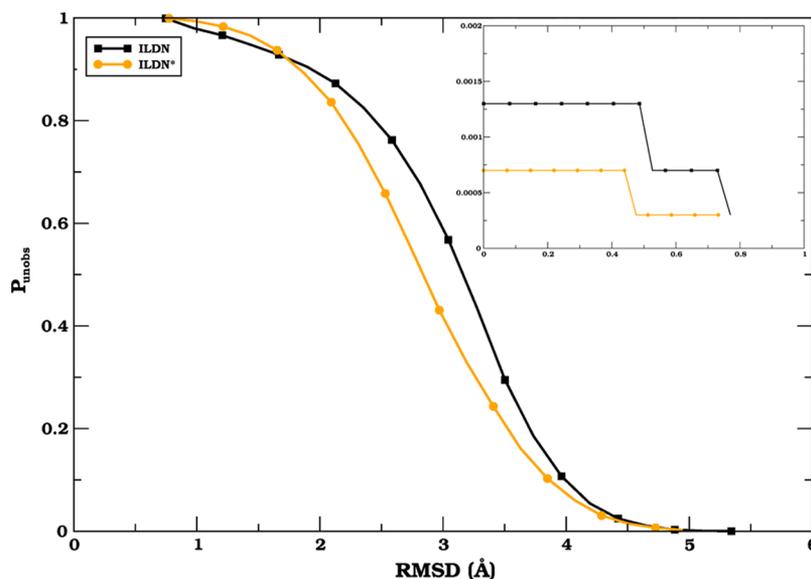

Figure 5. Extent of sampling through application of Good−Turing statistics. Probability of unobserved structures as a function of rmsd from the observed ones, for ILDN (black curve) and ILDN* (orange curve) force fields as produced by the GoodTuring MD program (https://github.com/pkoukos/GoodTuringMD). The main curves are the estimates obtained using structures of the whole trajectories whereas the curves in the inset are estimates for only structures that are associated with temperatures of $T < 340$ K (corresponding to structurally more stable peptide conformations).

structures, over the turn/coil ones, with minimal interstitial β-structures. This is nicely captured in Figure 3, where we calculate the per residue secondary structural preferences for the two force fields. A closer look reveals that the first and last couple of residues are quite flexible, as expected for such a short peptide. On the other hand, the central part of the peptide comprising the ^{651}LELDKWAS^{668} epitope sequence shows some discrepancy between the two force fields. The preference for turn conformation of each residue is preserved among them but there is difference in the preference for α-helix and 3_{10}-helix. The ILDN* force field is prone to helical conformations overall, as is also observed in the per residue secondary structure analysis over time (Figure S3, bottom panel). But this helical preference is closer to an α-helix and lesser to a 3_{10}-helix, the last one being more favoured by the ILDN force field. Likewise, most of the helical experimental structures (9 out of 30) are found in α-helical conformations, a couple of them present with a combination of turn and 3_{10}-helix and only one (out of 30) folds to a 3_{10}-helix, as discussed previously. From this perspective, ILDN* force field seems to

capture this behaviour more accurately. Nonetheless, the balance between these conformations is critical for accurate representation of the peptide's conformations which can play a role in antibody recognition and interaction.

To explore further the source of this discrepancy, we calculated the fraction of each secondary structure assignment for each residue as a function of temperature (Figure S4). This updated view reveals that α-helical conformations are favourably assigned to the central part of the peptide by both force fields, but they are more stable in the case of the ILDN* force field, as supported by the higher occupancy in the low temperature regime (see blue-coloured region in the difference-plot in the middle column). These same residues have a higher preference for 3_{10}-helix in the ILDN force field simulation, whilst ILDN* force field assigns a small 3_{10}-helical population only to residues in the right end part of the epitope sequence (notice the two small red and blue regions in the difference-plot in the middle column). On the other hand, β-structures are not observed in any of the experimentally determined peptide structures (Figure 1). The ILDN* force

field again seems to represent this feature better as they are only slightly visited at elevated temperatures (note also the almost complete absence of β-structures in the per residue secondary structure assignment over simulation time in Figure S3, bottom panel). On the contrary, ILDN force field allows the peptide to accommodate these structures at low temperatures, and particularly for residues in both the ends of the central epitope sequence, [660]LL[661] and [668]SL[669] (Figures 3, S3, and S4). On the other hand, turn and coil preferences among the two force fields seem very comparable with no outstanding differences, apart from some slightly higher preference for turn conformation on the N-terminal part and for coil conformation on the C-terminal part of the peptide by the ILDN* force field.

ILDN's bias over 3_{10}-helix is illustrated in the dihedral angle calculation presented in Figure 4 (see also the per residue Ramachandran plots and corresponding difference-Ramachandran plots in Figure S5). In these Ramachandran plots, the prevalence of helical populations is clear, but with subtle differences in the exact area that is occupied: ILDN's helical population tends more to 3_{10}-helix (note the slightly raised values in the helical region, coloured red in the middle difference Ramachandran plot), ILDN*'s helical population tends more to α-helix (note the characteristic blue-coloured region in the difference Ramachandran plot in the lower part of the helical region) and there is an oversampling of the β-region by ILDN force field. This behaviour is replicated if we repeat the calculation only for the stable structures of the trajectory associated with low temperatures (data not shown). Furthermore, it is transferred on a per residue level calculation (see Figure S5) and is particularly evident for the central epitope residues [665]KWAS[668], which are critical for antibody recognition. This implies that the calculated secondary structure preferences cannot be an artifact of the end-effect due to the peptide's small size or of the insufficient simulation length, as it is highly unlikely for both force fields to have not sampled a stable peptide structure (Figure 5), but rather represent an inherent better secondary structure balance of the ILDN* over the ILDN force field.

To conclude, it seems that the two force fields agree on the turn/coil representation even on a per-residue level, but the ILDN* force field is more α-helical whereas ILDN force field gives higher 3_{10}-helical content and emerges a minor yet unseen β-structure population. This trend has been observed recently in folding simulations using both helical and β-hairpin peptides.[32,33] Surprisingly, the force fields' agreement covers the end-residues of the peptide that are quite flexible, leaving the central supposedly structured part a point of controversy.

3. CONCLUSIONS

In this communication we tried to present the divergent structure determinations for this debatable small peptide and unravel its structural complexity by gathering the collection of representative available structures. From our perspective, the view of Anglister's group of a 3_{10}-helix (PDB entry 1LCX[15]) holds only partially true for this peptide, due to the imposition of a hydrogen bond, which seems that is a necessary but not a sufficient condition for the folding to a 3_{10} type of helix. On the other hand, Pessi's group argued that this 3_{10}-helix might be false and proposed instead that the peptide exists in solution as a complex mixture of conformers (PDB entry 1MZI[18]), with some local stable conformation for residues [660]Leu, [663]LD[664], [666]Trp and [670]Trp. In particular, they argued for a helical tendency for [659]ELLEL[663] residues followed by β-turn for

[664]DKW[666] and coil for [667]ASLWN[671] residues. According to our simulations, their ensemble of conformations (PDB entry 1MZI[18]) is not that often visited (10−20% of the low-temperature associated frames and 3−6% of the total simulation time, Figures 2 and S1). However, the local conformations that they proposed for certain residues are indeed supported by both our simulations and from other experimental structures (Figures 2 and 3). This highlights the difficulty in interpreting complex NMR data of disordered peptides to build accurate models that can cover the entire peptide's landscape.

In a recent MD study, the peptide was found to have significant amount of helical population, though less in explicit solvent compared to implicit solvent and visited the two experimental peptide structures that were examined (PDB entries 1LCX[15] and 1TJH[34]) only upon end-to-end distance restrains.[46] We believe that our unbiased molecular dynamics simulation with the properly selected force field successfully managed to capture the peptide's heterogeneity and offered the plethora of accessible peptide conformations with relative probabilities. Thus, our folding simulations offer an invaluable method to tackle such a disordered peptide that can accommodate different structures depending on the sequence context, the interaction partner and the surrounding environment.

What's more, the peptide's structural variance allowed us to compare the performance of two very closely related force field versions, ILDN and ILDN*, and look for signs of discrepancy. Both these refined force field versions reproduced all available experimental structures in sub-angstrom accuracy while preserving the disordered nature of the peptide. Nonetheless, the ILDN* force field offered a better representation of all the structures by accommodating them for a longer simulation time (although the actual simulation time was 3 times less compared to the ILDN force field). Last but not least, the backbone corrections in this modified ILDN* force field version provided a better balance between the different secondary structural elements with discrete differences in the helical region in the Ramachandran plot: ILDN force field shows an overestimation of 3_{10}-helical conformations wherein ILDN* force field favours α-helices instead.

Closing this section and based on this study, ILDN* force field emerges out of the AMBER99SB-ILDN family, as the choice of preference. The gp41[659−671] peptide through the availability of many divergent structure determinations served as an excellent candidate to examine the performance of molecular dynamics and current generation force fields not only to reproduce all of them but also to indicate structural preferences and nevertheless reveal sensitivities of the force fields. We believe that current generation force fields have matured enough in predicting peptide structure and dynamics that could complement current experimental schemes that suffer from the weakness of mixed folded/unfolded populations and the dependency on each experimental condition and aim to a better interpretation of experiments.

4. METHODS

4.1. Selection of Experimental Structures. Figure 1 illustrates the collection of experimental structures deposited in the PDB to date, that comprise the complete 13-mer peptide sequence [659]ELLELDKWASLWN[671] as well as smaller 7-mer, 9-mer and 11-mer peptide sequences that contain the complete central epitope sequence [662]ELDKWA[667]. PDB entries that

contain partially the sequence of gp41[659−671] (like PDB entry 2X7R[47]), were not included for structural comparison with the full 13-mer peptide sequence.

4.2. System Preparation and Simulation Protocol.

The starting peptide structure, [659]ELLELDKWASLWN[671], was in the fully extended form, as obtained from the program Ribosome with both terminal ends unprotected. Addition of missing hydrogen atoms and solvation−ionisation were performed with the program LEAP from the AMBER tools distribution.[48] The simulation was performed using periodic boundary conditions and a cubic unit cell sufficiently large to guarantee a minimum separation between the PBC-related images of the peptide of at least 16 Å, filled with pre-equilibrated TIP3P water molecules.[49] We followed the dynamics of the peptide's folding simulation using the program NAMD[50] for a grant total of 13 μs [10 μs with the AMBER99SB-ILDN,[51] hereafter referred to as "ILDN" and 3 μs with the AMBER99SB-STAR-ILDN[52] force field, hereafter referred to as "ILDN*"] and adaptive tempering with an inclusive temperature range of 280−480 K[53] as implemented in the program NAMD (adaptive tempering can be considered as a single-copy replica exchange method with a continuous temperature range).

The simulation protocol was the following: the system was first energy minimised for 1000 conjugate gradient steps. Then the temperature was increased with a ΔT step of 20 K until a final temperature of 300 K over a period of 32 ps. Subsequently the systems were equilibrated for 10 ps under NpT conditions without any restrains until the volume equilibrated. This was followed by the production NpT runs with the temperature and the pressure controlled using the Nosè−Hoover Langevin dynamics and Langevin piston barostat control methods as implemented by the NAMD program, with adaptive tempering applied through the Langevin thermostat, while the pressure was maintained at 1 atm. The Langevin damping coefficient was set to 1 ps^{-1} and the pistons oscillation period to 200 fs, with a decay time of 100 fs. The production run was performed with the impulse Verlet-I multiple timestep integration algorithm[54] as implemented by NAMD. The inner time step was 2 fs, with non-bonded interactions calculated every 1 time steps and full electrostatics every 2 time steps using the particle mesh Ewald method[55] with a grid spacing of approximately 1 Å and a tolerance of 10^{-6}. A cutoff for the van der Waals interactions was applied through a switching function, acting between 7 and 9 Å. The SHAKE algorithm[56] with a tolerance of 10^{-8} was used to restrain all bonds involving hydrogen atoms. Trajectories were obtained by saving the atomic coordinates of the whole system every 0.8 ps.

4.3. Extend of Sampling and Convergence.

Given the highly dynamic nature of the gp41[659−671] peptide in solution, even 13 μs of cumulative folding simulation time might be insufficient to characterise its folding landscape, especially the vast configurational space of the disordered/unfolded conformations. This can be nicely illustrated in Figure 5, where we demonstrate the application of a recently proposed probabilistic method that applies Good−Turing statistics to molecular dynamics trajectories.[57] According to that, there is a small probability (1 out of 10) to find structures that differ by more than 4 Å rmsd from the ones already observed, if you account for the whole trajectory and all-heavy atoms. This value drops down to 2.6−2.8 Å if you consider the central peptide epitope sequence [662]ELDKWA[667] or the backbone atoms of the 13mer peptide (data not shown).

However for the stable peptide conformations ($T < 340$ K), it seems improbable to find structures with statistically significant difference as there is a probability 1 out of 1000 to observe a structure with rmsd > 0.5 Å from the ones already observed (inset in Figure 5, 13mer, all-heavy atoms). This demonstrates the fine sampling of the stable peptide conformations, offering statistical significance to our observations regarding the experimentally observed structures (Figure 1 and Table S1). As for the unfolded fraction for which convergence is difficult to achieve, parameters that are affected by this shortcoming are explicitly discussed in the corresponding sections.

4.4. Trajectory Analysis.

The programs CARMA[58] and GRCARMA[59] together with custom scripts were used for most of the analyses, including removal of overall rotations/translations, calculation of the evolution of the Cα distances, calculation of rmsd's from a chosen reference structure, calculation of frame-to-frame rmsd matrices, principal component analysis in both Cartesian[60,61] and dihedral space[62,63] and corresponding cluster analysis, calculation of phi/psi dihedral angles, calculation of average structures and of the atomic root mean square fluctuations and production of PDB files from the trajectories. Secondary structure assignments were calculated with the program STRIDE.[64] All molecular graphics work and figure preparation was performed with the programs VMD[65] and CARMA.[58]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b01579.

> Summarised table of experimental structures found in the PDB; histogram rmsd distributions; hydrogen bond persistence over the trajectories; comparative molecular dynamics analysis over simulation time; per residue secondary structure preferences versus temperature; per residue Ramachandran and difference-Ramachandran plots (PDF)

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: panagiota.georgoulia@lnu.se. Phone: +46 (0)72 594 94 67.

ORCID Ⓓ
Panagiota S. Georgoulia: 0000-0003-4573-8052
Nicholas M. Glykos: 0000-0003-3782-206X

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Salzwedel, K.; West, J. T.; Hunter, E. A. 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.* **1999**, *73*, 2469−2480.

(2) Poumbourios, P.; el Ahmar, W.; McPhee, D. A.; Kemp, B. E. Determinants of human immunodeficiency virus type 1 envelope glycoprotein oligomeric structure. *J. Virol.* **1995**, *69*, 1209−1218.

(3) Wild, C.; Greenwell, T.; Matthews, T. A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virus-mediated cell-cell fusion. *AIDS Res. Hum. Retroviruses* **1993**, *9*, 1051−1053.

(4) Lawless, M. K.; Barney, S.; Guthrie, K. I.; Bucy, T. B.; Petteway, S. R., Jr.; Merutka, G. HIV-1 membrane fusion mechanism: structural

studies of the interactions between biologically-active peptides from gp41. *Biochemistry* **1996**, *35*, 13697−13708.

(5) Kilby, J. M.; Hopkins, S.; Venetta, T. M.; DiMassimo, B.; Cloud, G. A.; Lee, J. Y.; Alldredge, L.; Hunter, E.; Lambert, D.; Bolognesi, D.; Matthews, T.; Johnson, M. R.; Nowak, M. A.; Shaw, G. M.; Saag, M. S. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat. Med.* **1998**, *4*, 1302−1307.

(6) Veiga, A. S.; Santos, N. C.; Castanho, M. A. R. B. An Insight on the leading HIV entry inhibitors. *Recent Pat. Anti-Infect. Drug Discov.* **2006**, *1*, 67−73.

(7) Yu, F.; Lu, L.; Du, L.; Zhu, X.; Debnath, A.; Jiang, S. Approaches for identification of HIV-1 entry inhibitors targeting gp41 pocket. *Viruses* **2013**, *5*, 127−149.

(8) Conley, A. J.; Kessler, J. A.; Boots, L. J.; Tung, J. S.; Arnold, B. A.; Keller, P. M.; Shaw, A. R.; Emini, E. A. Neutralization of divergent human immunodeficiency virus type 1 variants and primary isolates by IAM-41-2F5, an anti-gp41 human monoclonal antibody. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3348−3352.

(9) Purtscher, M.; Trkola, A.; Gruber, G.; Buchacher, A.; Predl, R.; Steindl, F.; Tauer, C.; Berger, R.; Barrett, N.; Jungbauer, A.; Katinger, H. A broadly neutralizing human monoclonal antibody against gp41 of human immunodeficiency virus type 1. *AIDS Res. Hum. Retroviruses* **1994**, *10*, 1651−1658.

(10) Geffin, R. B.; Scott, G. B.; Melenwick, M.; Hutto, C.; Lai, S.; Boots, L. J.; McKenna, P. M.; Kessler, J. A., 2nd; Conley, A. J. Association of antibody reactivity to ELDKWA, a glycoprotein 41 neutralization epitope, with disease progression in children perinatally infected with HIV type 1. *AIDS Res. Hum. Retroviruses* **1998**, *14*, 579−590.

(11) Parker, C. E.; Deterding, L. J.; Hager-Braun, C.; Binley, J. M.; Schulke, N.; Katinger, H.; Moore, J. P.; Tomer, K. B. Fine Definition of the Epitope on the gp41 Glycoprotein of Human Immunodeficiency Virus Type 1 for the Neutralizing Monoclonal Antibody 2F5. *J. Virol.* **2001**, *75*, 10906−10911.

(12) Sattentau, Q. J.; Zolla-Pazner, S.; Poignard, P. Epitope exposure on functional, oligomeric HIV-1 gp41 molecules. *Virology* **1995**, *206*, 713−717.

(13) Gorny, M. K.; Zolla-Pazner, S. Recognition by human monoclonal antibodies of free and complexed peptides representing the prefusogenic and fusogenic forms of human immunodeficiency virus type 1 gp41. *J. Virol.* **2000**, *74*, 6186−6192.

(14) Koshiba, T.; Chan, D. C. The prefusogenic intermediate of HIV-1 gp41 contains exposed C-peptide regions. *J. Biol. Chem.* **2003**, *278*, 7573−7579.

(15) Biron, Z.; Khare, S.; Samson, A. O.; Hayek, Y.; Naider, F.; Anglister, J. A monomeric 3_{10}-Helix is formed in water by a 13-residue peptide representing the neutralizing determinant of HIV-1 on gp41. *Biochemistry* **2002**, *41*, 12687−12696.

(16) Mobley, P. W.; Pilpa, R.; Brown, C.; Waring, A. J.; Gordon, L. M. Membrane-perturbing domains of HIV type 1 glycoprotein 41. *AIDS Res. Hum. Retroviruses* **2001**, *17*, 311−327.

(17) Ahmed, Z.; Asher, S. A. UV resonance Raman investigation of a 3_{10}-helical peptide reveals a rough energy landscape. *Biochemistry* **2006**, *45*, 9068−9073.

(18) Barbato, G.; Bianchi, E.; Ingallinella, P.; Hurni, W. H.; Miller, M. D.; Ciliberto, G.; Cortese, R.; Bazzo, R.; Shiver, J. W.; Pessi, A. Structural analysis of the epitope of the anti-HIV antibody 2F5 sheds light into its mechanism of neutralization and HIV fusion. *J. Mol. Biol.* **2003**, *330*, 1101−1115.

(19) Tulip, P. R.; Gregor, C. R.; Troitzsch, R. Z.; Martyna, G. J.; Cerasoli, E.; Tranter, G.; Crain, J. Conformational plasticity in an HIV-1 Antibody Epitope. *J. Phys. Chem. B* **2010**, *114*, 7942−7950.

(20) Gregor, C. R.; Cerasoli, E.; Tulip, P. R.; Ryadnov, M. G.; Martyna, G. J.; Crain, J. Autonomous folding in the membrane proximal HIV peptide gp41(659−671): pH tuneability at micelle interfaces. *Phys. Chem. Chem. Phys.* **2011**, *13*, 127−135.

(21) Best, R. B.; Buchete, N.-V.; Hummer, G. Are current molecular dynamics force fields too helical? *Biophys. J.* **2008**, *95*, L07−L09.

(22) Freddolino, P. L.; Park, S.; Roux, B.; Schulten, K. Force field bias in protein folding simulations. *Biophys. J.* **2009**, *96*, 3772−3780.

(23) Matthes, D.; de Groot, B. L. Secondary structure propensities in peptide folding simulations: A systematic comparison of molecular mechanics interaction schemes. *Biophys. J.* **2009**, *97*, 599−608.

(24) Mittal, J.; Best, R. B. Tackling force field bias in protein folding simulations: folding of Villin HP35 and Pin WW domains in explicit water. *Biophys. J.* **2010**, *99*, L26−L28.

(25) Patapati, K. K.; Glykos, N. M. Order through disorder: Hyper-Mobile C-terminal residues stabilize the folded state of a helical peptide. A molecular dynamics study. *PLoS One* **2010**, *5*, No. e15290.

(26) Patapati, K. K.; Glykos, N. M. Three force fields' views of the 3_{10} helix. *Biophys. J.* **2011**, *101*, 1766−1771.

(27) Georgoulia, P. S.; Glykos, N. M. Using J-coupling constants for force field validation: application to hepta-alanine. *J. Phys. Chem. B* **2011**, *115*, 15221−15227.

(28) Patmanidis, I.; Glykos, N. M. As good as it gets? Folding molecular dynamics simulations of the LytA choline-binding peptide result to an exceptionally accurate model of the peptide structure. *J. Mol. Graphics Modell.* **2013**, *41*, 68−71.

(29) Georgoulia, P. S.; Glykos, N. M. On the foldability of tryptophan-containing tetra- and pentapeptides: an exhaustive molecular dynamics study. *J. Phys. Chem. B* **2013**, *117*, 5522−5532.

(30) Koukos, P. I.; Glykos, N. M. Folding molecular dynamics simulations accurately predict the effect of mutations on the stability and structure of a vammin-derived peptide. *J. Phys. Chem. B* **2014**, *118*, 10076−10084.

(31) Baltzis, A. S.; Glykos, N. M. Characterizing a partially ordered miniprotein through folding molecular dynamics simulations: comparison with the experimental data. *Protein Sci.* **2016**, *25*, 587−596.

(32) Serafeim, A.-P.; Salamanos, G.; Patapati, K. K.; Glykos, N. M. Sensitivity of folding molecular dynamics simulations to even minor force field changes. *J. Chem. Inf. Model.* **2016**, *56*, 2035−2041.

(33) Adamidou, T.; Arvaniti, K.-O.; Glykos, N. M. Folding simulations of a nuclear receptor box-containing peptide demonstrate the structural persistence of the LxxLL motif even in the absence of its cognate receptor. *J. Phys. Chem. B* **2018**, *122*, 106−116.

(34) Ofek, G.; Tang, M.; Sambor, A.; Katinger, H.; Mascola, J. R.; Wyatt, R.; Kwong, P. D. Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. *J. Virol.* **2004**, *78*, 10724−10737.

(35) Bryson, S.; Julien, J.-P.; Hynes, R. C.; Pai, E. F. Crystallographic definition of the epitope promiscuity of the broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2F5: vaccine design implications. *J. Virol.* **2009**, *83*, 11862−11875.

(36) Julien, J.-P.; Bryson, S.; Nieva, J. L.; Pai, E. F. Structural Details of HIV-1 Recognition by the Broadly Neutralizing Monoclonal Antibody 2F5: Epitope Conformation, Antigen-Recognition Loop Mobility, and Anion-Binding Site. *J. Mol. Biol.* **2008**, *384*, 377−392.

(37) de la Arada, I.; Julien, J.-P.; de la Torre, B. G.; Huarte, N.; Andreu, D.; Pai, E. F.; Arrondo, J. L. R.; Nieva, J. L. Structural constraints imposed by the conserved fusion peptide on the HIV-1 gp41 epitope recognized by the broadly neutralizing antibody 2F5. *J. Phys. Chem. B* **2009**, *113*, 13626−13637.

(38) Reardon, P. N.; Sage, H.; Dennison, S. M.; Martin, J. W.; Donald, B. R.; Alam, S. M.; Haynes, B. F.; Spicer, L. D. Structure of an HIV-1-neutralizing antibody target, the lipid-bound gp41 envelope membrane proximal region trimer. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 1391−1396.

(39) Serrano, S.; Araujo, A.; Apellániz, B.; Bryson, S.; Carravilla, P.; de la Arada, I.; Huarte, N.; Rujas, E.; Pai, E. F.; Arrondo, J. L. R.; Domene, C.; Jiménez, M. A.; Nieva, J. L. Structure and Immunogenicity of a Peptide Vaccine, Including the Complete HIV-1 gp41 2F5 Epitope. *J. Biol. Chem.* **2014**, *289*, 6565−6580.

(40) Sun, Z.-Y. J.; Oh, K. J.; Kim, M.; Yu, J.; Brusic, V.; Song, L.; Qiao, Z.; Wang, J.-h.; Wagner, G.; Reinherz, E. L. HIV-1 broadly neutralizing antibody extracts its epitope from a kinked gp41

ectodomain region on the viral membrane. *Immunity* **2008**, *28*, 52−63.

(41) Sun, Z.-Y. J.; Cheng, Y.; Kim, M.; Song, L.; Choi, J.; Kudahl, U. J.; Brusic, V.; Chowdhury, B.; Yu, L.; Seaman, M. S.; Bellot, G.; Shih, W. M.; Wagner, G.; Reinherz, E. L. Disruption of helix-capping residues 671 and 674 reveals a role in HIV-1 entry for a specialized hinge segment of the membrane proximal external region of gp41. *J. Mol. Biol.* **2014**, *426*, 1095−1108.

(42) Huang, J.; Ofek, G.; Laub, L.; Louder, M. K.; Doria-Rose, N. A.; Longo, N. S.; Imamichi, H.; Bailer, R. T.; Chakrabarti, B.; Sharma, S. K.; Alam, S. M.; Wang, T.; Yang, Y.; Zhang, B.; Migueles, S. A.; Wyatt, R.; Haynes, B. F.; Kwong, P. D.; Mascola, J. R.; Connors, M. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **2012**, *491*, 406−412.

(43) Ofek, G.; Zirkle, B.; Yang, Y.; Zhu, Z.; McKee, K.; Zhang, B.; Chuang, G.-Y.; Georgiev, I. S.; O'Dell, S.; Doria-Rose, N.; Mascola, J. R.; Dimitrov, D. S.; Kwong, P. D.; Doms, R. W. Structural basis for HIV-1 neutralization by 2F5-like antibodies m66 and m66.6. *J. Virol.* **2014**, *88*, 2426−2441.

(44) Biron, Z.; Khare, S.; Quadt, S. R.; Hayek, Y.; Naider, F.; Anglister, J. The 2F5 epitope is helical in the HIV-1 entry inhibitor T-20. *Biochemistry* **2005**, *44*, 13602−13611.

(45) Sia, S. K.; Carr, P. A.; Cochran, A. G.; Malashkevich, V. N.; Kim, P. S. Short constrained peptides that inhibit HIV-1 entry. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14664−14669.

(46) Zhang, Y.; Sagui, C. The gp41659-671 HIV-1 Antibody Epitope: A Structurally Challenging Small Peptide. *J. Phys. Chem. B* **2014**, *118*, 69−80.

(47) Buzon, V.; Natrajan, G.; Schibli, D.; Campelo, F.; Kozlov, M. M.; Weissenhorn, W. Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathog.* **2010**, *6*, No. e1000880.

(48) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The AMBER biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(50) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comp. Physiol.* **1999**, *151*, 283−312.

(51) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950−1958.

(52) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix−Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.

(53) Zhang, C.; Ma, J. Enhanced sampling and applications in protein folding in explicit solvent. *J. Chem. Phys.* **2010**, *132*, 244101.

(54) Izaguirre, J. A.; Reich, S.; Skeel, R. D. Longer time steps for molecular dynamics. *J. Chem. Phys.* **1999**, *110*, 9853−9864.

(55) Darden, T.; York, D.; Pedersen, L. An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(56) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(57) Koukos, P. I.; Glykos, N. M. On the application of Good-Turing statistics to quantify convergence of biomolecular simulations. *J. Chem. Inf. Model.* **2014**, *54*, 209−217.

(58) Glykos, N. M. Software news and updates carma: A molecular dynamics analysis program. *J. Comput. Chem.* **2006**, *27*, 1765−1768.

(59) Koukos, P. I.; Glykos, N. M. GRCARMA: a fully automated task-oriented interface for the analysis of molecular dynamics trajectories. *J. Comput. Chem.* **2013**, *34*, 2310−2312.

(60) Ichiye, T.; Karplus, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **1991**, *11*, 205−217.

(61) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412−425.

(62) Mu, Y.; Nguyen, P. H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins* **2005**, *58*, 45−52.

(63) Altis, A.; Nguyen, P. H.; Hegger, R.; Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126*, 244111.

(64) Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **1995**, *23*, 566−579.

(65) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.