Supporting Information

Folding molecular dynamics simulations of a gp41-derived peptide reconcile divergent structure determinations.

Panagiota S. Georgoulia[*], Nicholas M. Glykos

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

PDB code	Consensus Amino Acid Sequence	Comments
SIM	ELLELDKWASLWN	13-mer peptide
1LCX_A	ELLELDKWASLWN	Solution NMR, 13-mer peptide[15]
1MZI_A	ELLELDKWASLWN	Solution NMR[18]
1TJG_P	---ELDKWAS---	X-ray Diffraction, 7-mer in complex with 2F5[34]
1TJH_P	-LLELDKWASLW-	X-ray Diffraction, 9-mer in complex with 2F5[34]
1TJI_P	ELLELDKWASLW-	X-ray Diffraction, 11-mer, in complex with 2F5[34]
1U8I_C	---ELDKWA----	X-ray Diffraction, 7-mer in complex with 2F5[35]
1U8J_C	---ELDKWA----	X-ray Diffraction, 7-mer in complex with 2F5[35]
1U8K_C	--LELDKWASL--	X-ray Diffraction, 9-mer in complex with 2F5 (to Be Published, 2004)
2F5B_P	---ELDKWAS---	X-ray Diffraction, 7-mer in complex with 2F5[35]
2LP7_A	ELLELDKWASLWN	Solution NMR, 59-mer membrane associated trimer[38]
2M7W_A	ELLELDKWASLWN	Solution NMR, 59-mer, independently verified trimer[38]
2M8M_A	ELLELDKWASLWN	Solution NMR, 28-mer in hexafluoroisopropanol[39]
2M8O_A	ELLELDKWASLWN	Solution NMR, 28-mer in DPC[39]
2ME1_A	ELLELDKWASLW-	Solution NMR, 27-mer, alanine mutant in DPC[41]
2P8L_C	ELLELDKWASL--	X-ray Diffraction, 13-mer mutant, in complex with 2F5[36]
2P8M_C	ELLELDKWA----	X-ray Diffraction, 13-mer mutant, in complex with 2F5, different crystal form[36]
2P8P_C	--LELDKWASLW-	X-ray Diffraction, 11-mer mutant, capped, in complex with 2F5[36]
2PV6_A	---ELDKWASLWN	Solution NMR, 22-mer in DPC[40]
3D0L_C	--LELDKWASLW-	X-ray Diffraction, 35-mer construct, in complex with 2F5[36]
3D0V_C	-LLELDKWAS---	X-ray Diffraction, 11-mer peptide, in complex with 2F5[36]
3DRO_P	--LELDKWA----	X-ray Diffraction, 13-mer peptide, in complex with 2F5, in ammonium sulfate[36]
3DRQ_C	--LELDKWAS---	X-ray Diffraction, 35-mer construct, in complex with 2F5, in PEG/2-propanol[36]
3DRT_C	---ELDKWA----	X-ray Diffraction, 35-mer construct, in complex with 2F5[37]
3MNW_P	ELLELDKWASLW-	X-ray Diffraction, 21-mer peptide in complex with 13H11 (to Be Published, 2010)
3MNZ_P	ELLELDKWASLW-	X-ray Diffraction, 21-mer peptide in complex with 2F5 (to Be Published, 2010)
3MOA_P	ELLELDKWASLW-	X-ray Diffraction, 17-mer peptide in complex with 2F5 (to Be Published, 2010)
3MOB_P	-LLELDKWASLW-	X-ray Diffraction, 11-mer peptide in complex with 2F5 (to Be Published, 2010)
3MOD_P	-LLELDKWASLW-	X-ray Diffraction, 11-mer peptide in complex with 2F5 (to Be Published, 2010)
4G6F_F	ELLELDKWASLWN	X-ray Diffraction, 34-mer peptide in complex with 10E8[42]
4NRX_C	ELLELDKWAS---	X-ray Diffraction, 34-mer peptide in complex with m66[43]

Table S1: Summary of the experimental structures found in the PDB that entail the gp41[659-671] T-20 peptide, including the complete epitope (sequence ^{662}ELDKWA^{667}) for the broad-spectrum neutralizing 2F5 monoclonal antibody. For each one of the 30 structures, the PDB accession code is given (including the chain used for analysis), along with consensus amino acid sequence and the experimental method used for structure determination.
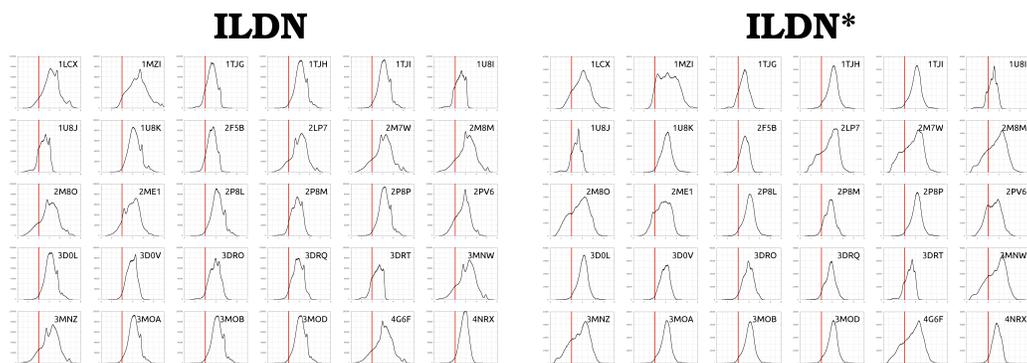
Figure S1: **Histogram RMSD distributions.** Histogram distributions of the RMSD from each experimental structure over the course of the simulation. All-backbone atoms were used for the calculation, excluding the fist/last two flexible residues. The red vertical line denotes the 2.0Å cut-off that was used for the calculation of the occupancies of each experimental structure over the stable conformations of the trajectory (T<340K).
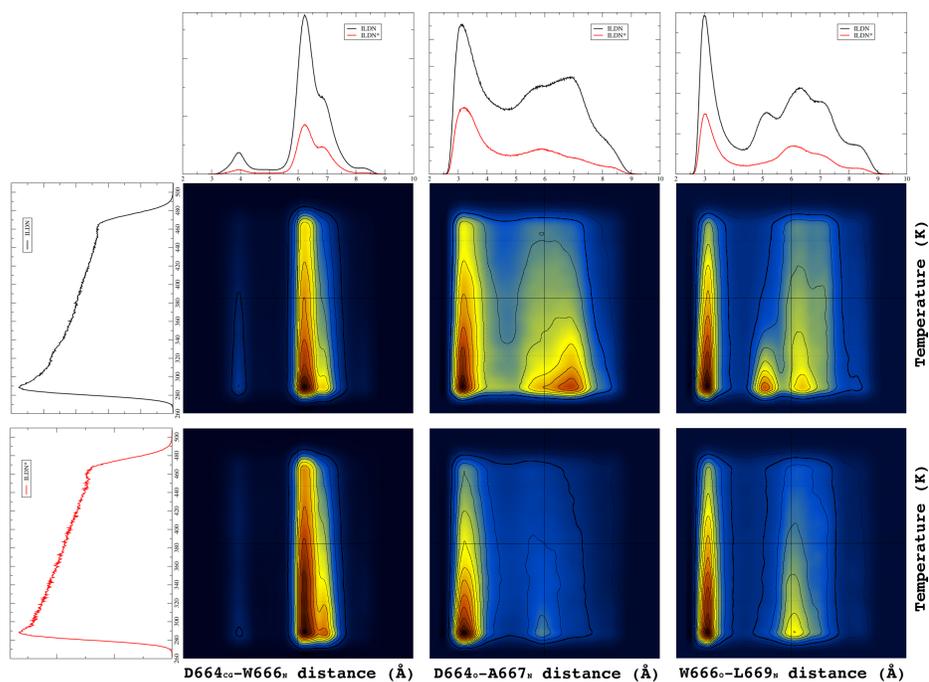
Figure S2: **Hydrogen bond persistence over the trajectories.** H_{bond}-T probability plots for 3 'key' hydrogen bonds, for ILDN (top row) and ILDN* (bottom row). Histograms of the two parameters are overlayed on the two axis of the plots.
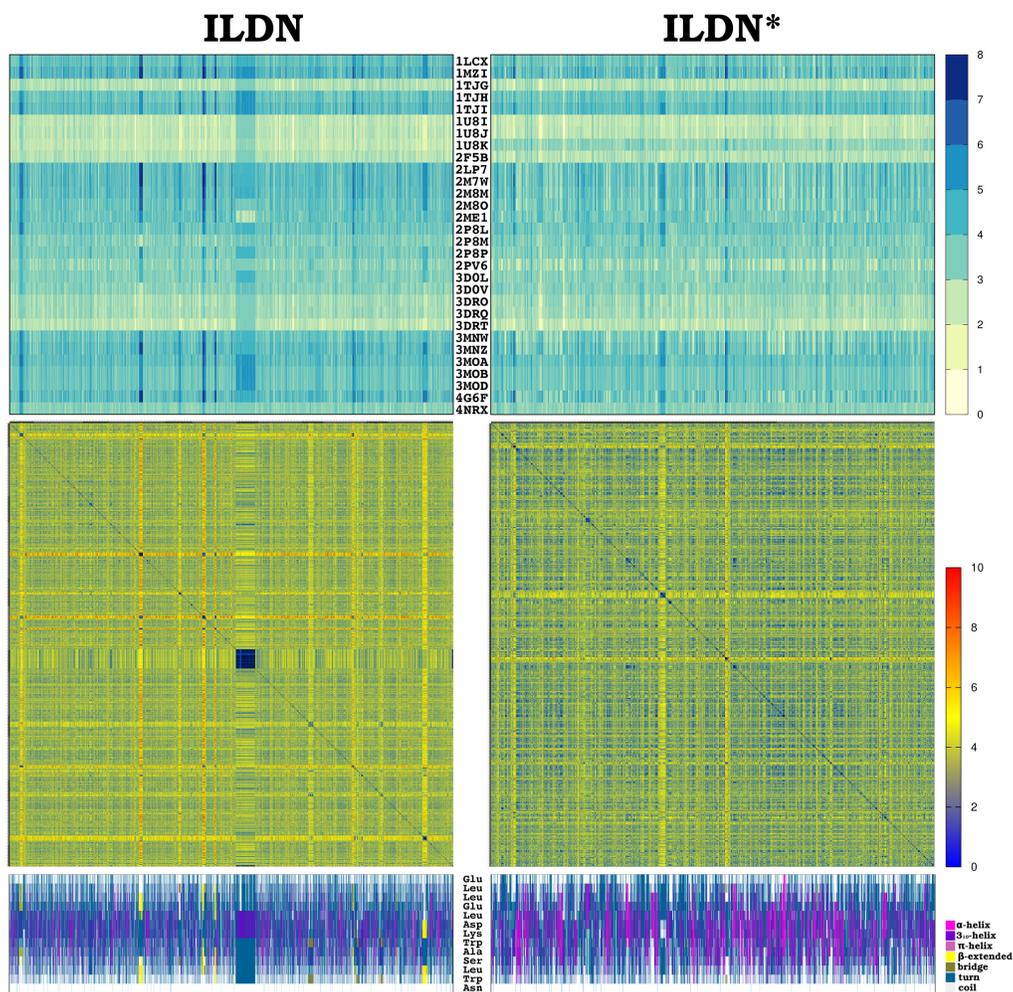
Figure S3: **Comparative molecular dynamics analysis for two force field variants ILDN (left column) and ILDN* (right column) over simulation time (x-axis).** From top to bottom (a) Evolution of RMSD (for all-heavy atoms) during the simulation from each experimental structure (individual rows in the matrix) using color coding from light green (0Å) to dark blue (8Å), (b) Matrices of RMSDs (for all-heavy atoms) between all possible structures from successive frames of the trajectory (the origin, t=0, is at the top left-hand corner) using linear color scale ranging from dark blue (0Å), through yellow to dark red (10Å), (c) Evolution of per residue secondary structure STRIDE assignments (one per row) using the color code shown in the right-hand panel (same as in Figure 1).
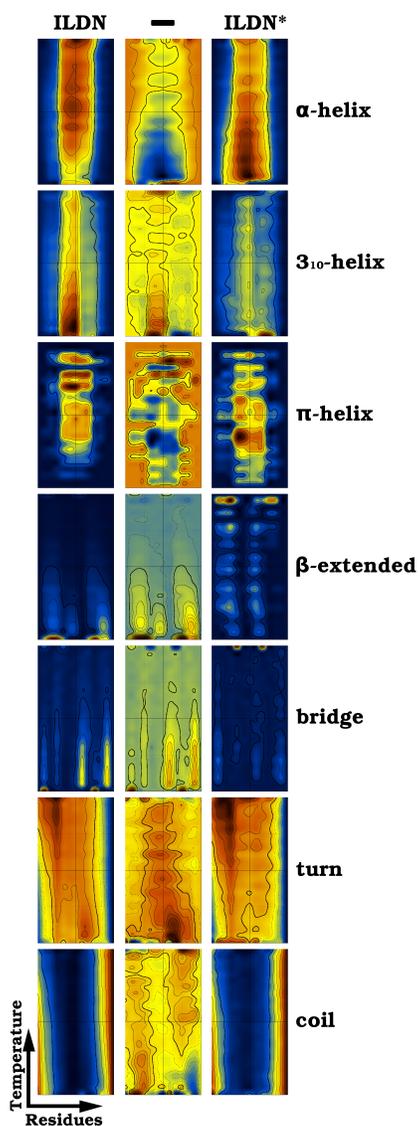
Figure S4: **Per residue secondary structure preferences vs Temperature.** Distribution of the fraction of each secondary structure type for each residue (x axis) as a function of temperature (y axis). The origin for each matrix is at the bottom left corner. The left column of matrices is calculated for the ILDN force field, the right one for ILDN* force field and the middle column of matrices is the difference between them, where cyan/blue coloring means is mostly occupied in the ILDN* force field (right) and yellow/red coloring means is mostly occupied in the ILDN (left) force field.

S6
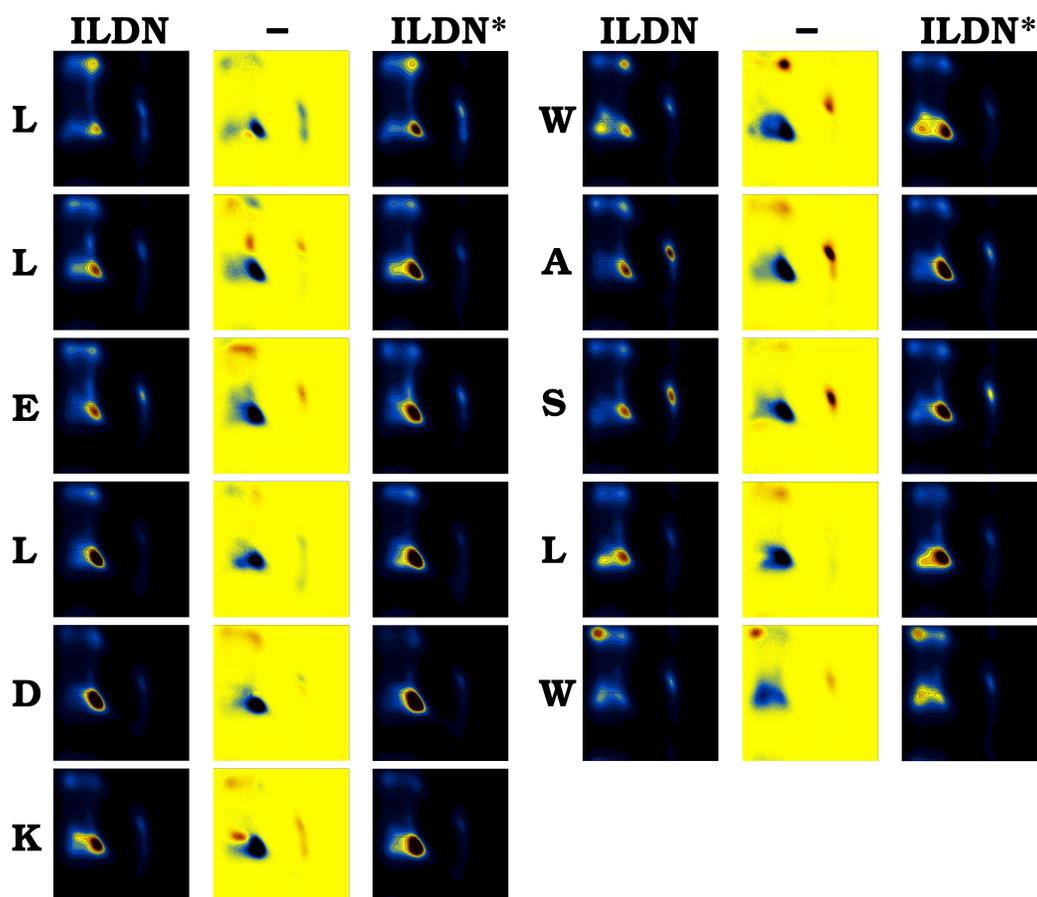
Figure S5: **Per-residue Ramachandran plots and difference-Ramachandran plots.** Ramachandran plots for residues 2-12 of the 13mer gp41[659-671] T-20 peptide calculated over the whole trajectories for ILDN (left column) and ILDN* (right column) force fields. The middle column is the difference between the two force fields where blue colors correspond to higher preference for ILDN* (right) and red colors correspond to higher preference for ILDN (left) force field.