




FULL PAPER

The balance between side-chain and backbone-driven association in folding of the α -helical influenza A transmembrane peptide

Ioannis Stylianakis¹ | Ariella Shalev² | Steve Scheiner³ | Michael P. Sigalas⁴  |
Isaiah T. Arkin² | Nikolas Glykos⁵  | Antonios Kolocouris¹ 

¹Section of Pharmaceutical Chemistry, Department of Pharmacy, National and Kapodistrian University of Athens, Athens, Greece

²Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus Givat-Ram, Jerusalem, Israel

³Department of Chemistry and Biochemistry, Utah State University, Logan, Utah

⁴Department of Chemistry, Laboratory of Applied Quantum Chemistry, Aristotle University of Thessaloniki, Thessaloniki, Greece

⁵Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece

Correspondence

Antonios Kolocouris, Department of Pharmaceutical Chemistry, Faculty of Pharmacy, National and Kapodistrian University of Athens, Athens, Greece.
Email: ankol@pharm.uoa.gr

Nikolas Glykos, Department of Molecular Biology and Genetics, Democritus University of Thrace, University Campus, Alexandroupolis 68100, Greece.
Email: glykos@mbg.duth.gr

Abstract

The correct balance between attractive, repulsive and peptide hydrogen bonding interactions must be attained for proteins to fold correctly. To investigate these important contributors, we sought a comparison of the folding between two 25-residue peptides, the influenza A M2 protein transmembrane domain (M2TM) and the 25-Ala (Ala₂₅). M2TM forms a stable α -helix as is shown by circular dichroism (CD) experiments. Molecular dynamics (MD) simulations with adaptive tempering show that M2TM monomer is more dynamic in nature and quickly interconverts between an ensemble of various α -helical structures, and less frequently turns and coils, compared to one α -helix for Ala₂₅. DFT calculations suggest that folding from the extended structure to the α -helical structure is favored for M2TM compared with Ala₂₅. This is due to CH \cdots O attractive interactions which favor folding to the M2TM α -helix, and cannot be described accurately with a force field. Using natural bond orbital (NBO) analysis and quantum theory atoms in molecules (QTAIM) calculations, 26 CH \cdots O interactions and 22 NH \cdots O hydrogen bonds are calculated for M2TM. The calculations show that CH \cdots O hydrogen bonds, although individually weaker, have a cumulative effect that cannot be ignored and may contribute as much as half of the total hydrogen bonding energy, when compared to NH \cdots O, to the stabilization of the α -helix in M2TM. Further, a strengthening of NH \cdots O hydrogen bonding interactions is calculated for M2TM compared to Ala₂₅. Additionally, these weak CH \cdots O interactions can dissociate and associate easily leading to the ensemble of folded structures for M2TM observed in folding MD simulations.

KEYWORDS

α -helix, adaptive tempering, B3LYP, CH \cdots O hydrogen bonding, circular dichroism, D95(d,p), folding, influenza A M2TM, molecular dynamics, NBO, ONIOM, QTAIM

1 | INTRODUCTION

Proteins with TM α -helices constitute 20–30% of all proteins encoded by sequenced genomes, and are critically involved in functionally rich processes, including bioenergetics, signal transduction, ion transmission, and catalysis^[1–4]. TM “single-pass” proteins, that span the

membrane bilayer with a single TM helix, are the largest class of integral membrane proteins^[4–6], and the TM spanning domains exhibit most of the functionalities of the full length proteins^[2, 4, 5]. Helix-helix packing from C–H \cdots O=C between the adjacent helices stabilize dimeric “single-pass” proteins, like glycoporphin A. Such forces are not present between the TM α -helices of the tetrameric or pentameric

proteins^[7], like the ion channels influenza A M2, influenza B BM2,^[8] hepatitis p7^[9] or the SARS E protein, respectively^[10]. These oligomeric bundles have enough conformational plasticity for ion channel activity and C—H...O=C can stabilize the secondary structure in the lipophilic peptide monomers.

M2 protein of the influenza A virus has 97 amino acid residues and forms homotetramers. It is a proton channel, with M2TM being the ion channel pore, and it forms an active, open state at low pH during endocytosis^[11–15]. The activation of M2TM tetrameric bundle ultimately leads to the unpacking of the influenza viral genome and to pathogenesis.^[16] In the influenza A M2TM tetramer, the hydrophobic α -helical monomer is the 25-residues M2TM monomeric peptide from Ser22 to Leu46 (SSDPLVVAASIIGILHLILWILDRL).

In this work, we explored the folding of M2TM. We performed circular dichroism (CD) experiments which are routinely carried out to probe the amount of secondary structure elements in equilibrium ensembles of proteins.^[17,18] We performed an extensive 2.2 μ s folding MD simulations, with adaptive tempering for sampling conformational space, of the 25-residues M2TM and Ala₂₅ peptides in trifluoroethanol (TFE), since TFE together with dimethyl sulfoxide are two established membrane-mimicking solvents.^[19] We aimed to show by means of these simulations that in the applied membrane-mimicking environment, M2TM forms an α -helix, even in its isolated monomeric form. Second, we investigated the stability of the hydrophobic α -helix of M2TM peptide and the propensity for folding from an extended, β -strand like structure, to the α -helix, compared to a reference Ala₂₅ peptide. Poly(alanine) peptides are model α -helical peptides, since alanine is the amino acid which most prefers α -helix. Subsequently, the energy stabilization from alanine residues to the poly(alanine) α -helix is higher compared to other polypeptides.^[20–25]

Toward the last aim, we applied DFT calculations to compare the difference in electronic energy between the fully minimized α -helical and extended structures for M2TM and Ala₂₅. The M2TM monomer has bulky adjacent aliphatic residue in the amino acid side chains, which are likely to have Pauli repulsive or London dispersion (LD) interactions, since LD always coexist with Pauli repulsive forces,^[26–29] with a resulting destabilization/stabilization effect in the α -helix. The LD can be considered to be C—H...H—C interactions between atoms of the side chains separated by 3–5 Å, which can be attractive due to the electron correlation at these distances.^[30–32] However, the C—H...O=C hydrogen bonding interactions are the most important attractive interactions.^[26–28] A hyperconjugative interaction $n(\text{C}=\text{O}) \rightarrow \sigma^*(\text{X}-\text{H})$ (X = N or C) that weakens the C—H bond is considered to be a useful measure for the identification of this hydrogen bonding interaction.^[26–29,32–39] The topological properties of the electron density at bond critical points (BCPs) from QTAIM have also been used for quantitatively characterizing hydrogen bonding interactions between two atoms.^[40,41] Using NBO^[34] and QTAIM^[40–42] calculations, we compared the C—H...O=C hydrogen bonding interactions with the backbone N—H...O=C (peptidic) hydrogen bonding interactions in the M2TM monomer. In Ala₂₅ peptide, the α -helix is stabilized mainly by N—H...O=C hydrogen bonding

interactions,^[25] and we compared also the N—H...O=C hydrogen bonding interactions between M2TM and Ala₂₅ peptides. Because Ala has a small side chain, poly(alanine) peptides are free of van der Waals interactions between residues, and helix formation is stabilized predominantly by the backbone peptide hydrogen bonding interactions.^[20–25]

We showed using DFT calculations that the M2TM is easier to fold from an extended structure to the α -helix compared with Ala₂₅. The NBO and QTAIM calculations suggested that the folding stabilization forces to the α -helical structures of M2TM are driven, not only by NH...O, but also by CH...O hydrogen bonding interactions. Our folding MD simulations showed that Ala₂₅ adopts a single α -helix while M2TM has many conformational variables in the side chains of M2TM, is more flexible and equilibrates between α -helical structures. In M2TM the weak C—H—O interactions can easily dissociate and reassociate,^[43] leading to an ensemble of α -helical and other folded structures, like turns and coils.

2 | METHODS

2.1 | System preparation and simulation protocol

We used as starting structures for the peptides M2TM and Ala₂₅ their fully extended conformations, with protected N- and C-termini by acetyl- and methylamino groups, respectively. The solvation of the peptides and the ionization states of the charges amino acids were carried out with program LEAP from the AMBER tools.^[44] For both simulations, periodic boundary conditions were applied with a cubic unit cell sufficiently large to guarantee the appropriate separation between the symmetry-related images of the peptides. This distance was set in 16 Å for the MD simulations as a minimum sufficient separation. The MD folding simulations of the peptides were performed using NAMD^[45] software for a grant total of 2.2 μ s using the TFE parameterization from the R.E.D. library^[46–48] and the AMBER99SB-STAR-ILDN force field.^[49–51] This force field has been shown to describe correctly the folding of numerous peptides,^[52–64] including peptides in mixed organic solvents (TFE/water).^[65]

For both MD simulations, adaptive tempering^[66] was applied as implemented in the program NAMD. Adaptive tempering is formally equivalent to a single-copy replica exchange MD simulation with a continuous temperature range. For the simulations, this temperature range was 280–380 K and was applied to the system through the Langevin thermostat.^[67] The simulation protocol has been previously described.^[54–56] In summary, the systems were first energy minimized for 1,000 conjugate gradient steps followed by a slow heating-up phase to a temperature of 380 K, with a temperature step of 20 K, over a period of 32 ps. Subsequently, the systems were equilibrated for 1,000 ps under NPT conditions without any restraints, until the volume was equilibrated. This was followed by the production NPT runs for 1.1 μ s with the temperature controlled using the Nosè-Hoover method^[68] and pressure with a Langevin piston barostat,^[69] which are the standard methods implemented by the NAMD program.

The adaptive tempering was applied through the Langevin thermostat and the pressure was maintained at 1 atm. The Langevin damping coefficient was set to 1 ps^{-1} , and the piston's oscillation period to 200 fs, with a decay time of 100 fs. The production runs were performed with the impulse Verlet-L^[70] multiple time-step integration algorithm as implemented by NAMD. The inner time step was 2.5 fs, with short-range nonbonded interactions being calculated in each time step, and long-range electrostatics interactions every two time steps using the particle mesh Ewald method^[71] with a grid spacing of approximately 1 \AA and a tolerance of 10^{-6} . A cutoff for the van der Waals interactions was set at 9 \AA through a switching function, and the SHAKE algorithm^[72] with a tolerance of 10^{-8} was used to restrain all X—H bonds. Trajectories were obtained by saving the atomic coordinates of the whole systems every 1.0 ps.

2.2 | Trajectory analysis

The analysis of the trajectories was carried out as previously described.^[54–56] The programs CARMA,^[73] GRCARMA^[74] and Cluster5D^[75] have been used for almost all of the analyses, including removal of overall rotations/translations, calculation of RMSDs from a chosen reference structure, calculation of average structures from clusters, production of PDB files from the trajectory, dihedral space principal component analysis (dPCA)^[76,77] and cluster analysis, calculation of the frame-to-frame RMSD matrices, calculation of similarity Q values, and so forth. Briefly, principal component analysis (PCA) is a method that takes the trajectory of a MD simulation and extracts the dominant modes in the motion of the molecule. Most of them contribute with small percentage to the overall motion and thus we choose for the description the major three components.^[76,77] Secondary structure assignments were calculated with the program STRIDE,^[78] while few other such programs have been developed like Mdtraj.^[79] All molecular graphics work and figure preparations were performed with the programs VMD,^[80] RASTER3D,^[81] PyMol,^[82] WebLogo,^[83] and CARMA.^[73]

2.3 | DFT calculations

2.3.1 | Geometry optimizations

Coordinates for the 17-residue Ala peptides were taken from Dannenberg's work.^[84] At the time this work started the structure of the 25-amino acids M2TM monomer, SSDPLVVAASIIGLHLILWILDRL, was taken from the X-ray structure of the M2TM tetramer-amantadine complex (PDB ID 3C9J)^[85] after removing additives from the X-ray structure, amantadine and the three other M2TM monomers. Structures for M2TM tetramer of higher resolution have been published^[86,87] but their Ca-carbons do not deviate significantly from the structure of M2TM monomer used here for the calculations.^[85] N- and C-termini of the M2TM peptide were capped by acetyl- and

methylamino groups, respectively. Calculations were performed with GAUSSIAN 09 package.^[88] All the structures in the α -helix or the β -strand conformation were fully energy optimized using the B3LYP functional.^[89] The basis sets tested were the 3-21G, 6-31G 6-31G (d,p), D95, and D95(d,p)^[90,91] (Tables 1 and S1). Geometry optimizations were performed with B3LYP/D95(d,p), B3LYP/6-31G, B3LYP/3-21G and the HF/D95(d,p), HF/6-31G, HF/3-21G. Geometry optimizations were performed also using the ONIOM method^[92,93] for the combinations B3LYP/ONIOM (6-31G, 3-21G), B3LYP ONIOM (6-31G, AM1) where in parenthesis are denoted the higher and lower level. The higher theory was applied to the backbone chain and the lower theory to the side chains. Frequency calculations were also performed at the B3LYP/D95(d,p) optimized structure using B3LYP/3-21G level of theory in order to characterize the stationary points found as minima.

2.3.2 | NBO analysis

The NBO 3.1 program^[34] was used as implemented in the GAUSSIAN 09 package. The NBO analysis of the molecular wave function provides a convenient way of exploring it, via chemically meaningful localized bond and lone pair orbitals. The NBO calculations were performed at B3LYP/D95(d,p) and M06/D95(d,p) theories. Second-order perturbation energies ($E(2)$) higher than $0.10 \text{ kcal mol}^{-1}$ were analyzed for the hyperconjugative interactions ($\text{C}=\text{O} \rightarrow \sigma^*(\text{N}-\text{H})$ and ($\text{C}=\text{O} \rightarrow \sigma^*(\text{C}-\text{H})$) (see Tables 2–4, Tables S2 and S4).

2.3.3 | QTAIM analysis

The existence^[40–42] of a (3,-1) BCP on the electron density at a $\text{D}-\text{H}\cdots\text{O}=\text{C}$ (D is N or C) bond path suggests the presence of a bonding between atoms H and O.^[94–98] The topological properties of the electron density at this BCP and especially the electron density and its Laplacian ($\nabla^2\rho_{\text{BCP}}$) are good descriptors of the strength of a particular bond. Typical values of hydrogen bonding interaction at $\text{H}\cdots\text{Y}$ critical points are 0.002–0.04 au for electron density and 0.02–0.15 au for its Laplacian.^[99,100] To gain a deeper insight into the nature of $\text{D}-\text{H}\cdots\text{O}=\text{C}$ contacts, QTAIM theory was applied

TABLE 1 DFT calculated electronic energies (kcal mol^{-1}) for the stabilization of energy minimized α -helix structure compared to the β -strand like structure (extended) using B3LYP functional and few levels of theory^a

Peptide	B3LYP D95(d,p)//B3LYP/6-31G, B3LYP 3-21G ^b	B3LYP/D95(d,p)
A ₂₅	—	34.23
M2TM	39.78	38.80

^aSee also Supporting Information.

^bONIOM theory applied with backbone calculated at the higher level and alkyl side chains at the lower level.

		R(H...O) (Å)	$\theta(\text{NH}\cdots\text{O}) (^{\circ})$	ρ_{BCP} (au)	$\nabla^2\rho_{\text{BCP}}$ (au)	E(2) (kcal mol ⁻¹)
O24 ^a	H88 ^a	2.136	164.1	0.0147	0.0667	7.35
	H72	2.140	137.8	0.0168	0.0718	2.39
O84	H136	1.952	162.4	0.0239	0.1041	13.87
O35	H102	1.935	168.3	0.0235	0.0103	13.13
O49	H112	1.956	163.4	0.0228	0.0976	11.28
O68	H123	1.904	169.1	0.0234	0.1029	9.44
O84	H136	1.932	162.4	0.0239	0.0104	9.77
O100	H155	1.905	166.0	0.0253	0.1098	10.57
O110	H170	1.876	164.5	0.0272	0.1164	11.79
O120	H180	1.889	163.8	0.0262	0.1140	10.89
O131	H199	1.866	168.7	0.0276	0.1191	12.01
O150	H220	1.843	166.9	0.0294	0.1257	12.44
O169	H235	1.880	163.1	0.0270	0.1158	10.64
O175	H254	1.911	162.9	0.0250	0.1086	10.28
O194	H273	1.905	167.2	0.0253	0.1091	10.79
O213	H298	1.889	167.7	0.0262	0.1136	10.82
O230	H316	1.846	169.7	0.0288	0.1253	11.53
O249	H335	2.070	142.0	0.0187	0.0811	2.80
	H353	2.022	156.8	0.0199	0.0845	7.73
O268	H368	2.132	139.3	0.0154	0.0727	3.38
O287	H368	2.966	126.1	0.0028	0.0116	—
O330	H401	1.930	175.3	0.0237	0.1029	8.87
Average		1.990	160.3	0.0226	0.0897	9.61
Sum				0.4975	1.9740	

^aIn the first column, it is shown the sp² donor orbital of the carbonyl oxygen and in the second column the hydrogen of the acceptor anti-bonding σ orbital of N—H bond.

TABLE 2 Geometrical and other characteristics of NH...O hydrogen bonds of M2TM, at B3LYP level with D95(d,p) basis set

to analyze the topology of the electron density. AIMALL^[101] software was used for identifying (3,-1) BCPs for the D—H...O=C segments of M2TM α -helix together with the corresponding bond paths. All the QTAIM analyses were carried out using the B3LYP/D95(d,p) wave functions.

2.4 | Experimental measurement of peptide thermal stability

In order to determine the thermal stability of the M2TM peptide, we used CD spectroscopy. In brief, the transmembrane peptide corresponding to residues Ser22-Leu46 of the M2 protein was synthesized using standard solid-phase F-moc chemistry and purified by reverse-phase HPLC according to established protocols.^[102,103] Following lyophilization of the peptide in fractions, the protein was reconstituted into 1% dodecyl maltoside (DDM) detergent-micelle using hexafluoro-isopropanol cosolubilization. Subsequently, CD spectra containing 1.7 mg ml⁻¹ (ca. 0.85 mM) of peptide were recorded using a J-810 spectropolarimeter (JASCO, Japan) in a 0.1 cm path length quartz cuvette. Peptide solutions contained 0, 1, and 4% sodium dodecyl sulfate (SDS).^[104]

3 | RESULTS AND DISCUSSION

3.1 | M2TM folds to a highly dynamic ensemble of α -helical structures

The starting structure of the 25-residues M2TM monomer and Ala₂₅ peptides for the folding MD simulations in TFE was a fully extended β -strand conformation. The simulations gave similar results starting from either the extended or the α -helical structure. In the former case, the peptide very quickly folded in the α -helical structure. The folding MD simulations were performed with adaptive tempering in the temperature range from 280 to 380 K (adaptive tempering is akin to a single copy replica exchange with a continuous temperature range). The implication is that this type of MD simulations never “converge” to a single structure (Figures S1 and S2). Instead, the MD simulation trajectory samples multiple folding and unfolding events, as the temperature continuously increases and decreases, and thus, the analysis of the conformational features is statistical. The statistical treatment included time-independent analyses and the results are shown in the WebLogo diagrams of Figure 1 and the average helical preferences are shown in Figure 2. Both of these analyses only deal with average secondary structure preferences and not atomic coordinates of snapshots.

TABLE 3 Geometrical and other characteristics of CH...O hydrogen bonds of M2TM, at B3LYP level with D95(d,p) basis set

		R(H...O) (Å)	$\theta(\text{CH}\cdots\text{O}) (^{\circ})$	ρ_{BCP} (au)	$\nabla^2\rho_{\text{BCP}}$ (au)	E(2) (kcal mol ⁻¹)
O24 ^a	H74 ^a	2.543	133.0	0.0080	0.0331	1.53
	H96	2.522	160.3	0.0079	0.0316	0.57
O84	H140	2.346	145.8	0.0113	0.0495	1.61
O49	H106	2.617	142.2	0.0071	0.0265	0.73
O68	H116	2.576	158.5	0.0076	0.0297	—
O84	H140	2.346	145.8	0.0113	0.0495	0.61
O100	H158	2.333	149.7	0.0115	0.0502	0.66
O110	H157	2.578	140.8	0.0080	0.0296	—
O120	H184	2.343	145.0	—	—	0.68
O131	H203	2.426	136.7	0.0096	0.0423	—
O150	H201	2.633	141.4	0.0071	0.0261	—
	H222	2.558	138.7	0.0074	0.0306	—
O169	H223	3.081	114.8	0.0032	0.0116	—
	H237	2.756	119.1	0.0053	0.0218	—
O175	H257	2.291	156.3	—	—	0.81
O194	H283	2.272	160.2	0.0129	0.0558	0.95
O213	H275	3.066	119.5	0.0076	0.0287	—
O213	H225	2.365	132.1	0.0119	0.0485	0.51
O230	H322	2.271	147.8	—	—	0.77
O268	H355	2.216	142.4	0.0166	0.0648	3.18
O287	H371	2.251	151.4	0.0144	0.0596	2.44
	H302	2.221	132.4	—	—	0.83
O311	H336	2.293	106.8	0.0171	0.0823	0.81
	H389	2.427	111.8	0.0110	0.0503	—
O360	H372	2.283	108.4	0.0166	0.0729	—
	H387	2.350	106.4	—	—	0.67
Average		2.460	136.4	0.0102	0.0426	1.09
Sum				0.2134	0.8950	

^aIn the first column, it is shown the sp² donor orbital of the carbonyl oxygen and in the second column the hydrogen of the acceptor anti-bonding σ orbital of N–H bond.

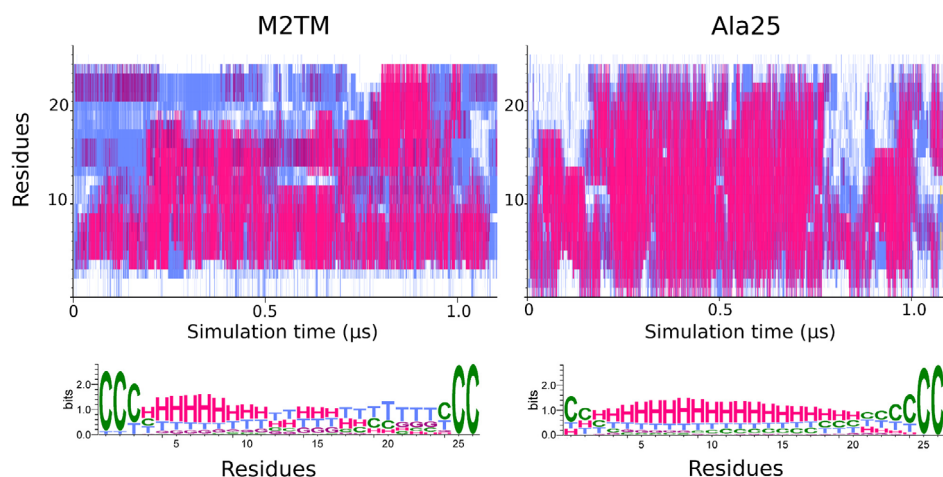


FIGURE 1 The two upper diagrams of the secondary structure analysis of the M2TM (left) and Ala₂₅ (right) peptides show the variation of the secondary structure per-residue. The color coding is red/magenta → $\alpha/3_{10}$ helical structure, cyan → turns, white → coil. The lower panels show the WebLogo-derived distributions corresponding to the STRIDE-derived secondary structure assignments as a function of simulation time with H → α helix, G → 3_{10} helix, T → turn, C → coil. The color coding is identical with the above except for coil which is depicted in green. For both peptides, the terminal protection groups were treated as discrete residues, with corresponding numbers of 0 and 26, respectively. The numbering refers to 1–25 amino acids of the 25-residue M2TM peptide which are residues 22–46 in the 98-residues full M2 protein sequence [Color figure can be viewed at wileyonlinelibrary.com]

The upper left diagram of Figure 1 shows the results obtained as per-residue secondary structure assignments of the M2TM peptide as a function of simulation time. Two observations are immediately

TABLE 4 Geometrical and other characteristics of $\text{NH}\cdots\text{O}$ hydrogen bonds of Ala_{25} , at B3LYP level with D95(d,p) basis set

		R(H \cdots O) (Å)	$\theta(\text{NH}\cdots\text{O}) (^{\circ})$	E(2) (kcal mol^{-1})
O53 ^a	H143 ^a	2.225	160.3	1.77
O55	H153	2.360	165.7	2.22
O57	H158	2.100	165.0067	3.94
O59	H163	2.291	165.6451	1.88
O61	H168	2.319	165.2411	1.77
O63	H173	2.206	163.2401	2.67
O65	H178	2.154	164.7158	3.33
O67	H183	2.184	166.0827	3.00
O69	H188	2.156	163.4445	3.20
O71	H193	2.149	164.6363	3.36
O73	H198	2.153	161.4066	3.20
O75	H203	2.263	163.2273	2.19
O77	H208	2.106	168.6601	4.15
O79	H257	2.305	151.0426	1.25
	H213	2.357	166.9166	2.37
O81	H218	2.142	160.4852	1.61
O112	H228	2.234	164.8095	3.67
O114	H233	2.217	163.6693	2.61
O116	H238	2.376	156.5034	2.60
O118	H243	2.441	155.1498	1.24
O120	H248	2.268	163.2853	1.01
O122	H258	2.356	140.5165	2.21
Average				2.51

^aIn the first column, it is shown the sp^2 donor orbital of the carbonyl oxygen and in the second column the hydrogen of the acceptor anti-bonding σ orbital of N–H bond.

obvious from Figure 1. The first is that the M2TM peptide is highly helical in TFE since most of the recorded structures have an α - or 3_{10} -helical secondary structure. The second observation is that the M2TM peptide appears to be dynamic in nature and quickly interconverts between various forms of helical structures, compared to Ala_{25} which includes mainly one α -helix, as recorded from the PCA analysis (Figures S3 and S4).

The variability in the M2TM helix propensity along the length of the peptide is shown in the WebLogo diagram, lower panel of Figure 1. The helix-forming propensity varies systematically for different parts of the peptide with residues Pro-4 to Ala-8 (Pro-25 to Ala-29 as numbered in the 98-residues full M2 protein) being always helical, whereas the N- and C-terminal residues remain flexible for the greatest part of the trajectory. In the edges of both peptides at the end of α -helices, the helicity is lowered. The capping residues have non-helical φ , ψ dihedral angles, although they form helical $i, i + 4$ hydrogen bonds.^[105] There is a discontinuity of α -helix centered at residue Gly13, since Gly is known to break strongly α -helix. In M2TM, the C-end part after Gly13, ILHLILWILDRL, is less helical compared to the N-end part SSDPLVVAASII before Gly13. The C-end includes the two polar residues His, Trp at Positions 16, 20 and the two charged residues Asp, Arg at Positions 23, 24 (37, 41 and 44, 45, respectively, as numbered in the full M2 protein) which reduce the α -helix character and the $\text{CH}\cdots\text{O}$ hydrogen bonding interactions. Further, Figure 1 shows not only the significant variation of helicity on a per-residue basis, but also indicates the presence of some minor preferences, such as for example, a tendency of Residues 14–16 and 21–23 (Residues 35–37 and 42–44 as numbered in the full M2 protein) at the C-end to form 3_{10} -helical structures.

Taken together, these two diagrams paint the following picture. As mentioned before, the M2TM peptide in TFE does not form one single α -helix. Instead, it interconverts between an ensemble of partially folded conformations, mainly of α -helical character, with the notable exception of its N-terminal half which is almost exclusively α -helical. In order to clarify the dynamic nature of M2TM monomer,

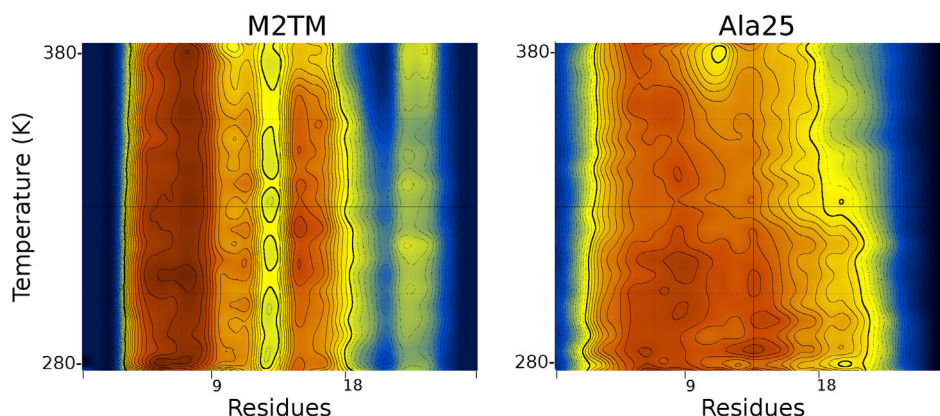


FIGURE 2 The two graphs describe the thermal stability of helical structures. Using a color representation, the graphs depict the values of the fractional helicity as a function of temperature for each residue of the two peptides. The color scale is the same for the two graphs and ranges from dark blue (fractional helicity of zero) to dark red (fractional helicity equal to 1.0). To help interpretation, contour lines have also been drawn (continuous lines for density values above the mean of the distribution, dotted lines for density values below the mean, thick solid line at the mean density). The maximum observed helicity of the M2TM peptide is 0.833 at low temperatures, and 0.762 at the upper temperature range [Color figure can be viewed at wileyonlinelibrary.com]

as equilibrium between different conformations, we analyzed the MD simulation trajectories using dPCA. The three main components of the trajectory describe 67% of conformational variance in M2TM and 87% in Ala₂₅. From the cluster analysis based on the RMSD of all carbons, we selected six conformations for M2TM and four conformations for Ala₂₅ (Figures S3 and S4). The higher conformational variance of the C-end part of M2TM monomer may contribute to the experimentally observed conformational variance of the M2TM tetramer's C-end inside the membrane bilayer environment. M2TM tetramer adopts closed and open conformations, with His16 (H37 as numbered in the 98-residues full M2 protein) tetrad neutral or fully protonated, both present in different percentage at high or low pH, respectively. The closed conformation has a higher percentage at high pH and the open conformation at low pH. In the closed or open states of M2TM tetramer, each helical M2TM has a kink at Gly13 or not, respectively.^[106–108] The M2TM tetramer plasticity is critical for its functional proton channel properties during the virus replication cycle.^[16]

3.2 | Comparison with the poly(alanine) peptide demonstrates the intricacies of complex sequence folding and establishes the stability of the M2TM peptide

The right panel of Figure 2 shows the results obtained from the Ala₂₅ peptide. The simplicity of the Ala₂₅ sequence is clearly reflected in its folding behavior. The poly(alanine) peptide spends most of the simulation time uniformly with folding and unfolding between full-length helical structures (Figures S2 and S4) which appear to be rather stable. The structural complexity observed in the M2TM peptide is no longer observed in Ala₂₅, as indicated by the uniform α -helical preferences shown in the WebLogo diagram, with the exception of the frayed termini. Noteworthy, the residues with the highest helical propensity are present in the M2TM peptide and not Ala₂₅, that is, 1.30 versus 1.25 bits for Residues 7 and 8 of M2TM (Residues 28 and 29 in the full M2 protein) and Ala₂₅, respectively. This finding prompted us to compare the thermal stability of the helical structures of the two peptides. Such an analysis was possible because the simulations were performed using adaptive tempering and a temperature range of 280–380 K. This implies that each of the structures recorded in the two trajectories has an associated temperature within this range. This allowed us to calculate, for every residue, the fraction of helicity as a function of the temperature. Figure 2 shows the results from this calculation, as graphical representation of how high or low, on a per residue basis, is the helical content as function of temperature. Warm colors in Figure 2 indicate high helical content, cold colors indicate low helical content, and both graphs for the two peptides are on the same color scale. What is immediately obvious from these graphs is the surprising finding that the residues with the highest thermal stability belong to the N-terminal part of the M2TM peptide, and not to Ala₂₅, in contrast to what might be expected from the overall α -helical character shown in Figure 2. Figure 2 shows again the

complex and nonuniform behavior of the M2TM peptide as analyzed in the previous section.

3.3 | Experimental evidence of α -helix folding

In order to determine whether M2TM goes through thermal denaturation, we used CD spectroscopy. We reconstituted the M2TM peptide in detergent-micelles (1% DDM) using organic solvent cosolubilization. Since M2 is helical in its native state,^[85–87,109–112] it exhibits a maximum CD signature at 222 nm.^[110–112] Therefore, we could conduct a thermal scan by monitoring the ellipticity at 222 nm as a function of temperature. Results shown in Figure 3 indicate no significant change, allowing us to conclude that thermal denaturation did not occur. Subsequently, we added SDS, a denaturing detergent to the solution and repeated the thermal scans. Once again, no difference in ellipticity was observed, indicating that under these conditions, M2TM secondary structure is stable, in agreement with previous experiments in DPC micelles.^[110]

It has been shown both experimentally and using MD simulations that in Ala₁₄–Ala₄₈ peptides at low concentrations (0.1–0.5 mM) and low temperatures (25°C) random coils assemble into α -helical structures and relatively very few β -sheet peptides.^[25,36,37] The β -sheet content increases to 20% at high temperatures (65°C). These concentrations correspond to the very dilute regime, in which most peptides do not interact with neighboring peptides. These, uninterrupted

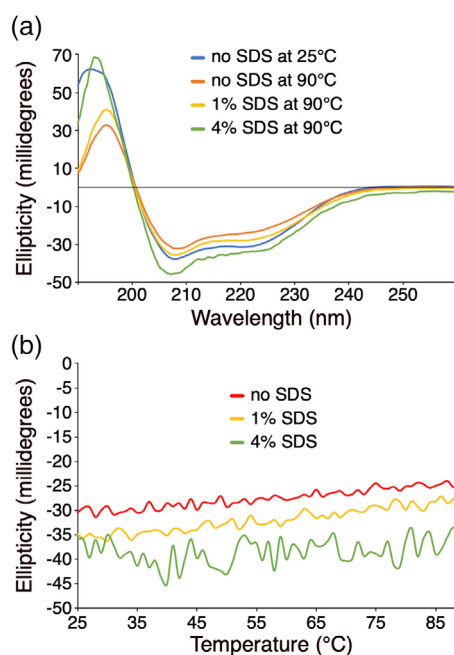


FIGURE 3 (a) Circular dichroism spectra of the M2TM peptide in three different detergent micelles containing *n*-dodecyl β -D-maltoside (DDM) and sodium dodecyl sulfate (SDS) in three micelles concentrations. Spectra were obtained at 25 or 90°C. (b) Ellipticity thermal profile for the M2TM peptide in three detergent micelles concentrations [Color figure can be viewed at wileyonlinelibrary.com]

alanine sequences are found to have the highest helix propensity characteristic of alanine between other peptides.^[24,25,38] Alanine is the only amino acid forming a standard peptide backbone that does not suffer a loss of side-chain conformational entropy on α -helix formation.^[24,25,38] Indeed, polyproline or polyglycine helices are frequently found in proteins, but are not considered as equivalent secondary structure elements, because they do not form a similar self-contained hydrogen-bonding network of the main chain atoms. As the peptide concentration increased to intermediate concentrations (1 mM) random coils assemble into α -helical structures at low temperatures and β -sheet peptides predominate at high temperatures 65°C). At high concentrations (5 mM) the β -sheet formation predominates at 25°C. Synthetic Ala₁₄-based peptides have been shown to undergo a transition from α -helical structures to β -sheet complexes in vitro,^[37] mimicking the structural transition that is believed to be a prerequisite for fibril nucleation and growth.^[18,39] Thus, poly(alanine)-based peptides have been used as model systems to study: (a) peptide hydrogen bonds and (b) interconformational processes as they may relate to the formation of anomalous filamentous intranuclear inclusions in human pathologies, like neurodegenerative diseases or notably in oculopharyngeal muscular dystrophy patients.^[113]

From the results obtained here and the previously published results,^[36,37] it is shown that both 25-residue peptides, M2TM or Ala₂₅, exist in a stable α -helix at 25°C and low concentration regime. The MD simulations suggest that both α -helices are spontaneously formed from an extended structure to α -helical peptides.

3.4 | Calculations of electronic energy for M2TM peptide folding

Understandably, although the MD simulations show how the highly complex membrane environment may affect folding, the comparison of the two peptides from MD simulations using a force field can be viewed as inherently limited. DFT calculations were applied to study how much spontaneous is the folding from the extended structure to the hydrophobic α -helix of the 25-residue M2TM monomer compared to Ala₂₅. Thus, we calculated the difference in electronic energy between the geometry optimized structures of the extended M2TM peptide and the M2TM α -helical monomer, and the difference in electronic energy between the geometry optimized structures of Ala₂₅ in the β -strand and the α -helix conformations. As a reference unfolded geometry, all dihedral angles along the chain were initially set to 180° as in the classical β -strand, and a full optimization begun at this point. Full geometry optimization of the 25-residue M2TM monomer from the experimental structure of the M2TM tetramer,^[85] led to the α -helix designation below (Figure 4). The minimized extended structure for M2TM is not an ideal β -strand conformation (Figure 5) and will be designated as extended M2TM structure below. From a general point of view, repulsions due to steric crowding between the bulky adjacent aliphatic residues of amino acid side chains would tend to destabilize the extended structure and the α -helix in M2TM. The Ala₂₅ peptide with a minimal side

chain groups is free of this kind of destabilizing steric repulsion for the α -helix and β -strand structures. These calculations do not account for the difference in energy between the α -helix and the 25 isolated residues of each peptide, since Ala is the amino acid which most preferred α -helix.

We also sought to explore a computationally economic level of theory for peptides using here the energetic comparison between α -helix and extended structure for peptides using here M2TM. We tested several levels of theory, including AM1, HF, DFT, and two-layer ONIOM methods with different basis sets, and we used also model poly(alanine) peptides, i.e., analogues of Ala₁₇ and Ala₂₅ (see Table S1 and Supporting Information). The combination of the B3LYP functional with the Dunning/Huzinaga full double- ζ D95(d,p)^[114] basis set was taken as reference level for full geometry optimization energies as it has been considered to be reliable in previous works.^[84] We showed that for M2TM, single point energies B3LYP D95(d,p) with fully geometry optimized structures with ONIOM B3LYP/6-31G, B3LYP 3-21G (the backbone was calculated at the higher level and alkyl side chains at the lower level) resulted in the closest agreement with B3LYP/D95(d,p) energy differences between extended structure and α -helix (see Table 1 and the relevant underlined values in Table S1).

The difference in electronic energy between the energy minimized α -helical monomer and the energy minimized extended structure was higher for M2TM compared to Ala₂₅ at all levels of theory (Tables 1 and S1). Ala₂₅ is a model α -helical peptide with an α -helix stabilized only by NH...O hydrogen bonding interactions. The MD simulations show that Ala₂₅ adopts a single α -helix in contrast to M2TM which has many different side chains and conformational variables. From the folding MD simulations, it was plotted a higher thermal stability for the N-terminal of M2TM (24-31) compared to the C-terminal part of the M2TM peptide. It has been shown that the energetics of α -helix formation from the unfolded nonhelical state are enthalpy dominated rather than entropy controlled.^[23] Consequently, the search centers on why folding M2TM α -helix from the extended structure is favored in M2TM compared to Ala₂₅, upon the energetics. Since the α -helix is held together in part by hydrogen bonds, it is here that the examination centers.

3.5 | Contribution of C—H...O hydrogen bonds to α -helical stability

The folding MD simulations from 280 to 380 K sample a wide region of conformational space, rather than structures around a minimum of an energy well, and the simulation never converges to a single average structure. An average structure for M2TM peptide based on the huge number of the structures from these folding MD simulations will not provide a useful mean to investigate the nature of the forces stabilizing M2TM α -helix. Thus, we choose a structure of M2TM peptide from the X-ray structure of the M2TM tetramer as a representative of the α -helix structure to investigate the presence of CH...O interactions.

In order to better understand the forces holding M2TM in the optimal α -helical conformation, an analysis has been carried out of all hydrogen bonds involving the carbonyl oxygens atoms. These interactions are of two sorts, those with the conventional NH proton donors, as well as those in which it is a CH group that serves as donor. Examples of some of these $\text{NH}\cdots\text{O}$ and $\text{CH}\cdots\text{O}$ hydrogen bonds for the M2TM peptide are displayed in Figure 4b, which constitutes an exploded section of the α -helix shown in Figure 4a, and which contains some of the relevant hydrogen bond lengths. Figure 5 displays the extended structure of which precludes the appearance of hydrogen bonds of either sort.

These N—H—O and C—H—O bonds are summarized in Tables 2 and S2, and Tables 3 and S3, respectively, for M2TM, and the $\text{NH}\cdots\text{O}$ hydrogen bonds in Tables 3 and S3, for Ala_{25} . The third and fourth

columns of Tables 2 and 3 report the geometrical aspects of each bond, both the hydrogen bond length and its deviation from linearity. The next two columns contain the salient characteristics of the QTAIM BCPs connecting the two pertinent atoms, both the density at that point, and its Laplacian. The final column is derived from the NBO treatment, wherein is the second-order perturbation energy $E^{(2)}$. This quantity measures the charge transfer involving the lone pair of the acceptor oxygen and the $\sigma^*(\text{NH/CH})$ anti-bonding orbital and may be overestimated,^[115] but nevertheless assures the presence of hydrogen bonding. All of these parameters are known to be strong indicators of the strength of a hydrogen bond (see also Section 1 and sub-section 2.3.2).

Comparison of Tables 2 and 3 shows quickly that the $\text{NH}\cdots\text{O}$ hydrogen bonds tend to be significantly shorter than their $\text{CH}\cdots\text{O}$ analogues. The former are generally less than 2 Å, while the latter are roughly 0.5 Å longer. The penultimate row of each Table provides an average, which is 1.990 and 2.460 Å for these two hydrogen bonds, respectively. The $\text{NH}\cdots\text{O}$ hydrogen bonds also tend closer to linearity. The average $\theta(\text{NH}\cdots\text{O})$ of 160.3° is within 20° of linearity, whereas $\text{CH}\cdots\text{O}$ hydrogen bonds deviate by an average of 44°. The wave function measures of hydrogen bond strength also show the stronger nature of $\text{NH}\cdots\text{O}$. The average ρ and $\nabla^2\rho$ are 0.0226 and 0.0897 for $\text{NH}\cdots\text{O}$, about double the same $\text{CH}\cdots\text{O}$ quantities. There is an even greater disparity of nearly an order of magnitude in the NBO average charge transfers.

Given the relative values described above, it is clear that conventional $\text{NH}\cdots\text{O}$ hydrogen bonds are a more important component of the optimal structure than are $\text{CH}\cdots\text{O}$. However, the data nonetheless raises the argument that the latter cannot be ignored in the M2TM 25-residue peptide helical structure which includes 16 lipophilic amino acids. There are fully 26 such $\text{CH}\cdots\text{O}$ interactions, between side chains of lipophilic amino acids and carbonyl oxygens present in this system, as compared to 22 $\text{NH}\cdots\text{O}$ bonds. The cumulative effects may be gauged by adding the sum of their electronic components. As reported in the last row of Tables 2 and 3, the sum of all 22 values of ρ_{BCP} for the $\text{NH}\cdots\text{O}$ H-bonds is 0.498 au, which compares to 0.213 au for $\text{CH}\cdots\text{O}$. This ratio of 2:1 is also true for the Laplacian sum, 1.974 au vs 0.895 au. Within the context of

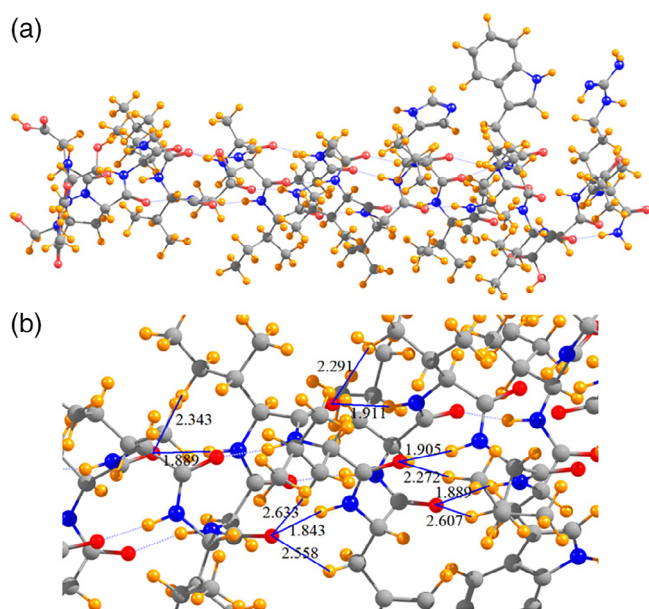


FIGURE 4 Diagram of (a) full model of M2TM and (b) subsection that indicates selected $\text{NH}\cdots\text{O}$ and $\text{CH}\cdots\text{O}$ hydrogen bonds, and their lengths in Å [Color figure can be viewed at wileyonlinelibrary.com]

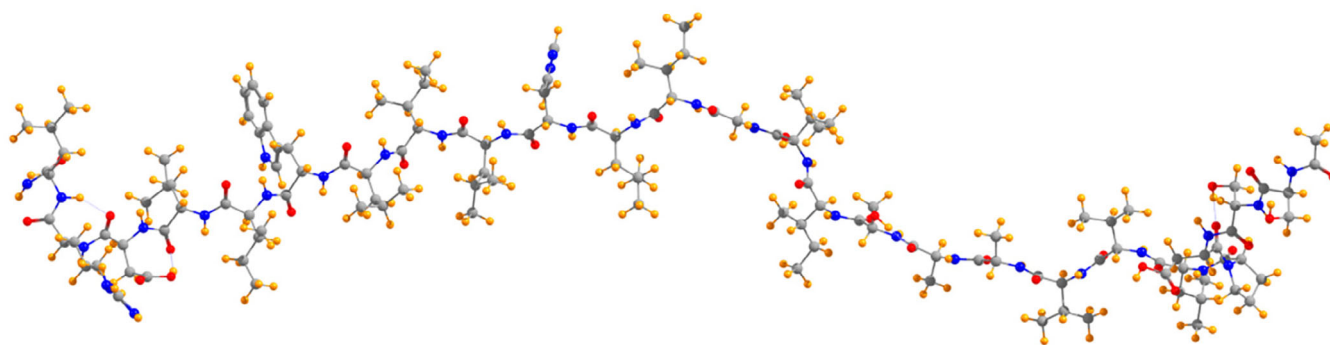


FIGURE 5 Optimized structure of M2TM resulted after energy optimization an extended β -sheet structure [Color figure can be viewed at wileyonlinelibrary.com]

QTAIM analysis, one might estimate that the CH \cdots O hydrogen bonds contribute roughly half as much as the NH \cdots O hydrogen bonds in M2TM, certainly not negligible. In Table 4, geometrical characteristics and E(2) values are shown for Ala₂₅ peptide. It is observed that N—H \cdots O hydrogen bonds are calculated 0.2–0.4 Å longer, and thus weaker, and E(2) mean value is much smaller in Ala₂₅ compared to M2TM. The E(2) values were also calculated with M06/D95(d,p) for NH \cdots O of both peptides and for CH \cdots O of M2TM and were similar to the values calculated with B3LYP/D95(d,p) (see Tables S2–S4).

4 | CONCLUSION

Upon folding, the protein minimizes the free energy of the protein-water system by clustering hydrophobic groups and forming intramolecular hydrogen bonds.^[43,102,107] It has been concluded using both experiments and simulations that in proteins where peptide hydrogen bonding interactions are needed for folding, a correct balance between side-chain-driven and backbone-driven self-association must be attained for proteins to fold correctly.^[43,102,107]

We performed CD experiments which show that M2TM is a stable α -helical peptide, as was shown previously for poly(alanine) peptides with CD experiments. MD simulations with adaptive tempering show that the influenza A M2TM 25-residues peptide appears to be more dynamic in nature than Ala₂₅ and quickly interconverts between an ensemble of various folded structures. The folding MD simulations show that although Ala₂₅ adopts mainly a single α -helix, M2TM equilibrates mainly between different α -helices and less frequently other folded structures, like turns and coils. A significant finding is that the residues with the highest thermal stability belong to the N-terminal part of the M2TM peptide, and not in Ala₂₅.

Our DFT calculations showed that folding from an extended structure to an α -helix of M2TM is favored compared with Ala₂₅ folding from β -strand to α -helix. Folding in M2TM is driven by NH \cdots O hydrogen bonding interactions and double in number CH \cdots O hydrogen bonding interactions between the side chains of the amino acid residues and peptidic carbonyls of the main chain, as well as the repulsive forces between amino acid residues.^[43,116] While CH \cdots O are individually weaker than NH \cdots O hydrogen bonding interactions, their cumulative effect cannot be ignored, and can contribute as much as half of the total hydrogen bonding interaction energy, when compared to NH \cdots O, to the higher propensity for folding to α -helix of M2TM compared to Ala₂₅. Apart CH \cdots O interactions, a strengthening of NH \cdots O hydrogen bonds in M2TM is also calculated. The CH \cdots O hydrogen bonding interactions are weaker than NH \cdots O hydrogen bonds,^[43,116] and can dissociate and reassociate easily. This in combination with the repulsive Pauli forces between the side chains lead to the higher flexibility and the observed ensemble of folded structures, mainly α -helical, for M2TM compared to Ala₂₅ peptide.

ACKNOWLEDGMENTS

Antonios Kolocouris gratefully acknowledges supporting of this research from Chiesi Hellas. The authors thank Prof Carlos Silva

Lopez for providing access to G09 program. The authors thank Olalla Nieto-Faza for running the QTAIM calculations.

AUTHOR CONTRIBUTIONS

Antonios Kolocouris designed the project; Nikolas Glykos performed the folding MD simulations and analysis of the results; Ioannis Stylianakis, Michael P. Sigalas, and Antonios Kolocouris performed the DFT, ONIOM, and NBO calculations; Steve Scheiner help on recording the NBO and QTAIM data; Antonios Kolocouris, Nikolas Glykos, and Steve Scheiner wrote the manuscript.

ORCID

Michael P. Sigalas  <https://orcid.org/0000-0002-4166-2057>

Nikolas Glykos  <https://orcid.org/0000-0003-3782-206X>

Antonios Kolocouris  <https://orcid.org/0000-0001-6110-1903>

REFERENCES

- [1] J. Liu, B. Rost, *Protein Sci.* **2001**, *10*, 1970.
- [2] M. Sallman Almen, K. Nordstrom, R. Fredriksson, H. Schioth, *BMC Biol.* **2009**, *7*, 50. <https://doi.org/10.1186/1741-7007-7-50>.
- [3] P. Hubert, P. Sawma, J. P. Duneau, J. Khao, J. Hénin, D. Bagnard, J. Sturgis, *Cell Adh. Migr.* **2010**, *4*, 313.
- [4] I. D. Pogozheva, A. L. Lomize, *Biochim. Biophys. Acta - Biomembr.* **2018**, *1860*, 364.
- [5] I. T. Arkin, A. T. Brunger, *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1998**, *1429*, 113.
- [6] R. F. S. Walters, W. F. DeGrado, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13658.
- [7] G. G. Rhys, C. W. Wood, J. L. Beesley, N. R. Zaccari, A. J. Burton, R. L. Brady, A. R. Thomson, D. N. Woolfson, *J. Am. Chem. Soc.* **2019**, *141*, 8787.
- [8] J. A. Mould, R. G. Paterson, M. Takeda, Y. Ohigashi, P. Venkataraman, R. A. Lamb, L. H. Pinto, *Dev. Cell* **2003**, *5*, 175.
- [9] S. D. C. Griffin, L. P. Beales, D. S. Clarke, O. Worsfold, S. D. Evans, J. Jaeger, M. P. G. Harris, D. J. Rowlands, H. D. Klenk, *FEBS Lett.* **2003**, *175*.
- [10] E. Arbely, Z. Khattari, G. Brotons, M. Akkawi, T. Salditt, I. T. Arkin, *J. Mol. Biol.* **2004**, *341*, 769.
- [11] C. Wang, K. Takeuchi, L. H. Pinto, R. A. Lamb, *J. Virol.* **1993**, *67*, 5585.
- [12] I. V. Chizhnikov, F. M. Geraghty, D. C. Ogden, A. Hayhurst, M. Antoniou, A. J. Hay, *J. Physiol.* **1996**, *494*, 329.
- [13] M. Yi, T. A. Cross, Zhou, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13311.
- [14] J. Wang, J. X. Qiu, C. Soto, W. F. DeGrado, *Curr. Opin. Struct. Biol.* **2011**, *21*, 68.
- [15] X. Jing, C. Ma, Y. Ohigashi, F. A. Oliveira, T. S. Jardetzky, L. H. Pinto, R. A. Lamb, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10967.
- [16] A. Helenius, *Cell* **1992**, *69*, 577.
- [17] A. Ianeselli, S. Orioli, G. Spagnoli, P. Faccioli, L. Cupellini, S. Jurinovich, B. Mennucci, *J. Am. Chem. Soc.* **2018**, *140*, 3674.
- [18] R. Woody, *Biomed. Spectrosc. Imaging* **2015**, *5*, 5.
- [19] A. M. S. Duarte, C. P. M. Van Mierlo, M. A. Hemminga, *J. Phys. Chem. B* **2008**, *112*, 8664.
- [20] A. Chakrabarty, T. Kortemme, R. L. Baldwin, *Protein Sci.* **1994**, *3*, 843.
- [21] M. Goodman, F. Toda, N. Ueyama, *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70*, 331.
- [22] N. W. Schmidt, A. Mishra, J. Wang, W. F. DeGrado, G. C. L. Wong, *J. Am. Chem. Soc.* **2013**, *135*, 13710.
- [23] R. L. Baldwin, *J. Biol. Chem.* **2003**, *278*, 17581.
- [24] K. T. O'Neil, W. F. DeGrado, *Science* **1990**, *250*, 646.

- [25] P. Luo, R. L. Baldwin, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 4930.
- [26] G. R. Desiraju, *Acc. Chem. Res.* **1996**, *4842*, 441.
- [27] G. R. Desiraju, T. Steiner, *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford University Press, New York **1999**.
- [28] S. K. Panigrahi, G. R. Desiraju, *Proteins Struct. Funct. Genet.* **2007**, *67*, 128.
- [29] J. Joseph, E. D. Jemmis, *J. Am. Chem. Soc.* **2007**, *129*, 4620.
- [30] S. Grimme, P. R. Schreiner, *Angew. Chem. - Int. Ed.* **2011**, *50*, 12639.
- [31] A. A. Fokin, L. V. Chernish, P. A. Gunchenko, E. Y. Tikhonchuk, H. Hausmann, M. Serafin, J. E. P. Dahl, R. M. K. Carlson, P. R. Schreiner, *J. Am. Chem. Soc.* **2012**, *134*, 13641.
- [32] J. P. Wagner, P. R. Schreiner, *Angew. Chem. - Int. Ed.* **2015**, *54*, 12274.
- [33] F. Weinhold, R. A. Klein, *Mol. Phys.* **2012**, *110*, 565.
- [34] A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* **1988**, *88*, 899.
- [35] I. V. Alabugin, M. Manoranhan, S. Peabody, F. J. Weinhold, *Am. Chem. Soc.* **2003**, *125*, 5973.
- [36] H. D. Nguyen, C. K. Hall, *J. Am. Chem. Soc.* **2006**, *128*, 1890.
- [37] S. E. Blondelle, B. Forood, R. A. Houghten, E. Pérez-Payá, *Biochemistry* **1997**, *36*, 8393.
- [38] E. J. Spek, C. A. Olson, Z. Shi, N. R. Kallenbach, *J. Am. Chem. Soc.* **1999**, *121*, 5571.
- [39] J. D. Sipe, A. S. Cohen, *J. Struct. Biol.* **2000**, *130*, 88.
- [40] R. F. W. Bader, H. Essén, *J. Chem. Phys.* **1984**, *80*, 1943.
- [41] R. F. W. Bader, *Chem. Rev.* **1991**, *91*, 893.
- [42] R. F. W. Bader, *J. Phys. Chem. A* **2009**, *113*, 10391.
- [43] A. W. Fitzpatrick, T. P. J. Knowles, C. A. Waudby, M. Vendruscolo, C. M. Dobson, *PLoS Comput. Biol.* **2011**, *7*, e1002169.
- [44] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668.
- [45] L. Kalé, R. Skeel, Bhandarkar, M. Brunner, R. A. Gursoy, N. Krawetz, J. Phillips, A. Shinzaki, K. Varadarajan, K. Schulten, *J. Comput. Phys.* **1999**, *151*, 283.
- [46] R. Chitra, P. E. Smith, *J. Chem. Phys.* **2001**, *114*, 426.
- [47] Y. Yu, J. Wang, Q. Shao, J. Shi, W. Zhu, *Sci. Rep.* **2016**, *6*, 1.
- [48] F. Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, P. Cieplak, *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821.
- [49] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins Struct. Funct. Bioinforma.* **2006**, *65*, 712.
- [50] L. Wickstrom, A. Okur, C. Simmerling, *Biophys. J.* **2009**, *97*, 853.
- [51] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 1950.
- [52] P. S. Georgoulia, N. M. Glykos, *J. Phys. Chem. B* **2011**, *115*, 15221.
- [53] K. K. Patapati, N. M. Glykos, *Biophys. J.* **2011**, *101*, 1766.
- [54] A. P. Serafeim, G. Salamanos, K. K. Patapati, N. M. Glykos, *J. Chem. Inf. Model.* **2016**, *56*, 2035.
- [55] P. I. Koukos, N. M. Glykos, *J. Phys. Chem. B* **2014**, *118*, 10076.
- [56] A. S. Baltzis, N. M. Glykos, *Protein Sci.* **2016**, *25*, 587.
- [57] P. S. Georgoulia, N. M. Glykos, *J. Phys. Chem. B* **2013**, *117*, 5522.
- [58] K. K. Patapati, N. M. Glykos, *PLoS One* **2010**, *5*, e15290.
- [59] I. Patmanidis, N. M. Glykos, *J. Mol. Graph. Model.* **2013**, *41*, 68.
- [60] A. M. Razavi, V. A. Voelz, *J. Chem. Theory Comput.* **2015**, *11*, 2801.
- [61] R. B. Best, W. W. Zheng, J. Mittal, *J. Chem. Theory Comput.* **2014**, *10*, 5113.
- [62] S. Piana, A. G. Donchev, P. Robustelli, D. E. Shaw, *J. Phys. Chem. B* **2015**, *119*, 5113.
- [63] D. Mercadante, S. Milles, G. Fuertes, D. I. Svergun, E. A. Lemke, F. Gräter, *J. Phys. Chem. B* **2015**, *119*, 7975.
- [64] S. Piana, J. L. Klepeis, D. E. Shaw, *Curr. Opin. Struct. Biol.* **2014**, *24*, 98.
- [65] T. Adamidou, K.-O. Arvaniti, N. M. Glykos, *J. Phys. Chem. B* **2018**, *122*, 106.
- [66] C. Zhang, J. Ma, *J. Chem. Phys.* **2010**, *132*, 244101.
- [67] J. A. Izaguirre, D. P. Catarella, J. M. Wozniak, R. D. Skeel, *J. Chem. Phys.* **2001**, *114*, 2090.
- [68] G. J. Martyna, M. L. Klein, M. Tuckerman, *J. Chem. Phys.* **1992**, *97*, 2635.
- [69] S. E. Feller, Y. Zhang, R. W. Pastor, B. R. Brooks, *J. Chem. Phys.* **1995**, *103*, 4613.
- [70] J. A. Izaguirre, S. Reich, R. D. Skeel, *J. Chem. Phys.* **1999**, *110*, 9853.
- [71] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089.
- [72] J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327.
- [73] N. M. Glykos, *J. Comput. Chem.* **2006**, *27*, 1765.
- [74] P. I. Koukos, N. M. Glykos, *J. Comput. Chem.* **2013**, *34*, 2310.
- [75] A. S. Baltzis, P. I. Koukos, N. M. Glykos, *arXiv* **2015**, 04024.
- [76] A. Amadei, A. B. M. Linssen, H. J. C. Berendsen, *Proteins Struct. Funct. Genet.* **1993**, *17*, 412.
- [77] C. C. David, D. J. Jacobs, *Meth. Mol. Biol.* **2014**, 193.
- [78] D. Frishman, P. Argos, *Proteins Struct. Funct. Bioinforma.* **1995**, *23*, 566.
- [79] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L. P. Wang, T. J. Lane, V. S. Pande, *Biophys. J.* **2015**, *109*, 1528.
- [80] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33.
- [81] D. Bacon, W. Anderson, *J. Mol. Graph.* **1988**, *6*, 219.
- [82] W. L. DeLano, *Protein Crystallogr.* **2002**, *40*, 82.
- [83] G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, *Genome Res.* **2004**, *14*, 1188.
- [84] P. Salvador, A. Asensio, J. J. Dannenberg, *J. Phys. Chem. B* **2007**, *111*, 7462.
- [85] A. L. Stouffer, R. Acharya, D. Salom, A. S. Levine, L. Di Costanzo, C. S. Soto, V. Tereshko, V. Nanda, S. Stayrook, W. F. DeGrado, *Nature* **2008**, *451*, 596.
- [86] J. L. Thomaston, M. Alfonso-Prieto, R. A. Woldeyes, J. S. Fraser, M. L. Klein, G. Fiorin, W. F. DeGrado, *Proc. Natl. Acad. Sci.* **2015**, *112*, 14260.
- [87] J. L. Thomaston, R. A. Woldeyes, T. Nakane, A. Yamashita, T. Tanaka, K. Koizumi, A. S. Brewster, B. A. Barad, Y. Chen, M. Uervirojnangkoorn, T. Arima, J. Kobayashi, T. Masuda, M. Suzuki, M. Sugahara, N. K. Sauter, R. Tanaka, O. Nureki, K. Tono, Y. Joti, E. Nango, S. Iwata, F. Yumoto, J. S. Fraser, W. F. X. F. E. L. DeGrado, S. Iwata, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 13357.
- [88] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09*, Gaussian, Inc., Wallingford, CT **2009**.
- [89] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
- [90] P. Hohenberg, *Phys. Rev.* **1964**, *136*, 864.
- [91] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, 1133.
- [92] T. Vreven, K. Morokuma, Ö. Farkas, H. B. Schlegel, M. J. Frisch, *J. Comput. Chem.* **2003**, *24*, 760.
- [93] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H. B. Li, L. Ding, K. Morokuma, *Chem. Rev.* **2015**, *115*, 5678.
- [94] J. N. Woodford, *J. Phys. Chem. A* **2007**, *111*, 8519.

- [95] S. J. Grabowski, *J. Phys. Chem. A* **2001**, 105, 10739.
- [96] L. F. Pacios, *J. Phys. Chem. A* **2004**, 108, 1177.
- [97] R. Parthasarathi, S. Sundar Raman, V. Subramanian, T. Ramasami, *J. Phys. Chem. A* **2007**, 111, 7141.
- [98] S. M. Lapointe, S. Farrag, H. J. Boho, R. J. Boyd, *J. Phys. Chem.*, Vol. 113 **2009**, 10957.
- [99] U. Koch, P. L. A. Popelier, *J. Phys. Chem.* **1995**, 99, 9747.
- [100] P. L. A. Popelier, P. L. A. Popelier, *J. Phys. Chem. A* **1998**, 5639, 1873.
- [101] T. A. Keith, *AIMALL (version 13.11.04)*, TK Gristmill Software, Overland Park, KS **2013**.
- [102] E. S. Brielle, I. T. Arkin, *J. Phys. Chem. Lett.* **2018**, 9, 4059.
- [103] E. S. Feldblum, I. T. Arkin, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, 111, 4085.
- [104] J. Manor, E. S. Feldblum, M. T. Zanni, I. T. Arkin, *J. Phys. Chem. Lett.* **2012**, 3, 939.
- [105] D. A. E. Cochran, *Protein Sci.* **2001**, 10, 463.
- [106] F. Hu, W. Luo, S. D. Cady, M. Hong, *Biochim. Biophys. Acta - Bio-membr.* **2011**, 1808, 415.
- [107] W. Luo, M. Hong, *J. Am. Chem. Soc.* **2010**, 132, 2378.
- [108] V. S. Mandala, M. D. Gelenter, M. Hong, *J. Am. Chem. Soc.* **2018**, 140, 1514.
- [109] J. L. Thomaston, M. Alfonso-Prieto, R. A. Woldeyes, J. S. Fraser, M. L. Klein, G. Fiorin, W. F. DeGrado, *Proc. Natl. Acad. Sci.* **2015**, 112, 14260.
- [110] D. Salom, B. R. Hill, J. D. Lear, W. F. DeGrado, *Biochemistry* **2000**, 39, 14160.
- [111] L. Cristian, J. D. Lear, W. F. DeGrado, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100, 14772.
- [112] B. N. Markiewicz, T. Lemmin, W. Zhang, I. A. Ahmed, H. Jo, G. Fiorin, T. Troxler, W. F. DeGrado, F. Gai, *Phys. Chem. Chem. Phys.* **2016**, 18, 28939.
- [113] B. Brais, G. A. Rouleau, J. P. Bouchard, M. Fardeau, F. M. S. Tomé, *Semin. Neurol.* **1999**, 19, 59.
- [114] T. H. Dunning, P. J. Hay, in *Modern Theoretical Chemistry*, III ed. (Ed: H. F. Schaefer), Plenum, New York **1976**.
- [115] J. Stone, *J. Phys. Chem. A* **2017**, 121, 1531.
- [116] T. X. Hoang, A. Trovato, F. Seno, J. R. Banavar, A. Maritan, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 7960.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Stylianakis I, Shalev A, Scheiner S, et al. The balance between side-chain and backbone-driven association in folding of the α -helical influenza A transmembrane peptide. *J Comput Chem.* 2020;1–12. <https://doi.org/10.1002/jcc.26381>