JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

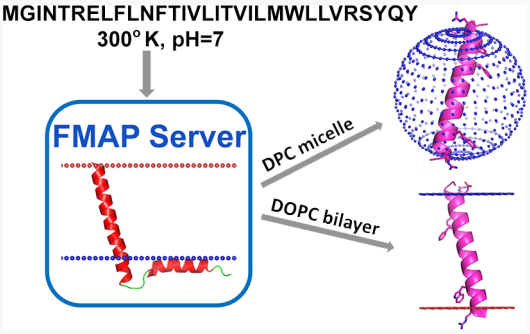
pubs.acs.org/jcim Article

Thermodynamics-Based Molecular Modeling of α-Helices in Membranes and Micelles

Andrei L. Lomize,* Kevin A. Schnitzer, Spencer C. Todd, and Irina D. Pogozheva

Cite This: https://doi.org/10.1021/acs.jcim.1c00161 Read Online

ABSTRACT: The Folding of Membrane-Associated Peptides (FMAP) method was developed for modeling α-helix formation by linear peptides in micelles and lipid bilayers. FMAP 2.0 identifies locations of α-helices in the amino acid sequence, generates their three-dimensional models in planar bilayers or spherical micelles, and estimates their thermodynamic stabilities and tilt angles, depending on temperature and pH. The method was tested for 723 peptides (926 data points) experimentally studied in different environments and for 170 single-pass transmembrane (TM) proteins with available crystal structures. FMAP 2.0 detected more than 95% of experimentally observed α-helices with an average error in helix end determination of around 2, 3, 4, and 5 residues per helix for peptides in water, micelles, bilayers, and TM proteins, respectively. Helical and nonhelical residue states were predicted with an accuracy from 0.86 to 0.96, and the Matthews correlation coefficient ~~had an accuracy~~ from 0.64 to 0.88 depending on the environment. Experimental micelle- and membrane-binding energies and tilt angles of peptides were reproduced with a root-mean-square deviation of around 2 kcal/mol and 7°, respectively. The TM and non-TM states of hydrophobic and pH-triggered α-helical peptides in various lipid bilayers were reproduced in more than 95% of cases. The FMAP 2.0 web server (https://membranome.org/fmap) is publicly available to explore the structural polymorphism of antimicrobial, cell-penetrating, fusion, and other membrane-binding peptides, which is important for understanding the mechanisms of their biological activities.



■ INTRODUCTION

Membrane-interacting peptides play important roles in many vital cellular processes, including cell defense, molecular transport, membrane fission and fusion, enzymatic regulation, and signaling.[1] They belong to different classes, such as antimicrobial,[2] cell-penetrating[3] and fusion[4] peptides, toxins,[5] and others. To perform their biological functions, these peptides usually form α-helices and multihelical complexes. For example, transmembrane (TM) α-helices of signal peptides direct localization and translocation across membranes of secreted and numerous integral membrane proteins.[6] Individually stable TM α-helices of single-pass (bitopic) membrane proteins constitute their membrane-anchoring domains which play important functional roles in the formation of signaling complexes and TM pores and in guiding protein sorting and intracellular localization.[7]

The quickly expanding universe of biologically active peptides requires the advancement of experimental and computational methods for their structural studies. Such methods are needed for understanding the molecular mechanisms of peptide activities and for the development of peptide-based drugs. Experimental studies have produced a vast set of data, including peptide secondary and three-dimensional (3D) structures, energies of membrane binding, and spatial arrangements of peptides in micelles and lipid bilayers. These studies demonstrated that linear peptides frequently do not have a unique folded structure. Their conformations are highly flexible and strongly dependent on environmental conditions, such as solvent, temperature, and pH. Many peptides are unstructured in aqueous solutions at physiological conditions but can fold into α-helices or other structures upon binding to proteins, micelles, or membranes.[8] Formation of β-structures usually requires peptide aggregation in water or on the membrane surface or stabilization by disulfide cross-linking.

In addition to experimental approaches, diverse computational methods for de novo structure prediction of peptides and small proteins (up to 50 amino acids) have been developed.[9] Despite the apparent progress in peptide modeling, the currently available web servers and software, such as PEPstrMOD,[10] Bhageerath-H,[11] PEP-FOLD3,[12] and RosettaMP,[13] are aimed at generating a unique structure of a peptide of interest without considering the environment-dependent equilibrium of multiple alternative structures and

Received: February 10, 2021

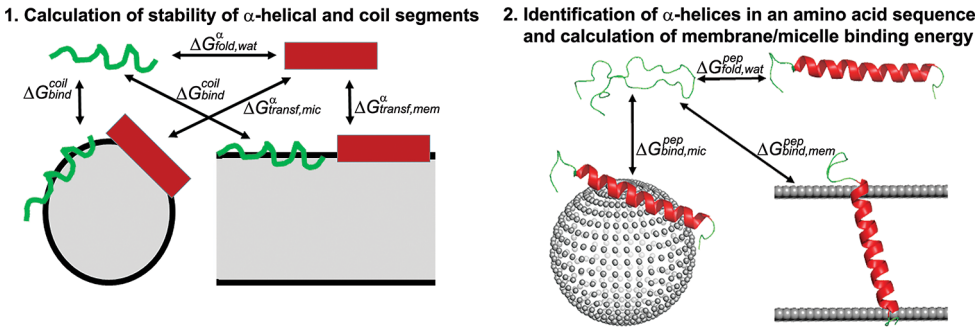


Figure 1. Two steps of the computational procedure of the FMAP 2.0 method. (1) Calculation of stabilities of helical and coil states for every segment of a polypeptide chain using eqs 5−13. (2) Identification of experimentally detectable α-helical and coil regions of a peptide, calculation of the peptide binding energy to the membrane or micelle, and generation of the 3D structures of α-helices (excluding coil segments) positioned in lipid bilayers or a spherical micelle.

the unfolded state. In contrast, the thermodynamic theories of the helix−coil transition, such as modified Zimm−Bragg, Lifson−Roig[14] or AGADIR[15] models, operate with free energy and can reproduce α-helix formation depending on the ionic strength, temperature, and pH. However, these methods have been developed and parametrized to analyze α-helicity only in aqueous solutions but not in membrane-like environments. The structural flexibility of peptides in lipid membranes can be explored using molecular dynamics (MD) simulations with explicit lipids or continuum models of lipid bilayers.[9] However, MD simulations are rather complex for general use by nonexperts, highly computationally expensive, and not available as online resources.

To overcome the limitations of existing computational methods, we previously developed Framework,[16] a thermodynamics-based method for predicting α-helices in peptides in aqueous solutions, in the protein molten globule state, or in the presence of micelles, depending on the temperature and pH. However, Framework did not allow the generation of 3D models of peptides. More recently, we developed the first version of our FMAP (Folding of Membrane-Associated Peptides) software to predict TM α-helices of bitopic proteins using a whole-residue approximation.[17] FMAP 1.0 generated approximate 3D models of TM but not surface α-helices without optimization of the α-helix geometry and side-chain conformers. Therefore, it did not allow for accurate evaluation of the thermodynamic stabilities and spatial arrangement of α-helices in membranes and micelles.

Here, we present version 2 of the FMAP method and a web server for the modeling and structural analysis of α-helical peptides in lipid membranes and membrane mimetics. The method was significantly advanced by employing an all-atom approximation to enable, for the first time, the following new features: (a) predicting stable α-helical segments in peptides depending on the experimental system, temperature, and pH, (b) generating all-atom 3D models of α-helices with proper adjustment of the α-helix geometry and optimization of side-chain rotamers simultaneously with α-helix positioning in planar membranes or spherical micelles; (c) accurate estimation of helix tilt angles and binding free energies of peptides to such systems; (d) calculating peptide properties in four types of micelles, four types of artificial membranes, and seven types of natural membranes. The performance of FMAP 2.0 was tested for α-helical peptides studied in water, micelles, and lipid bilayers (118, 460, and 348 data points, respectively) and for 170 bitopic membrane proteins with known crystal structures.

METHODS

Overview of FMAP 2.0. The FMAP 2.0 method employs a thermodynamic model of α-helix formation with an empirical parametrization of various free energy contributions that have been previously implemented in the Framework,[16] PPM,[18] TMDOCK,[19] and FMAP 1.0[17] methods. Framework defines the α-helix stability in an aqueous solution as the sum of the free energy contributions arising from formation of backbone hydrogen bonds, α-helix propensities, capping and other structural motifs, and interactions of side chains with each other and helix macrodipoles.[16] The contributions have been previously derived primarily from experimental data, similar to that in the AGADIR method.[15] The improved PPM method calculates the transfer energy of an α-helix from water to a planar lipid bilayer or a spherical micelle using an empirical parametrization of the first-shell solvation effects, long-range electrostatic contributions of dipole moments and charges, and a deionization penalty. The parameters of the solvation model were derived from a large set of partition coefficients of small molecules,[18,20] while the distributions of various lipid segments along the normal of the lipid bilayer were taken from X-ray scattering studies.[18,21] Parameters of interatomic potentials for the local energy minimization in condensed media were derived from stabilities of protein mutants.[22]

Computational Procedure. Different α-helical and coil segments in a peptide molecule can compete with each other for binding to a micelle or a lipid bilayer. This can be described as an equilibrium of various partitions of a polypeptide chain into α-helical and coil segments.[16] Assuming that α-helical and coil segments do not interact with each other, the energy of an α-helix−coil partition l (ΔG_l) is the sum of the energies of the corresponding membrane- or micelle-bound α-helical and coil segments

$$\Delta G_l^{\text{mem}} = \sum_i \Delta G^{\alpha}_{\text{fold,mem}}(k_i, m_i) + \sum_j \Delta G^{\text{coil}}_{\text{bind}}(k_j, m_j) \quad (1)$$

$$\Delta G_l^{\text{mic}} = \sum_i \Delta G^{\alpha}_{\text{fold,mic}}(k_i, m_i) + \sum_j \Delta G^{\text{coil}}_{\text{bind}}(k_j, m_j) \quad (2)$$

where k_i is the number of the first residue in segment i and m_i is the number of residues in the segment. The ΔG_l value defines which set of α-helical segments will have a relatively

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

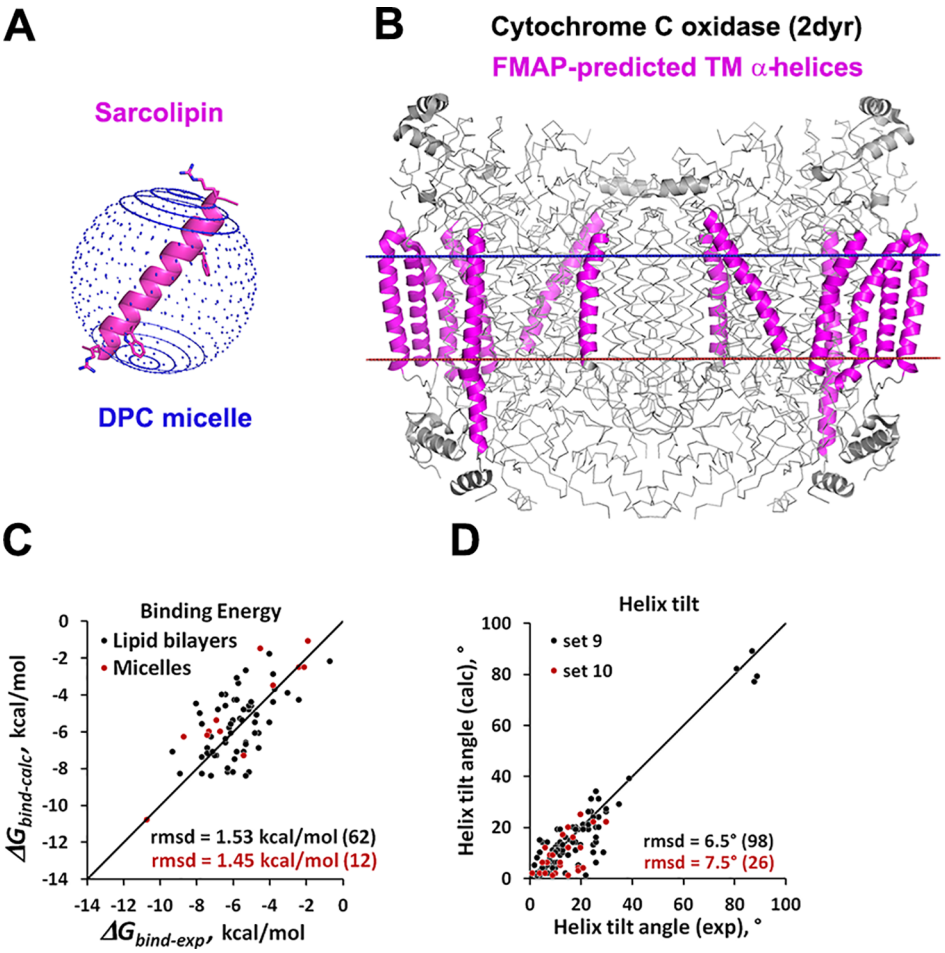


Figure 2. FMAP 2.0 features. (A) Modeling of peptides in spherical micelles: purple color indicates a predicted α-helix and its aromatic and charged residues (shown by sticks). (B) Identification of TM α-helices of bitopic proteins (purple) and their positioning in planar lipid bilayers, extramembrane helices of bitopic proteins (from X-ray structure) are colored gray, and other structural elements of polytopic proteins are shown by thin gray lines. (C) Evaluation of the peptide binding energies to lipid bilayers (black circles) and micelles (red circles). Correlation coefficient (R^2) is 0.39 for both sets combined. (D) Calculation of helix tilt angles for peptides in lipid bilayers. R^2 is 0.84 for sets 9 and 10 combined. Number of peptides is indicated in parentheses.

low energy or a significant statistical weight and therefore can be experimentally detected. The folding free energy of α-helices, $\Delta G^{\alpha}_{\text{fold}}(k,m)$, and energies of bound coil segments in eqs 1 and 2 are defined relative to the coil in water, which is considered as the reference state.

The stabilities of α-helices are defined relative to the corresponding bound coil segments

$$\Delta G^{\alpha}_{\text{stab,mem}} = \Delta G^{\alpha}_{\text{fold,mem}}(k, m) - \Delta G^{coil}_{\text{bind,mem}}(k, m) \quad (3)$$

$$\Delta G^{\alpha}_{(\text{stab,mic})} = \Delta G^{\alpha}_{(\text{fold,mic})}(k, m) - \Delta G^{coil}_{(\text{bind,mic})}(k, m) \quad (4)$$

The first step of the computational procedure includes calculating the stabilities of the α-helical and coil states for every segment of a peptide using eqs 5−13 (Figure 1).

The folding free energy of a membrane- or micelle-bound α-helix (first term in eqs 3 and 4) is calculated as the sum of its folding energy in water and the energy of its transfer from water to a membrane or a micelle

$$\Delta G^{\alpha}_{\text{fold,mem}}(k, m) = \Delta G^{\alpha}_{\text{fold,wat}}(k, m) + \Delta G^{\alpha}_{\text{transf,mem}}(k, m) \quad (5)$$

$$\Delta G^{\alpha}_{\text{fold,mic}}(k, m) = \Delta G^{\alpha}_{\text{fold,wat}}(k, m) + \Delta G^{\alpha}_{\text{transf,mic}}(k, m) \quad (6)$$

The transfer energy of an α-helix from water to a membrane or a micelle in eqs 5 and 6 is calculated using either the whole-residue or the all-atom approximation as described in the next two sections. The folding free energy of an α-helix in water is calculated as the sum of the backbone free energy changes, α-helix propensities of residues (the backbone−side-chain interactions), and energy of interactions between side chains

$$\Delta G^{\alpha}_{\text{fold,wat}}(k, m) = (m - 2)\Delta H_{\text{bb}} - mT\Delta S_{\text{bb}} + \sum_i \Delta\Delta G_i^{\text{prop}} + \sum_i \sum_j \Delta\Delta G_{ij}^{\text{sch-sch}} \quad (7)$$

where k is the number of the first residue in the α-helix, m is the total number of residues in the helix, and T is the temperature (K). The backbone energy includes the enthalpy of hydrogen bonds between peptide groups in the poly-Ala α-helix (ΔH_{bb}) and conformational entropy loss per residue during the helix−coil transition (ΔS_{bb}). The $\Delta\Delta G_i^{\text{prop}}$ term is calculated as the sum of the experimentally determined free energy changes to the α-helix stability due to replacement of

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

the host Ala by different amino acid residues in the α-helix,[23−25] "N-capping", "C-capping", and "hydrophobic staple" positions, and from interactions between charged side chains and helix microdipoles.[26−28] The stabilizing energies of the ions pairs and interacting hydrophobic side chains ($\Delta\Delta G_{ij}^{\text{sch-sch}}$ values) were taken from experimental studies[25,29,30] or estimated from the calculated buried nonpolar surface areas between different types of residues in the α-helix.[16]

The binding free energy of a micelle- or membrane-bound coil segment (second term in eqs 3 and 4) is calculated as

$$\Delta G_{\text{bind}}^{\text{coil}}(k, m) = \sum_{i} (\Delta G_{\text{bind},i}^{\text{coil}} - \Delta G_{\text{bind,ref}}^{\text{coil}}) \tag{8}$$

where $\Delta G_{\text{bind},i}^{\text{coil}}$ are the membrane-binding free energies for different types of amino acid residues in the unfolded state and $\Delta G_{\text{bind,ref}}^{\text{coil}}$ is the binding energy of the reference Ala residue. These energies (Table S 1) were chosen based on studies of peptide binding to 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) bilayers[31,32] and sodium dodecyl sulfate (SDS) micelles.[33]

The second step of the computational procedure (Figure 1) includes identification of experimentally detectable α-helices in the amino acid sequence, generation of the final 3D structures of stable α-helices positioned in a planar lipid bilayer or a spherical micelle (Figure 2 A), calculation of the membrane- or micelle-binding energies (Figure 2 C) for the whole peptide containing all α-helical and coil regions, and estimation of the tilt angles of the α-helices in the lipid bilayer (Figure 2 D).

In the second step, two different approaches to identifying α-helices are used. In the first approach, FMAP 2.0 calculates the energies of different helix−coil partitions (eq 1 and 2) and performs their statistical (Boltzmann) averaging as described previously.[16] This allows calculation of the average α-helix occupancy for every residue, P_i, which varies between 0 and 1, similar to the AGADIR method.[15] Then, the NMR-detectable α-helices are identified as continuous segments with the occupancy of all helix turns P_i larger than a cutoff (P_d). We found that solution NMR data for peptides in water, micelles, and bicelles were reproduced best with $P_d = 0.2$. However, to analyze the membrane binding energies and tilt angles of peptides in membranes, we used only α-helices calculated with $P_d = 0.5$. The total binding energy of a peptide to a lipid bilayer ($\Delta G_{\text{bind,mem}}^{\text{pep}}$) or a micelle ($\Delta G_{\text{bind,mic}}^{\text{pep}}$) is calculated as the sum of the binding energies of the identified α-helical and coil segments.

It the second approach, the lowest energy partition (LEP) of a polypeptide chain containing M residues into α-helical and coil segments was calculated in a recurrent manner using the dynamic programming algorithm, i.e., by considering its fragments growing from the C- to the N-terminus, $[M - 1, M]$, $[M - 2, M]$,⋯, $[M - n, M]$,⋯, $[1,M]$, and calculating the corresponding LEP for each fragment, as previously described.[16] This approach is preferred when the locations of the α-helices in the amino acid sequence are not dynamic, but uniquely defined, as in protein structures. We are using this method for detecting α-helices in bitopic proteins and very long peptides (>50 residues).

Calculating the Transfer Energies of α-Helices with a Whole-Residue Approximation. A whole-residue approximation is used for TM α-helices of bitopic membrane proteins to assess their transfer energies from water to the lipid bilayer. In such an approximation, the transfer energy term in eqs 5 and 6 is calculated as the sum of the membrane-depth (z)-dependent transfer energies of the residues ($\Delta G_i^{\alpha}(z_i)$) and the energy of membrane deformation (ΔG_{def}) due to structural changes in response to a hydrophobic mismatch and a TM helix tilting

$$\Delta G_{\text{transf}}^{\alpha}(k, m) = \sum_{i=k}^{k+m-1} \Delta G_i^{\alpha}(z_i) + \Delta G_{\text{def}} \tag{9}$$

where i is the number of the amino acid residue, k is the number of the first residue in the helix, m is the number of residues in the helix, and $\Delta G_i^{\alpha}(z_i)$ are the energy profiles for different types of residues which were precalculated and tabulated with a step of 0.5 Å for different residues scanned along of the poly-Ala TM α-helix immersed into the 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) bilayer.[18]

The bilayer deformation energy is calculated as

$$\Delta G_{\text{def,res}} = f_{\text{mism}} |D - D_0| + f_{\text{tilt}} L \sin(\tau) \tag{10}$$

where D_0 and D represent the hydrophobic thicknesses of a lipid bilayer in equilibrium and after deformation, respectively, L is the length of the hydrophobic TM segment of an α-helix, and τ is the helix tilt angle with respect to the bilayer normal.

The linear dependence on the extent of mismatch ($D - D_0$) was taken based on experimental studies of the protein−lipid interaction energy dependence on the lipid acyl chain length.[34] The $L \sin(\tau)$ expression was chosen because the deformation energy is expected to depend on the projection of the tilt vector to the membrane plane.[35] The optimal values of the adjustable parameters f_{tilt} (0.04 kcal/mol Å^2) and f_{mism} (0.8 or 0.03 kcal/mol Å^2 for negative or positive mismatch, respectively) were determined in our previous work[17] by minimizing the average helix boundary deviation for a set of bitopic proteins with known 3D structures.

For each TM α-helix, the value of $\Delta G_{\text{transf}}^{\alpha}(k,m)$ is optimized by a grid scan with respect to three variables: the shift of the first residue of the helix along the membrane normal (z_o), membrane thickness (D_0), and helix tilt angle (τ). The optimization is constrained by values of $\tau < 30°$ and $D = D_0 \pm 5$ Å with $D_0 = 30$ Å.

This fast whole-residue approach is also used for peptides in micelles in two cases: (1) to analyze peptides longer than 35 residues (in web server version) and (2) to identify the micelle-buried arc of each α-helix for selecting the initial side-chain conformers. The size of the buried helix arc is determined by minimizing the sum of the water−micelle transfer energies of the residues located in the arc, as previously described.[16] These energies were taken from our previous work as those representing the transfer from water to the bilayer center for all residues except Tyr and Trp, where we used transfer energies from water to the membrane interface[18] (Table S1).

Calculating the Transfer Energies of α-Helices with an All-Atom Approximation. To improve the accuracy in calculating the transfer energies of the α-helices from water to a micelle or a lipid bilayer, all-atom models of α-helices are generated during both steps of the computational procedure (Figure 1). First, each α-helix is generated with the initial side-chain conformers, optimized in the space of the torsion angles, and positioned in a micelle or a lipid bilayer. Subsequently, the side-chain conformer with the lowest transfer energy is selected individually for each residue. Then, the helix geometry and its spatial position are optimized once more. To speed-up calculations, only side-chain conformers that are energetically

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

preferred in the α-helix and represent distinct combinations of χ^1 and χ^2 torsion angles are used during the optimization. The initial rotamers are chosen to provide orientations of nonpolar side chains (Leu and Met) toward the hydrophobic core of a bilayer or a micelle and "snorkeling" of polar side chains (Asn, Gln, Asp, Glu, Lys, Arg) toward water or the interfacial region.

Three-dimensional models of α-helices are generated and optimized in the space of the φ, ψ, and χ torsion angles using modules from ConforNMR with the modified ECEPP/2 force field implemented in TMDOCK.[19] Spatial positioning of the α-helices in membranes and micelles is performed using a new version of the PPM program[18,36] and the anisotropic solvent model for the bilayer[18,21] and micelles. The advanced PPM method has several new features: (1) positioning in spherical micelles; (2) including the membrane deformation penalty due to helix insertion; (3) a faster simplified optimization of the transfer energy (including deformation) by a grid scan with a gradually decreasing step.

The deformation energy due to insertion of a peptide into a lipid bilayer is calculated as

$$\Delta G_{\text{def,all-atom}} = C_s \sum_i \text{ASA}_i + N_L f_{\text{mism}} (D - D_0)^2 + N_L f_{\text{tilt}} (D \tan(\tau))^2 \quad (11)$$

where C_s is the effective solvation parameter, ASA_i is the accessible surface area of an atom i inserted into the hydrocarbon core of the bilayer, and N_L is the number of annular lipids in two leaflets around a TM domain (for a TM α-helix, $N_L = 10$); all other parameters are defined in eq 10. Only the first, ASA-dependent term in eq 11 is used for micelles. The optimal values of adjustable parameters f_{mism} (0.02 kcal/mol/Å^2) and f_{tilt} (0.0005 kcal/mol/Å^2) were determined in this work by minimizing the deviation of the experimental and calculated tilt angles for a subset of model peptides studied by solid-state NMR. We found that the quadratic dependence reproduces the experimental data better than the linear one, as expected from theoretical considerations of membrane deformations.[35,37]

As described previously,[18] PPM calculates the free energy of transfer of a molecule from water to the anisotropic membrane-like environment. This energy is a sum of short-range ASA-dependent contributions for all atoms (H bonds, van der Waals, and hydrophobic interactions with solvent), long-range electrostatic contributions of dipole moments and charged groups, and the ionization penalty for ionizable groups

$$\Delta G^{\alpha}_{\text{transf}}(z) = \sum_i \sigma_i^{\text{wat}\to\text{bil}}(z_i)\text{ASA}_i + \sum_j \eta^{\text{wat}\to\text{bil}}(z_j)\mu_j + \sum_k \min\{E_k^{\text{ion}}, E_k^{\text{neutr}}\} \quad (12)$$

where the z coordinate defines the position of each atom along the bilayer normal, $\alpha_i^{\text{wat}\to\text{bil}}(z_i)$ is the solvation parameter that depends on the type of atom and describes its transfer energy (per Å^2) from water to point z_i, μ_j is a group dipole moment, $\eta^{\text{wat}\to\text{bil}}(z_j)$ is the energy cost of transferring the dipole moment of 1 D from water to point z_j, and E_k^{ion} and E_k^{neutr} are the energies of ionizable group k in ionized and neutral states, respectively.

Parameters $\alpha_i^{\text{wat}\to\text{bil}}(z_i)$ and $\eta^{\text{wat}\to\text{bil}}(z_j)$ are functions of the polarity of the environment. They are defined by transbilayer profiles of the hydrogen-bond donor and acceptor capacities and the dielectric constant, $\alpha_{\text{bil}}(z)$, $\beta_{\text{bil}}(z)$, and $\varepsilon_{bil}(z)$, respectively. For example

$$\sigma_i^{\text{wat}\to\text{bil}}(z) = \sigma_i^0 - e_i\left(\frac{1}{\varepsilon_{\text{bil}}(z)} - \frac{1}{\varepsilon_{\text{wat}}}\right) + a_i(\alpha_{\text{bil}}(z) - \alpha_{\text{wat}}) + b_i(\beta_{\text{bil}}(z) - \beta_{\text{wat}}) \quad (13)$$

where α_{wat}, β_{wat}, and ε_{wat} are the corresponding values in water. The coefficients σ_i^0, e_i, a_i, and b_i were derived for different types of atoms from the partition coefficients of small organic compounds between water and 19 organic solvents during development of the corresponding universal solvation model.[20]

The transbilayer profiles of the polarity parameters were calculated based on the distributions of different lipid groups in the DOPC bilayer determined by X-ray scattering.[21] These polarity profiles were also applied to micelles using the spherical coordinate system, i.e., as functions of the distance r from the center of a micelle instead of the distance z from the middle plane of the bilayer. The values of the equilibrium hydrophobic thickness (D_0) of planar phosphatidylcholine (PC) bilayers were taken as 28.8 Å for DOPC and POPC, 25.7 Å for 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), and 21.7 Å for 1,2-dilauroyl-*sn*-glycero-3-phosphocholine (DLPC) bilayers.[38,39] On the basis of the experimental studies of micelles, the hydrophobic diameters were fixed as 37 Å for SDS, 26 Å for 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC),[40,41] and 39 Å for *n*-dodecyl-phosphocholine (DPC) micelles.[42]

Performance Evaluation Measures. To estimate the performance of FMAP 2.0 for α-helix prediction, we used several standard measures,[43,44] such as the precision of helix prediction (PRE_H) and the recall of helix prediction (REC_H) that are defined as

$$\text{PRE}_\text{H} = 100 \times \frac{\text{no. of correctly predicted α-helices}}{\text{no. of predicted α-helices}} \quad (14)$$

$$\text{REC}_\text{H} = 100 \times \frac{\text{no. of correctly predicted α-helices}}{\text{no. of observed α-helices}} \quad (15)$$

The residue-based precision (PRE_R, %) and recall (REC_R, %) are defined as[44]

$$\text{PRE}_\text{R} = 100 \times \frac{\text{no. of correctly predicted helical residues}}{\text{no. of predicted helical residues}} \quad (16)$$

$$\text{REC}_\text{R} = 100 \times \frac{\text{no. of correctly predicted helical residues}}{\text{no. of observed helical residues}} \quad (17)$$

To measure the quality of binary classification into helical and nonhelical residues, we use the accuracy (ACC) and the Matthews correlation coefficient (MCC)[44]

$$\text{ACC} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})) \quad (18)$$

$$\text{MCC} = (\text{TP} \times \text{TN} - \text{FP} \times \text{FN}) / \sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})} \quad (19)$$

where TP, TN, FP, and FN represent the numbers of true positives, true negatives, false positives, and false negatives, respectively.

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Performance of the FMAP 2.0 Method for Predicting α-Helices in Different Environments

system	water	SDS or DPC micelles	various micelles	various micelles	various micelles	lipid bilayers[a]	bitopic proteins
approximation	whole residue	all atom	all atom	all atom	all atom	all atom	whole residue
helix detection method	Boltzmann	Boltzmann	Boltzmann	LEP	Boltzmann	Boltzmann	LEP
data set	1	2a	2b	2c	2a−2b	3	11
no. of peptides[b]	118 (65,49,4,0,0)	255 (3,230,22,0,0)	152 (12,91,49,0,0)	10 (0,1,5,3,1)	407 (15,321,71,0,0)	34 (3,31,0,0,0)	170 (0,170,0,0,0)
no. of observed α-helices	57	274	189	24	463	31	170
prediction of α-helices							
PRE_H (%)	100	98	95	92	96	100	95
REC_H (%)	95	100	99	96	100	100	100
helix end errors (residue/helix)	2.1	2.0	3.4	5.5	2.6	4.0	4.8
prediction of α-helical state per residue (HR and NHR[c])							
no. of residues	2239	5255	4014	675	9269	916	20458
no. of observed HR	659	3988	2741	503	6729	634	4951
α-helicity (%)	29.4	75.9	68.3	74.5	72.6	69.2	24.2
PRE_R (%)	88	90	84	92	88	90	90
REC_R (%)	88	96	92	80	94	79	92
ACC	0.93	0.89	0.83	0.80	0.86	0.86	0.96
MCC	0.82	0.69	0.59	0.55	0.64	0.70	0.88

[a]Including bicelles, high-density lipoprotein nanodiscs, and lipid vesicles. [b]Number of peptides with zero, one, two, three, and four α-helices (shown in parentheses). All short (≤5 residues) unstable ($\Delta G^{\alpha}_{stab} > 0$) α-helices were excluded from this analysis for sets 2a, 2b, 2c, and 3. [c]HR, residue in helical state; NHR, residue in nonhelical state. NHR corresponds to a coil for sets 1−3 and to extramembrane domains of bitopic proteins (set 11).

FMAP 2.0 Web Server. Public access to our method is provided through the FMAP 2.0 web server (Figure S1) located at the Membranome database Web site (https://membranome.org/fmap). The server predicts the formation of individually stable α-helices by peptides and proteins under different experimental conditions and produces their all-atom 3D models oriented in membrane-like environments.

The FMAP input includes an amino acid sequence of a protein or a peptide of interest (with free or modified termini), experimental conditions (pH and temperature, K), and a choice of one of five modeling options: (1) peptides in water; (2) peptides in micelle; (3) peptides in membrane; (4) TM protein; (5) a water-soluble protein (molten globule). The all-atom approximation is employed for modeling of peptides in micelles and lipid bilayers (options 2 and 3), but the whole-residue approximation was implemented for fast identification of TM α-helices in bitopic membrane proteins (option 4) and for long peptides in micelles. The "peptide in water" and the "molten globule" options are the same as in Framework.[16] The server allows selection from four types of micelles with predefined diameters, SDS (37 Å), DPC (39 Å), DHPC (26 Å), and LPS (50 Å), or submission of a user-specified diameter for micelles. It also allows a choice between four types of artificial lipid bilayers (DOPC; DMPC; DLPC and DEuPC) and seven types of biological membranes, such as eukaryotic plasma membrane (PM), endoplasmic reticulum (ER) membrane, inner membrane (IM) of Gram-negative bacteria, PM of archaeabacteria, PM of Gram-positive bacteria, mitochondrial IM, and thylakoid membrane. The server uses previously estimated hydrophobic thicknesses (D_0) for these membranes[21] and the membrane deformation parameters determined in this work. The server also has an option with user-predefined α-helical segments to be modeled, analyzed, and oriented in membranes or micelles.

The output consists of a list of stable α-helical segments (only TM α-helices for option 4), the stabilities of α-helices relative to a coil in water and to a membrane-bound coil (ΔG^{α}_{fold} and ΔG^{α}_{stab}, respectively, kcal/mol) in the specified environment (water, micelles, or lipid bilayers) under the defined experimental conditions (temperature and pH), the binding energy of a peptide (ΔG^{pep}_{bind}, kcal/mol) to membranes or micelles, and the helix tilt angle (τ, degrees) relative to the normal of a lipid bilayer. The server also generates downloadable coordinate files (in PDB format) of all-atom 3D models of α-helices positioned with respect to the lipid bilayer or micelle and provides visualization of individual or multiple α-helices using iCn3D or GLmol. The computational efficiency of the server is relatively high: modeling of a 20-residue peptide requires about 20 min on a single CPU.

Data and Software Availability. Experimental data sets used for method development and testing are available in the Supporting Information. The software is incorporated into the public web server at https://membranome.org/fmap. The source code of FMAP 2.0 for using it as a standalone program in batch mode and all libraries are also available for download as defined in the instructions for the web server.

RESULTS AND DISCUSSION

Database of Experimentally Studied Peptides. To thoroughly test the method, we created a database that collected various published structural data for peptides in water, micelles, bicelles, and lipid bilayers, including locations of α-helical segments in amino acid sequences, spatial arrangement of peptides in lipid bilayers, and their binding energies to PC bilayers and micelles. The database contains 14 tables containing experimental and FMAP-generated data and references for 723 linear peptides (1107 data points) studied by solution and solid-state NMR, oriented circular dichroism (OCD), attenuated total reflection-Fourier transform infrared (ATR-FTIR) spectroscopy, fluorescence, and X-ray scattering (see the Microsoft Excel file in the Supporting Information). However, the following cases were excluded: (a) peptides

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

studied in the presence of trifluoroethanol or other organic solvents; (b) peptides forming oligomers of aggregates; (c) peptides forming α-helices upon binding to proteins, (d) cyclic peptides with disulfides or other covalent bonds; (e) peptides with metal clusters. The database also included a set of 170 TM bitopic proteins with structures determined by X-ray crystallography.

A significant fraction (55%) of the peptides in our database were unstructured in the aqueous solution. However, only around 4% of peptides did not form any α-helices upon binding to micelles. Nevertheless, a significant part (up to 30%) of the amino acid residues in micelle-bound peptides remain unfolded or form nonhelical structures. Therefore, we evaluated the performance of the method at the residue level by calculating the accuracy of α-helical state prediction (ACC) and the Matthews correlation coefficient (MCC) after identifying residues as helical (HR) or nonhelical (NHR) within every studied peptide.

The results of the calculations depend on the temperature and pH. The pH value affects the intrinsic stability of the α-helices through electrostatic interactions of ionizable residues with helix macrodipole (included in the third, propensity term in eq 7) and also through the transfer energy of ionizable groups from water to the micelle or membrane (last term in eq 12). The temperature is included in eq 7 and significantly affects the α-helix stability.

Peptides in Water. The FMAP 2.0 method and web server provides several options for modeling linear α-helical peptides in different environments. The "peptide in water" option was developed previously,[16] but retested here using the same parameters for a larger set of peptides studied by NMR spectroscopy in aqueous solutions at different temperatures and pH (Table S_water1). We found that FMAP 2.0 correctly predicted the presence or absence of stable α-helices in water in 54 of 57 (95%) and 63 of 65 (97%) cases, respectively, falsely predicted 2 α-helices, and did not predict 3 α-helices (Table S2). For the 54 correctly identified α-helices, the average error in prediction of helix ends was 2.1 residues per helix (Table 1). This helix prediction accuracy was similar to that in our previous assessment for the smaller set of peptides.[16] At the residue level, ACC was 0.93, MCC was 0.82, the precision of helical residue prediction (PRE_R) was 88%, and the recall of helical residue prediction (REC_R) was 88%.

Peptides in Micelles. The computational method for modeling peptides in micelles in this work is different from our previous "whole-residue" approach that was tested only for 36 peptides.[16] To date, much more experimental data are available. These data were separated into 4 sets based on data quality and micelle type. Set 2a (Table S_micelles2a) included 255 peptides with unequivocally defined α-helices that were studied by solution NMR in the presence of SDS or DPC micelles. Set 2b (Table S_micelles2b) included 152 peptides studied in various micelles with a more ambiguous interpretation of NMR data. Set 2c (Table S_micelles2c) contained 10 long peptides (with more than 50 residues) studied in micelles, most of which have long-range NOEs, i.e., some tertiary structure. Set 2d included 31 peptides studied by solution NMR in the presence of LPS aggregates (Table S-LPS 2d).

Having a more advanced method and much more experimental data, we decided to refine a few adjustable parameters of our previous model[16] using set 2a. The refined parameters were used in the calculations for all peptides in micelles and membranes. These parameters were as follows: (1) the enthalpy of a main-chain H bond in the helix (ΔH_{bb}); (2) the conformational entropy per residue due to fixing the main-chain torsion angles in an α-helix (ΔS_{bb}); (3) the helix detectability cutoff (P_d), i.e., the minimal α-helix turn occupancy that can be usually detected in solution NMR studies; (4) the binding energy to a micelle of the reference Ala residue in a coil ($\Delta G^{coil}_{bind,ref}$). To describe the mechanical deformation of a micelle due to peptide insertion, we introduced an additional parameter, $C_{s,surf}$ (eq 11). These parameters were determined by minimizing the deviations of the calculated and experimental helix boundaries by grid scan with a gradually decreasing step for peptides from set 2a.

The refined values of parameters ΔH_{bb} and ΔS_{bb} describe formation of an α-helix in water. They were found to be close to those previously determined for smaller sets of peptides in water[16] and bitopic proteins.[17] The value of ΔH_{bb} (−1.30 kcal/mol) was in between the enthalpy of the helix−coil transition determined by calorimetry (around −1 kcal/mol)[45] and the energy of the H bonds buried in the protein interior as follows from analysis of protein-engineering data (around −1.5 kcal/mol).[22] The obtained helix detectability cutoff (P_d = 0.2) indicates that α-helices with predicted occupancy greater than 20% can usually be detected based on the presence of the corresponding medium-range NOEs and other data. The optimal value of $C_{s,surf}$ for micelles was found to be 0.003 kcal/mol Å^2. The refined values of the adjustable parameters (Table S3) were used in all subsequent calculations.

The performance of FMAP 2.0 for α-helix prediction was assessed using different subsets of peptides in micelles: 2a, 2b, 2c, and a combined 2a−2b set (Table 1). The performance was better for set 2a containing peptides characterized by high-quality NMR data and worse for set 2b with more ambiguous NMR data and for set 2c containing long peptides with tertiary interactions. The combined 2a−2b set included 407 peptides, where 462 of 463 of the experimentally detected α-helices were identified in the calculations, even though 17 of them were merged to a single helix, while 17 long α-helices were predicted to be broken into two shorter α-helical fragments (Table S2). The average error in determination of the helix ends was 2.6 residues per helix. False positive predictions of α-helices occurred for 2 of 15 nonhelical peptides and for 15 of 393 helical peptides. At the residue level, the performance of FMAP 2.0 in the prediction of the α-helical residue state was satisfactory with ACC = 0.86, MCC= 0.64, PRE_R = 88%, and REC_R = 94%. Noteworthy, the ends of the α-helices are commonly determined in solution NMR studies with an error of a few residues per helix that can be close to the error in the calculations. Moreover, it is sometimes difficult to experimentally distinguish one continuous kinked helix from two adjacent helices.

In summary, these results demonstrate that FMAP 2.0 can identify α-helices that are stable in micelles at a specified temperature and pH. Furthermore, the locations of the helix ends in the amino acid sequences of these peptides are calculated with an average precision of around 3 residues per helix.

Peptides in LPS Complexes. To evaluate the method's performance for peptides in large lipid aggregates, we tested FMAP 2.0 for peptides studied in the presence of LPS complexes (set 2d). LPS aggregates are much larger than typical micelles and have an uncertain shape and a negatively

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

charged surface.[46,47] Calculations were conducted assuming that an LPS complex can be approximated by a spherical micelle with a diameter of 50 Å. For set 2d, all 27 helices were correctly identified with a helix end prediction error of 2.9 residues per helix. However, it also falsely predicted short low-stability α-helical segments for 8 of 10 nonhelical peptides, where compact nonhelical structures stabilized by tertiary interactions were evident from the presence of long-range-transferred NOEs. Thus, a more advanced computational model should be developed to account for the formation of tertiary structures by peptides in LPS complexes.

Peptides in Lipid Bilayers and Bicelles. The option for modeling peptides in membranes was developed in this work for the first time. Therefore, we thoroughly tested this option using several sets of data obtained by different experimental methods for 202 natural and synthetic peptides.

The first step was testing the ability of FMAP 2.0 to predict the locations of the α-helical segments in the amino acid sequences of membrane-bound peptides. Here, we collected data for 27 peptides in bicelles, 3 peptides in lipid vesicles, and 1 peptide in lipid nanodiscs, where the locations of the helices were determined in solution NMR studies (set 3, Table S_peptides_membranes). In bicelles, peptides can associate with either the planar central region or the curved micelle-like rim, whichever is energetically preferred. Therefore, we conducted calculations for each peptide in two systems: a bilayer formed by lipids with longer fatty acyl chains (such as DMPC) and a micelle formed by detergents or lipids with shorter fatty acyl chains (such as DHPC). Then, for every calculated α-helix, we selected the type of environment that provided the lowest helix energy ($\Delta G^{\alpha}_{\mathrm{stab}}$), which allows assigning α-helix localization to the central or the rim region.

The results are shown in Tables 1, S2, and S_peptides_membranes. FMAP 2.0 correctly predicted 3 nonhelical peptides as nonhelical and 31 single-helical peptides with a helix end prediction error of 4.0 residues per helix. However, it falsely predicted an additional α-helix in 3 single-helical peptides and suggested that the long kinked α-helix of the insulinotropic hormone was broken into two shorter fragments on the bicelle surface. The performance measures at residue level were as follows: ACC = 0.86, MCC = 0.70, $\mathrm{PRE_R}$ = 90%, and $\mathrm{REC_R}$ = 79%.

Furthermore, our calculations indicated that all predicted TM α-helices of these peptides were localized in the bilayer-like central region of bicelles, consistent with NMR studies of such peptides,[48] whereas amphiphilic helices were usually bound to the surface of the rim region. In this regard, S4 peptide from a potassium channel represents an additional interesting example. According to a solid-state NMR study, S4 peptide adopts a TM orientation in the DMPC/6-O-PC bicelles causing a ∼9 Å local thinning of the DMPC region.[49] However, calculations of S4 peptide suggested its TM arrangement only in the DLPC bilayer, which is ∼4 Å thinner than that of the DMPC bilayer. At the same time, S4 peptide had a lower calculated energy in micelles than that in the lipid bilayer, indicating a preferred localization in the rim region. Hence, we assumed that the amphiphilic α-helix of S4 peptide could form a TM α-helix near the rim region, where the effective DMPC thickness is smaller.

These results demonstrate that FMAP 2.0 can correctly identify α-helices in amino acid sequences of peptides in membrane systems (bicelles, nanodiscs, and vesicles) with a helix end prediction error of around 4 residues per helix, and it can also identify the preferential α-helix location in the central or the rim region of bicelles.

Estimation of Micelle and Membrane Binding Energies. The ability of FMAP 2.0 to reproduce the membrane binding energies of peptides was assessed using data for 12 peptides studied in various micelles (set 4, Table S_micelle_binding) and 62 peptides studied in liposomes (set 5, Table S_liposome_binding). Set 5 included only data for peptides that bind to the surface of uncharged vesicles formed by zwitterionic PC lipids. The binding energy of cationic peptides to anionic membranes has an additional electrostatic component that was not included in FMAP 2.0. Experimental binding energies were calculated from published molar partition coefficients of peptides between aqueous and lipid phases.[50] Data for the peptides with probable aggregation in water were not included.

FMAP 2.0 reproduced the experimental binding energies with root-mean-square deviations (rmsd) of 1.45 and 2.3 kcal/mol for peptides in micelles and PC bilayers, respectively, using a deformation parameter $C_{\mathrm{s,surf}}$ equal to 0.003 kcal/mol, as it was defined for micelles. However, for peptides bound to liposomes, the consistency of the calculated and observed binding energies was improved (rmsd = 1.53 kcal/mol) using the increased value of $C_{\mathrm{s,surf}}$ to 0.005 kcal/mol Å^2 (Figure 2 C). The larger optimal value of the deformation parameter $C_{\mathrm{s,surf}}$ in planar bilayers seems to be realistic, as it may reflect the stronger disturbances in the lipid headgroup region caused by insertion of surface α-helices as compared to deformations of micelles.

While the agreement was satisfactory for surface-bound peptides, the binding energies of TM α-helical states were typically lower than those in the experimental studies: −8.3 versus −6.7 kcal/mol for TMX-3 peptide, −13.9 versus −9.0 kcal/mol for pHLIP peptide, and −15.6 versus −8.2 kcal/mol, respectively, for $(\mathrm{AALALAA})_3$ peptide.[51−53] Such discrepancies may be explained by deficiencies of our method or by experimental challenges in studying highly hydrophobic peptides that are prone to aggregation.[54]

Arrangement of α-Helices in the Lipid Bilayer (TM versus non-TM States). At the next step of verification, we compared FMAP 2.0 predictions of TM and non-TM peptide arrangements in lipid bilayers with published experimental data. The test set 6 included synthetic pH-triggered membrane peptides with ionizable residues within hydrophobic α-helices studied by solid-state NMR,[55−64] ATR-FTIR spectroscopy,[65] and OCD[51,66−68] at different pH values (50 data points for 32 peptides). These peptides were designed to examine the pH-dependent equilibrium between membrane-spanning TM α-helices and surface-bound non-TM states in model PC bilayers.[69] The observed pH-dependent TM/non-TM interconversions ~~c~~ for all of these peptides ~~are found in~~ Table S_TM_surface_pH.

These experimental data were reproduced by implementing a smaller value of the membrane deformation parameter for TM α-helices ($C_{\mathrm{s,TM}}$ = 0.001 kcal/mol Å^2) than for peripheral helices ($C_{\mathrm{s,surf}}$ = 0.005 kcal/mol Å^2). The decreased value of $C_{\mathrm{s,TM}}$ relative to $C_{\mathrm{s,surf}}$ may reflect the smaller cost of bilayer deformation by a TM α-helix, which is roughly parallel to lipid acyl chains, compared to the asymmetric deformation of one leaflet by insertion of a surface helix. These values of $C_{\mathrm{s,TM}}$ and $C_{\mathrm{s,surf}}$ were used during further testing of the method for peptides in bicelles (see above) and bilayers (below).

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

The next two test sets included Lys-flanked peptides of various lengths and hydrophobicity studied in PC bilayers of different widths. Set 7 (23 data points for 7 peptides) contained a peptide series studied by solid-state NMR in DLPC (di12:0), DMoPC (di14:1Δ9c), DOPC (di18:1Δ9c), and DEuPC (di22:1Δ13c) bilayers.[70] In agreement with NMR data,[70] FMAP 2.0 predicted a non-TM state for the short peptide hΦ10 ($K_2(LA)_5K_2$) in all bilayers. For the longer hΦ16 peptide ($K_2(LA)_8K_2$), a TM state was calculated in DLPC, DMoPC, and DOPC bilayers but a non-TM state was suggested in the DEuPC bilayer, which agreed with experiments.[70] Besides, FMAP 2.0 predicted a TM state in various bilayers for longer peptides $K_3A_{18}K_3$, hΦ20 ($K_2(LA)_{10}K_2$), and hΦ25 ($K_2A(LA)_{12}K_2$), which was also consistent with experimental observations.[70,71] Some discrepancies appeared only for the longest peptide hΦ30 ($K_2(LA)_{15}K_2$), where calculations indicated formation of a tilted TM α-helix in DMoPC, DOPC, and DEuPC bilayers but suggested splitting the 30-residue segment into two short 15-residue TM α-helices in the DLPC bilayer. However, in experiments, a TM state was observed only for a minor fraction (25−30%) of the hΦ30 peptide in all bilayers, except the DEuPC bilayer, where a TM state was observed for the major fraction of hΦ30 (∼70%)[70] (see Table S_TM_nonTM_ssNMR for details).

We also compared our predictions with experimental studies of a natural peptide, a membrane-permeabilizing peptide melittin. According to our calculations, melittin forms a TM α-helix in DLPC, DMPC, and DOPC bilayers. This is consistent with solid-state NMR studies of melittin in DLPC, DMPC, and DPPC bilayers, where a kinked α-helix in TM orientation was detected.[72,73] However, X-ray scattering studies demonstrate a surface-bound state of melittin in the DOPC bilayer.[74] The authors of this study suggest that melittin has an interfacial location in the monomeric state but adopts a TM state and self-associates at higher peptide concentrations,[74] thus creating large barrel stave pores.[75]

Set 8 (55 data points for 26 peptides) included peptides studied by Trp fluorescence and fluorescence quenching.[71,76−78] The membrane penetration depth of Trp located in the middle of a peptide sequence was evaluated based on the Trp fluorescence maximum (λ_{max}) and the fluorescence quenching ratio (Q ratio) by hydrophobic and hydrophilic quenchers. For example, peptides with Q ratio < 1.5 were assigned to a TM state or a mixture of TM and non-TM states, whereas peptides with Q ratio >2.5 were assigned to non-TM states.[71] Using the Q ratio of 1.6 and λ_{max} of 340 nm as cutoff values to distinguish the TM and non-TM arrangements, we observed a good agreement between experimental data and our calculations for these peptides (see Table S_TM_noTM_fluorescence).

Hence, FMAP can properly assign the overall TM or non-TM arrangement of α-helices in PC bilayers for the majority of α-helical peptides (in more than 95% of cases). A few discrepancies can be explained by the insufficiently precise evaluation of energy by our method.

Evaluation of the Tilt Angles of α-Helical Peptides in Membranes. The ability of FMAP 2.0 to correctly evaluate the tilt angles of α-helices inserted into lipid bilayers was assessed using two sets of synthetic peptides studied by solid-state NMR in model PC bilayers of different widths. Set 9 had 98 data points for 40 synthetic peptides studied by Koeppe and co-workers (Table S_Tilt1). It was used for method testing and parametrization to optimize the value of two parameters (f_{mism} and f_{tilt}) that characterize the membrane deformation penalty due to the helix mismatch and tilting in the lipid bilayer, respectively (eq 11). Set 10 included 26 additional data points for 14 natural and synthetic peptides (Table S_Tilt2). It was used for method testing using the obtained values of two membrane deformation parameters (f_{mism} = 0.02 kcal/mol Å^2 and f_{tilt} = 0.005 kcal/mol Å^2).

The correlation coefficient between the calculated and the experimental values of the helix tilt angles for both sets combined (R^2 of 0.84) demonstrates the reasonable performance of FMAP 2.0 in the prediction of helix tilt angles (Figure 2 D). The rmsd values for helix tilt prediction were 6.5°, 7.5°, and 6.7° for sets 9, 10, and both sets combined, respectively. It is noteworthy that all peptides from these sets were correctly predicted as TM or located at the surface.

TM α-Helices of Bitopic Proteins. We previously developed a simplified FMAP 1.0 version for fast identification of TM α-helices in bitopic proteins that employs the transbilayer energy profiles for different types of amino acid residues (i.e., the whole-residue approximation) and calculates the locations of TM α-helices in sequence using the lowest energy helix−coil partition (LEP) approach.[18] In this work, we retested FMAP 1.0 using an expanded and updated set of 170 bitopic membrane proteins taken from 72 crystal structures of protein complexes from the Protein Data Bank (PDB)[79] with resolution < 3.2 Å but excluding NMR models that were used in other data sets (set 11, Table S_TM_proteins).

Taking bitopic protein sequences from the corresponding PDB files as input, FMAP 1.0 correctly detected all 170 TM α-helices with an average error in helix end prediction of 4.8 residues per TM α-helix (Tables 1, Figure 2 B). In addition, the method identified 9 hydrophobic α-helical segments which belong to water-soluble domains (Table S2). These false-positive predictions can be filtered out by comparing FMAP predictions with annotations of the protein domains in UniProt.[80] Similar results were obtained earlier while testing FMAP 1.0 for a different set of bitopic proteins.[17]

Furthermore, we tested the performance of FMAP 2.0 for the same set of bitopic proteins but using the more rigorous and complex all-atom "peptide in membrane" model with Boltzmann averaging of the helix−coil partitions. To speed up calculations, the input sequences included only residues from TM α-helices previously predicted by the whole-residue approach with eight additional residues from each side. The calculations were initially performed for the DOPC bilayer using different values (0.2, 0.35, and 0.5) of the helix detectability cutoff (P_d). We found that using an intermediate P_d value of 0.35, FMAP 2.0 performed better in predicting the ends of TM α-helices observed in crystal structures of bitopic proteins (see Table S_TM_proteins for details). Although the average errors in the determination of helix ends were not improved (Table S4), the all-atom approach decreased the number of falsely predicted TM α-helices in the set (from 9 to 4) and allowed one to optimize the geometry and side-chain rotamers of TM α-helices.

The average rmsd between the FMAP-generated models and the X-ray structures of bitopic protein complexes was found to be 1.6 Å for common Cα atoms. Hence, FMAP 2.0 correctly reproduced the TM α-helix geometry observed in protein complexes. However, only 58% of the side-chain conformations (as defined by torsion angle χ^1) were identical in the models and the corresponding crystal structures. The percentage of identical χ^1 conformers was ∼70−80% for

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

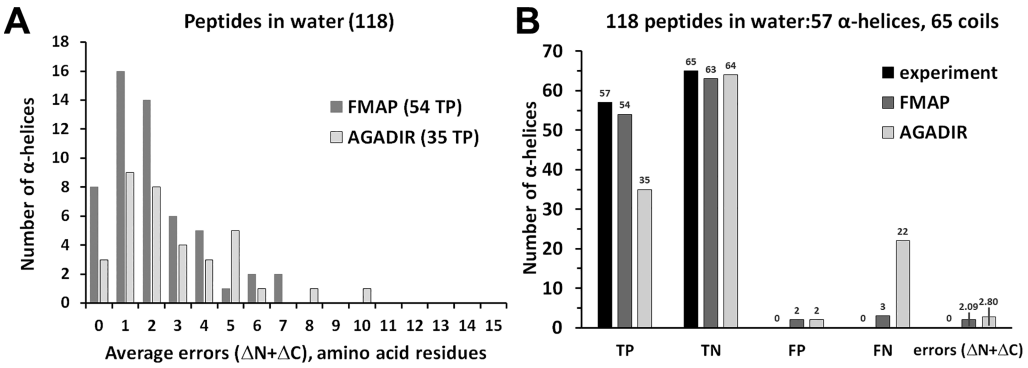


Figure 3. Comparison of the performance of FMAP 2.0 and AGADIR in the prediction of α-helices in peptides in water. (A) Distribution of helix end prediction errors for both termini (ΔN + ΔC) in calculations by FMAP (dark gray) and AGADIR (light gray). Numbers of studied peptides and correctly predicted (TP) α-helices are in parentheses. (B) Comparison of data experimentally obtained (black) and calculated by FMAP 2.0 (dark gray) and AGADIR (light gray): correctly predicted α-helices (TP), correctly predicted coils (TN), falsely predicted α-helices (FP), and missing α-helices (FN). Data set includes 118 peptides: 65 nonhelical peptides, 49 peptides with 1 α-helix, and 4 peptides with 2 α-helices. α-Helices predicted by AGADIR were defined as continuous segments of a peptide chain with the α-helicity of each residue larger than 10% cutoff. Error bars represent standard deviations for sample sizes of 54 helices (FMAP 2.0: errors = 2.09 ± 1.75 residues/helix) and 35 helices (AGADIR: errors = 2.80 ± 2.27 residues/helix).

sterically constrained β-branched side chains (Ile, Val, and Thr) but close to 40−50% for other residues (Table S5). Moreover, we observed only a poor correlation of calculated and experimental tilt angles of TM α-helices for this data set (the average deviation was around 11°). Such discrepancies presumably appeared because our calculations were performed for individual TM α-helices in the fluid lipid environment, while the corresponding X-ray structures represented protein complexes with tightly packed TM α-helices. According to our calculations, long side chains of single α-helices may adopt a number of isoenergetic conformations, whereas helix tilt angles may fluctuate by up to 10° within 1 kcal/mol around the global energy minimum. In crystallized protein complexes, the close packing of TM α-helices represents a major factor that defines the helix tilt angles and side-chain conformers. Therefore, FMAP-calculated tilt angles can be properly compared only with experimental values determined for isolated TM α-helices (as in Figure 2D).

We also investigated whether adjustment of the membrane deformation parameters (f_{mism}, f_{tilt}, $C_{s,surf}$, $C_{s,TM}$) could improve the accuracy of prediction of TM helices in bitopic proteins associated with different types of biological membranes. We found that the mechanical parameters of the DOPC bilayer perform well for predicting the ends of TM helices in eukaryotic PM, ER, and Golgi membranes. However, these parameters were reduced to improve the prediction of TM α-helices in prokaryotic cell membranes (Gram-negative and Gram-positive Bacteria and Archaea) and mitochondrial and thylakoid membranes (Table S6).

Performance of the FMAP 2.0 Server As Compared to Other Web Tools. The FMAP method and web tool are difficult to compare with other in silico methods available online because such methods were developed for a different purpose, i.e., predicting and modeling the unique structures of peptides and small proteins in aqueous solution rather than exploring the structural polymorphism of peptides in micelles or lipid bilayers under different experimental conditions. A direct comparison can be made only with AGADIR, another web server that implements a thermodynamics-based method to assess the α-helicity of water-soluble peptides depending on the pH, temperature, and ionic strength.[15] Since the current AGADIR version is applicable only to peptides in water, we compared the performance of both web servers for 118 peptides studied by NMR in aqueous solutions. We found that AGADIR correctly predicted only 61% of the experimentally observed α-helices, as compared to 95% of the α-helices identified by FMAP 2.0 (Figure 3). The average error in helix end determination by FMAP 2.0 was better than that calculated by AGADIR: 2.1 versus 2.8 residue per helix, respectively.

On the basis of our results (Figure 2 B, Table S_TM_proteins), FMAP 2.0 can be used for identifying hydrophobic TM α-helices in amino acid sequences of bitopic membrane proteins. Our previous validation indicated that FMAP 1.0 performed similarly to Phobius and slightly better than TMHMM and TopPred for a set of more than 4000 single-helical membrane proteins.[17] Nevertheless, the FMAP method was not developed for multihelical membrane proteins where some TM α-helices may not be stable in isolation but are stabilized by interactions with neighboring helices.

CONCLUSIONS

The α-helix is the most common type of structure found in membrane-bound peptides and proteins. An adequate theoretical description of the helix−coil transition in polar and nonpolar environments is essential for understanding the folding of membrane proteins and for the analysis and design of membrane-active α-helical peptides with desired biological activity. Despite the progress in the development of web tools for peptides, none of them can provide a fast and reliable assessment of the highly flexible and marginally stable structures of linear membrane-associated peptides, where small changes in the amino acid sequences, polarity of the environment, or experimental conditions may dramatically change the structure of a peptide.[9]

Here, we developed FMAP 2.0, a unique method to explore the structural plasticity and energetics of α-helical peptides in various experimental conditions and different environments, including membranes and micelles. This is a physics-based approach that uses previously developed energy terms and empirical parameters. Importantly, FMAP 2.0 not only identifies stable α-helices in the amino acid sequence but also evaluates their membrane-binding energy and generates

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

all-atom 3D models of α-helical fragments arranged in a membrane-like milieu.

The current FMAP version has a number of limitations. In general, it should be used only for linear peptides that do not form a tertiary structure and do not undergo aggregation. The method does not account for structure-stabilizing covalent bonds or metal-binding clusters, specific tertiary interactions, or formation of β-hairpins or other nonhelical structures. Moreover, FMAP 2.0 does not account for the influence of the lipid composition. It accounts only for the differences in the hydrophobic thickness and empirical deformation parameters for PC bilayers and several types of biological membranes (Table S6). The polarity profiles for different types of biological membranes[21] may be included into the future version of the method.

Despite its limitations, FMAP was useful for modeling of TM α-helices in more than 6000 bitopic membrane proteins collected in the Membranome database.[17] It was also included in software for modeling TM α-helical dimers.[19] We assume it will also be helpful for modeling and analysis of α-helical peptides in micelles and lipid bilayers. The provided web tool will make it easier for researchers to explore the structures, spatial orientations, and membrane-binding affinities of α-helical peptides in lipid membranes in different conditions, which is necessary for understanding the mechanisms of the biological activity of antimicrobial, cell-penetrating, fusion, and other membrane-associated peptides.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00161.

FMAP calculations and illustrations of the technical details (PDF)

Database containing 14 tables collecting experimental and FMAP-calculated data and corresponding references (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Andrei L. Lomize − *Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann Arbor, Michigan 48109-1065, United States*; orcid.org/0000-0002-3044-7597; Phone: +1(734) 615-7194; Email: almz@umich.edu

Authors

Kevin A. Schnitzer − *Department of Electrical Engineering and Computer Science, College of Engineering, University of Michigan, Ann Arbor, Michigan 48109-2102, United States*

Spencer C. Todd − *Department of Electrical Engineering and Computer Science, College of Engineering, University of Michigan, Ann Arbor, Michigan 48109-2102, United States*

Irina D. Pogozheva − *Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann Arbor, Michigan 48109-1065, United States*; orcid.org/0000-0002-3024-9574

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c00161

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Division of Biological Infrastructure of the National Science Foundation [Award nos. 1458002, 1855425, and 2010851 to A.L. and I.P.].

■ ABBREVIATIONS

3D, three-dimensional; ASA, accessible surface area; ATR-FTIR, attenuated total reflection-Fourier transform infrared (spectroscopy); AWS, Amazon Web Services; DEuPC, 1,2-dierucoyl-*sn*-glycero-3-ghosphocholine; DHPC, 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine; DLPC, 1,2-dilauroyl-*sn*-glycero-3-phosphocholine; DMPC, 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine; DMoPC, 1,2-dimyristoleoyl-*sn*-glycero-3-phosphocholine; DOPC, 1,2-dioleoyl-*sn*-glycero-3-phosphocholine; DPS, *n*-dodecylphosphocholine; ER, endoplasmic reticulum; FMAP, Folding of Membrane-Associated Peptides (method); HR, helical residue; IM, inner membrane; LPS, lipopolysaccharide; MD, molecular dynamics; NOE, nuclear Overhauser effect; NHR, nonhelical residue; OCD, oriented circular dichroism; PDB, Protein Data Bank; PC, phosphatidylcholine; POPC, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine; PM, plasma membrane; PPM, positioning of proteins in membranes (method); rmsd, root-mean-square deviation; SDS, sodium dodecyl sulfate; TM, transmembrane

■ REFERENCES

(1) Avci, F. G.; Akbulut, B. S.; Ozkirimli, E. Membrane active peptides and their biophysical characterization. *Biomolecules* **2018**, *8*, 77.

(2) Haney, E. F.; Straus, S. K.; Hancock, R. E. W. Reassessing the Host Defense Peptide Landscape. *Front. Chem.* **2019**, *7*, 43.

(3) Kauffman, W. B.; Fuselier, T.; He, J.; Wimley, W. C. Mechanism Matters: A Taxonomy of Cell Penetrating Peptides. *Trends Biochem. Sci.* **2015**, *40*, 749−764.

(4) Tamm, L. K.; Han, X.; Li, Y. L.; Lai, A. L. Structure and function of membrane fusion peptides. *Biopolymers* **2002**, *66*, 249−260.

(5) Wang, L.; Dong, C.; Li, X.; Han, W.; Su, X. Anticancer potential of bioactive peptides from animal sources (Review). *Oncol. Rep.* **2017**, *38*, 637−651.

(6) Lumangtad, L. A.; Bell, T. W. The signal peptide as a new target for drug design. *Bioorg. Med. Chem. Lett.* **2020**, *30*, 127115−127115.

(7) Pogozheva, I. D.; Lomize, A. L. Evolution and adaptation of single-pass transmembrane proteins. *Biochim. Biophys. Acta, Biomembr.* **2018**, *1860*, 364−377.

(8) Cardoso, M. H.; Oshiro, K. G. N.; Rezende, S. B.; Candido, E. S.; Franco, O. L. The structure/function relationship in antimicrobial peptides: what can we obtain from structural data? *Adv. Protein Chem. Struct. Biol.* **2018**, *112*, 359−384.

(9) Thevenet, P.; Rey, J.; Moroy, G.; Tuffery, P. De novo peptide structure prediction: an overview. *Methods Mol. Biol.* **2015**, *1268*, 1−13.

(10) Singh, S.; Singh, H.; Tuknait, A.; Chaudhary, K.; Singh, B.; Kumaran, S.; Raghava, G. P. S. PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues. *Biol. Direct* **2015**, *10*, 73.

(11) Jayaram, B.; Dhingra, P.; Mishra, A.; Kaushik, R.; Mukherjee, G.; Singh, A.; Shekhar, S. Bhageerath-H: a homology/ab initio hybrid

server for predicting tertiary structures of monomeric soluble proteins. *BMC Bioinf.* **2014**, *15*, S7.

(12) Lamiable, A.; Thevenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tuffery, P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* **2016**, *44*, W449−454.

(13) Koehler Leman, J.; Mueller, B. K.; Gray, J. J. Expanding the toolkit for membrane protein modeling in Rosetta. *Bioinformatics* **2017**, *33*, 754−756.

(14) Doig, A. J. Recent advances in helix-coil theory. *Biophys. Chem.* **2002**, *101*, 281−293.

(15) Lacroix, E.; Viguera, A. R.; Serrano, L. Elucidating the folding problem of α-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters11Edited by A. R. Fersht. *J. Mol. Biol.* **1998**, *284*, 173−191.

(16) Lomize, A. L.; Mosberg, H. I. Thermodynamic model of secondary structure for alpha-helical peptides and proteins. *Biopolymers* **1997**, *42*, 239−269.

(17) Lomize, A. L.; Lomize, M. A.; Krolicki, S. R.; Pogozheva, I. D. Membranome: a database for proteome-wide analysis of single-pass membrane proteins. *Nucleic Acids Res.* **2017**, *45*, D250−D255.

(18) Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. *J. Chem. Inf. Model.* **2011**, *51*, 930−946.

(19) Lomize, A. L.; Pogozheva, I. D. TMDOCK: an energy-based method for modeling alpha-helical dimers in membranes. *J. Mol. Biol.* **2017**, *429*, 390−398.

(20) Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. Anisotropic solvent model of the lipid bilayer. 1. Parameterization of long-range electrostatics and first solvation shell effects. *J. Chem. Inf. Model.* **2011**, *51*, 918−929.

(21) Pogozheva, I. D.; Tristram-Nagle, S.; Mosberg, H. I.; Lomize, A. L. Structural adaptations of proteins to different biological membranes. *Biochim. Biophys. Acta, Biomembr.* **2013**, *1828*, 2592−2608.

(22) Lomize, A. L.; Reibarkh, M. Y.; Pogozheva, I. D. Interatomic potentials and solvation parameters from protein engineering data for buried residues. *Protein Sci.* **2002**, *11*, 1984−2000.

(23) O'Neil, K. T.; DeGrado, W. F. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **1990**, *250*, 646−651.

(24) Blaber, M.; Zhang, X. J.; Lindstrom, J. D.; Pepiot, S. D.; Baase, W. A.; Matthews, B. W. Determination of alpha-helix propensity within the context of a folded protein. Sites 44 and 131 in bacteriophage T4 lysozyme. *J. Mol. Biol.* **1994**, *235*, 600−624.

(25) Chakrabartty, A.; Baldwin, R. L. Stability of alpha-helices. *Adv. Protein Chem.* **1995**, *46*, 141−176.

(26) Doig, A. J.; Chakrabartty, A.; Klingler, T. M.; Baldwin, R. L. Determination of free energies of N-capping in alpha-helices by modification of the Lifson-Roig helix-coil therapy to include N- and C-capping. *Biochemistry* **1994**, *33*, 3396−3403.

(27) Munoz, V.; Serrano, L. Analysis of i,i+5 and i,i+8 hydrophobic interactions in a helical model peptide bearing the hydrophobic staple motif. *Biochemistry* **1995**, *34*, 15301−15306.

(28) Munoz, V.; Blanco, F. J.; Serrano, L. The hydrophobic-staple motif and a role for loop-residues in alpha-helix stability and protein folding. *Nat. Struct. Mol. Biol.* **1995**, *2*, 380−385.

(29) Stapley, B. J.; Rohl, C. A.; Doig, A. J. Addition of side chain interactions to modified Lifson-Roig helix-coil theory: application to energetics of phenylalanine-methionine interactions. *Protein Sci.* **1995**, *4*, 2383−2391.

(30) Shalongo, W.; Stellwagen, E. Incorporation of pairwise interactions into the Lifson-Roig model for helix prediction. *Protein Sci.* **1995**, *4*, 1161−1166.

(31) Wimley, W. C.; White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Mol. Biol.* **1996**, *3*, 842−848.

(32) Hristova, K.; White, S. H. An experiment-based algorithm for predicting the partitioning of unfolded peptides into phosphatidylcholine bilayer interfaces. *Biochemistry* **2005**, *44*, 12614−12619.

(33) Imamura, T.; Konishi, K. Interaction of Tryptophan Dipeptides with Sodium Dodecyl Sulfate Micelles. *J. Colloid Interface Sci.* **1998**, *198*, 300−307.

(34) Lee, A. G. Lipid-protein interactions in biological membranes: a structural perspective. *Biochim. Biophys. Acta, Biomembr.* **2003**, *1612*, 1−40.

(35) May, S.; Kozlovsky, Y.; Ben-Shaul, A.; Kozlov, M. M. Tilt modulus of a lipid monolayer. *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2004**, *14*, 299−308.

(36) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **2012**, *40*, D370−D376.

(37) Marsh, D. Energetics of hydrophobic matching in lipid-protein interactions. *Biophys. J.* **2008**, *94*, 3996−4013.

(38) Kučerka, N.; Nagle, J. F.; Sachs, J. N.; Feller, S. E.; Pencer, J.; Jackson, A.; Katsaras, J. Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. *Biophys. J.* **2008**, *95*, 2356−2367.

(39) Kučerka, N.; Nieh, M. P.; Katsaras, J. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochim. Biophys. Acta, Biomembr.* **2011**, *1808*, 2761−2771.

(40) Oliver, R. C.; Lipfert, J.; Fox, D. A.; Lo, R. H.; Doniach, S.; Columbus, L. Dependence of micelle size and shape on detergent alkyl chain length and head group. *PLoS One* **2013**, *8*, e62488.

(41) Lipfert, J.; Columbus, L.; Chu, V. B.; Lesley, S. A.; Doniach, S. Size and shape of detergent micelles determined by small-angle x-ray scattering. *J. Phys. Chem. B* **2007**, *111*, 12427−12438.

(42) Pambou, E.; Crewe, J.; Yaseen, M.; Padia, F. N.; Rogers, S.; Wang, D.; Xu, H.; Lu, J. R. Structural features of micelles of zwitterionic dodecyl-phosphocholine (C12PC) surfactants studied by small-angle neutron scattering. *Langmuir* **2015**, *31*, 9781−9789.

(43) Bernhofer, M.; Kloppmann, E.; Reeb, J.; Rost, B. TMSEG: Novel prediction of transmembrane helices. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 1706−1716.

(44) Feng, S.-H.; Zhang, W.-X.; Yang, J.; Yang, Y.; Shen, H.-B. Topology prediction improvement of α-helical transmembrane proteins through helix-tail modeling and multiscale deep learning fusion. *J. Mol. Biol.* **2020**, *432*, 1279−1296.

(45) Makhatadze, G. I. Thermodynamics of alpha-helix formation. *Adv. Protein Chem.* **2005**, *72*, 199−226.

(46) Santos, N. C.; Silva, A. C.; Castanho, M. A. R. B.; Martins-Silva, J.; Saldanha, C. Evaluation of lipopolysaccharide aggregation by light scattering spectroscopy. *ChemBioChem* **2003**, *4*, 96−100.

(47) Yu, L.; Tan, M.; Ho, B.; Ding, J. L.; Wohland, T. Determination of critical micelle concentrations and aggregation numbers by fluorescence correlation spectroscopy: Aggregation of a lipopolysaccharide. *Anal. Chim. Acta* **2006**, *556*, 216−225.

(48) Mineev, K. S.; Nadezhdin, K. D.; Goncharuk, S. A.; Arseniev, A. S. Characterization of small isotropic bicelles with various compositions. *Langmuir* **2016**, *32*, 6624−6637.

(49) Doherty, T.; Su, Y. C.; Hong, M. High-resolution orientation and depth of insertion of the voltage-sensing s4 helix of a potassium channel in lipid bilayers. *J. Mol. Biol.* **2010**, *401*, 642−652.

(50) He, Y.; Lazaridis, T. Activity determinants of helical antimicrobial peptides: a large-scale computational study. *PLoS One* **2013**, *8*, e66440.

(51) Ladokhin, A. S.; White, S. H. Interfacial folding and membrane insertion of a designed helical peptide. *Biochemistry* **2004**, *43*, 5782−5791.

(52) Reshetnyak, Y. K.; Andreev, O. A.; Segala, M.; Markin, V. S.; Engelman, D. M. Energetics of peptide (pHLIP) binding to and folding across a lipid bilayer membrane. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 15340−15345.

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(53) Yano, Y.; Shimai, N.; Matsuzaki, K. Design of a soluble transmembrane helix for measurements of water-membrane partitioning. *J. Phys. Chem. B* **2010**, *114*, 1925−1931.

(54) Yano, Y.; Matsuzaki, K. Membrane insertion and dissociation processes of a model transmembrane helix. *Biochemistry* **2002**, *41*, 12407−12413.

(55) McKay, M. J.; Fu, R.; Greathouse, D. V.; Koeppe, R. E. Breaking the backbone: central arginine residues induce membrane exit and helix distortions within a dynamic membrane peptide. *J. Phys. Chem. B* **2019**, *123*, 8034−8047.

(56) Rajagopalan, V.; Greathouse, D. V.; Koeppe, R. E. Influence of glutamic acid residues and pH on the properties of transmembrane helices. *Biochim. Biophys. Acta, Biomembr.* **2017**, *1859*, 484−492.

(57) Martfeld, A. N.; Greathouse, D. V.; Koeppe, R. E. Ionization properties of histidine residues in the lipid bilayer membrane environment. *J. Biol. Chem.* **2016**, *291*, 19146−19156.

(58) Afrose, F.; Koeppe, R. E., II Comparing interfacial Trp, interfacial His and pH dependence for the anchoring of tilted transmembrane helical peptides. *Biomolecules* **2020**, *10*, 273.

(59) Afrose, F.; Martfeld, A. N.; Greathouse, D. V.; Koeppe, R. E. Examination of pH dependency and orientation differences of membrane spanning alpha helices carrying a single or pair of buried histidine residues. *Biochim. Biophys. Acta, Biomembr.* **2021**, *1863*, 183501.

(60) Thibado, J. K.; Martfeld, A. N.; Greathouse, D. V.; Koeppe, R. E. Influence of high ph and cholesterol on single arginine-containing transmembrane peptide helices. *Biochemistry* **2016**, *55*, 6337−6343.

(61) Lipinski, K.; McKay, M. J.; Afrose, F.; Martfeld, A. N.; Koeppe, R. E., II; Greathouse, D. V. Influence of lipid saturation, hydrophobic length and cholesterol on double-arginine-containing helical peptides in bilayer membranes. *ChemBioChem* **2019**, *20*, 2784−2792.

(62) Gleason, N. J.; Vostrikov, V. V.; Greathouse, D. V.; Koeppe, R. E., 2nd Buried lysine, but not arginine, titrates and alters transmembrane helix tilt. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1692−1695.

(63) Bechinger, B. Towards Membrane Protein Design: pH-sensitive topology of histidine-containing polypeptides. *J. Mol. Biol.* **1996**, *263*, 768−775.

(64) Vogt, B.; Ducarme, P.; Schinzel, S.; Brasseur, R.; Bechinger, B. The Topology of lysine-containing amphipathic peptides in bilayers by circular dichroism, solid-state NMR, and molecular modeling. *Biophys. J.* **2000**, *79*, 2644−2656.

(65) Aisenbrey, C.; Goormaghtigh, E.; Ruysschaert, J.-M.; Bechinger, B. Translocation of amino acyl residues from the membrane interface to the hydrophobic core: thermodynamic model and experimental analysis using ATR-FTIR spectroscopy. *Mol. Membr. Biol.* **2006**, *23*, 363−374.

(66) Nguyen, V. P.; Alves, D. S.; Scott, H. L.; Davis, F. L.; Barrera, F. N. A novel soluble peptide with ph-responsive membrane insertion. *Biochemistry* **2015**, *54*, 6567−6575.

(67) Reshetnyak, Y. K.; Andreev, O. A.; Segala, M.; Markin, V. S.; Engelman, D. M. Energetics of peptide (pHLIP) binding to and folding across a lipid bilayer membrane. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 15340−15345.

(68) Weerakkody, D.; Moshnikova, A.; Thakur, M. S.; Moshnikova, V.; Daniels, J.; Engelman, D. M.; Andreev, O. A.; Reshetnyak, Y. K. Family of pH (low) insertion peptides for tumor targeting. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 5834−5839.

(69) London, E.; Shahidullah, K. Transmembrane vs. non-transmembrane hydrophobic helix topography in model and natural membranes. *Curr. Opin. Struct. Biol.* **2009**, *19*, 464−472.

(70) Harzer, U.; Bechinger, B. Alignment of lysine-anchored membrane peptides under conditions of hydrophobic mismatch: A CD, 15N and 31P solid-state NMR spectroscopy investigation. *Biochemistry* **2000**, *39*, 13106−13114.

(71) Krishnakumar, S. S.; London, E. Effect of sequence hydrophobicity and bilayer width upon the minimum length required for the formation of transmembrane helices in membranes. *J. Mol. Biol.* **2007**, *374*, 671−687.

(72) Toraya, S.; Nishimura, K.; Naito, A. Dynamic structure of vesicle-bound melittin in a variety of lipid chain lengths by solid-state NMR. *Biophys. J.* **2004**, *87*, 3323−3335.

(73) Naito, A.; Nagao, T.; Norisada, K.; Mizuno, T.; Tuzi, S.; Saitô, H. Conformation and dynamics of melittin bound to magnetically oriented lipid bilayers by solid-state (31)P and (13)C NMR spectroscopy. *Biophys. J.* **2000**, *78*, 2405−2417.

(74) Hristova, K.; Dempsey, C. E.; White, S. H. Structure, location, and lipid perturbations of melittin at the membrane interface. *Biophys. J.* **2001**, *80*, 801−811.

(75) Ladokhin, A. S.; Selsted, M. E.; White, S. H. Sizing membrane pores in lipid vesicles by leakage of co-encapsulated markers: pore formation by melittin. *Biophys. J.* **1997**, *72*, 1762−1766.

(76) Ren, J.; Lew, S.; Wang, J.; London, E. Control of the transmembrane orientation and interhelical interactions within membranes by hydrophobic helix length. *Biochemistry* **1999**, *38*, 5905−5912.

(77) Caputo, G. A.; London, E. Cumulative effects of amino acid substitutions and hydrophobic mismatch upon the transmembrane stability and conformation of hydrophobic α-helices. *Biochemistry* **2003**, *42*, 3275−3285.

(78) Caputo, G. A.; London, E. Position and ionization state of Asp in the core of membrane-inserted alpha helices control both the equilibrium between transmembrane and nontransmembrane helix topography and transmembrane helix positioning. *Biochemistry* **2004**, *43*, 8794−8806.

(79) Burley, S. K.; Berman, H. M.; Kleywegt, G. J.; Markley, J. L.; Nakamura, H.; Velankar, S. Protein Data Bank (PDB): The single global macromolecular structure archive. *Methods Mol. Biol.* **2017**, *1607*, 627−641.

(80) The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **2015**, *43*, D204−D212.

https://doi.org/10.1021/acs.jcim.1c00161
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX