Prediction of amphipathic helix – membrane interactions with Rosetta

Alican Gulsevin[1] and Jens Meiler[1,2*]

[1]Department of Chemistry, Center for Structural Biology, Vanderbilt University, TN 37212, USA

[2]Institute for Drug Discovery, Leipzig University Medical School, 04103 Leipzig, Germany

Abstract

Amphipathic helices have hydrophobic and hydrophilic/charged residues situated on opposite faces of the helix. They can anchor peripheral membrane proteins to the membrane, be attached to integral membrane proteins, or exist as independent peptides. Despite the widespread presence of membrane-interacting amphipathic helices, there is no computational tool within Rosetta to model their interactions with membranes. In order to address this need, we developed the AmphiScan protocol with PyRosetta, which runs a grid search to find the most favorable position of an amphipathic helix with respect to the membrane. The performance of the algorithm was tested in benchmarks with six engineered and 44 naturally occurring amphipathic helices using membrane coordinates from the OPM and PDBTM databases, OREMPRO server, and MD simulations as reference. The AmphiScan protocol predicted the coordinates of amphipathic helices within less than 3Å of the reference structures and identified membrane-embedded residues with a Matthews Correlation Constant (MCC) up to 0.61. Overall, AmphiScan stands as fast, accurate, and highly-customizable protocol that can be pipelined with other Rosetta and Python applications.

Keywords: Amphipathic helices; Rosetta; PyRosetta; molecular dynamics; membrane proteins

Introduction

Definition and importance of amphipathic helices

Amphipathic (or amphiphilic) helices are peptides that have hydrophobic and hydrophilic/charged residues situated on opposite faces of a helix. They can interact with membranes by embedding the hydrophobic residues into the hydrophobic core of the membrane, which leaves hydrophilic/charged residues on the other side interacting with the polar lipid head groups or the solvent. Membrane-interacting amphipathic helices with diverse structure and functions can be found as part of integral and peripheral membrane proteins, and as α-helical peptides with antimicrobial properties (1). The membrane selectivity profile of amphipathic helices may vary based on charge, length, and the amino acid composition of the helix (2) or lipid

33 composition of the membrane (3). Different classes of amphipathic helix motifs have been
34 proposed to fine-tune the behavior of amphipathic helices (4).

35 In integral and peripheral membrane proteins, helices of amphipathic character play a role
36 in sensing changes at the membrane surface, alteration of membrane curvature, and insertion of
37 proteins into membranes (5). As standalone peptides, positively charged amphipathic helices
38 have been shown to have antimicrobial properties, thus called antimicrobial peptides (AMPs).
39 AMPs disrupt the membrane either by altering the membrane curvature or through formation of
40 pores in the cell membranes (6,7).

41 **Methods for modeling and analyzing amphipathic helix structures**

42 Despite the widespread presence of amphipathic helices, there are very few established
43 methods tailored for modeling of amphipathic helices. 3D-HM is a server that takes as input a
44 sequence which is modeled as an ideal α-helix, or a user-provided structure file (8). It calculates
45 parameters including the magnitude and orientation of the hydrophobic moment, the angle
46 between the hydrophobic moment vector and the membrane normal, and electrostatic potentials
47 of the helices. Images showing the helix orientation and electrostatic maps are also provided. The
48 HELIQUEST server takes as input a sequence and follows a sliding window approach to calculate
49 structural parameters of an ideal helix (9). It calculates parameters including hydrophobicity,
50 hydrophobic moment magnitude, polar/nonpolar residue composition, and net charge. Images of
51 hydrophobicity plots are also provided to determine the membrane-facing side of the helix.

52 Overall, these methods give information about the structural features of a helix, but they
53 either fail to model helix – membrane interactions in the presence of a membrane or have
54 limitations in customization or integration with other methods. Further, while these tools are useful
55 to predict amphipathicity of a small number of peptides and/or their membrane coordinates, they
56 usually are not suitable for high-throughput screening of putative amphipathic sequences.

57 **Monte-Carlo-based modeling of amphipathic helix – membrane interactions**

58 Different than the tools that focus solely on structural modeling of amphipathic helices, the
59 MCPep server utilizes an implicit membrane model with a Monte Carlo approach to predict the
60 position of peptides in the membrane starting with peptide sequences (10,11). The output files
61 include the helical content of the given sequence, trajectory of example simulations, free energy
62 decomposition of the peptide residues, and details on favorable orientations with respect to the
63 membrane. An image file showing the position of the helix with respect to the membrane planes
64 is also provided. One caveat of this method is that the algorithm was not benchmarked on
65 amphipathic helices, so its accuracy with helices that bind to the membrane surface is unknown.

66 **Molecular-dynamics-based modeling of amphipathic helix – membrane interactions**

67 Molecular dynamics (MD) simulations have also been used extensively to model
68 amphipathic helix – membrane interactions (12,13). These simulations were run both to explore
69 anchoring of peripheral membrane proteins to membranes through their amphipathic helices and
70 to explore the interactions between amphipathic peptides and membranes. One advantage of the
71 MD simulations is that the dynamic nature of helix – membrane interactions are reflected in the
72 MD trajectories. Further, the presence of an explicit membrane environment allows observation
73 of specific interaction of amino acids and lipids including changes in the membrane curvature. On
74 the other hand, MD simulations can be time-consuming. Also, computational modeling of peptide
75 – membrane systems for MD simulations can be challenging for the non-experts.

76 **RosettaMembrane is a fast and accurate alternative to existing amphipathic helix modeling**
77 **methods**

78 Considering the caveats of the existing methods, we suggest an alternative approach to
79 obtain fast, accurate, and detailed information on the amphipathic helix – membrane interactions:
80 The RosettaMembrane score function (14–16) is a good candidate for evaluating these
81 interactions rapidly considering its success in modeling membrane protein structures in the past
82 (17). The RosettaMembrane score function is based on the implicit membrane model IMM1 (18)

with a default bilayer hydrophobic thickness of 30Å. The model consists of a hydrophobic core region of 18Å that represents the non-polar tail groups of lipid molecules and a mid-polar "transition" region of 6Å flanking the hydrophobic core on either side of the membrane that represents the more polar glycerol and carbonyl groups of the lipid molecules. There is no explicit representation of the polar headgroups since these are considered to have the same dielectric constant as the solvent.

Although the RosettaMembrane energy function was originally conceived to model integral membrane proteins, we hypothesized that it could be used to score depth and orientations of amphipathic helices as well. Based on this premise, we designed a grid search protocol called AmphiScan that can be customized and pipelined with other Rosetta applications. We benchmarked the performance of AmphiScan with six ideal helices consisting only of leucine and lysine residues (LK peptides) and 44 amphipathic helix structures belonging to integral and peripheral membrane proteins (Supporting Table 1) using the helix membrane coordinates from the OPM, PDBTM, OREMPRO, and MD simulations as reference. The accuracy of the best energy poses generated by Rosetta was determined by the RMSD from the reference structures and the membrane-embedded residue accuracies.

Results

AmphiScan calculations utilize three search parameters to define helix depth, tilt angle, and rotation angle

The AmphiScan protocol was first tested on LK peptides with lengths of 15 to 22 amino acids (19) to assess its success with engineered sequences. The six LK peptides LK15, LK18, LK19, LK20, LK21, and LK22 (Table 1) were modeled as ideal α-helices with PyMOL (20). The membrane was placed along the xy-plane with the z-axis as the membrane normal. The rotation angle of the helix around its screw axis was defined as α and the angle of the screw axis with respect to the xy-plane was defined as β (

Figure 1). The LK peptides were placed at a β angle of 0° to 180° for independent AmphiScan calculations and α values between 0° and 360° were scored at each β and z value. Multiple step sizes were tested for depth (movement along the z-axis), α, and β parameters. A

depth step size of 0.1Å, α angle of 5º, and β angle of 1º was selected as a smaller step-size failed to further improve the calculation results but increased the time required. These step-size parameters were used for the naturally occurring helix sequences as well.

AmphiScan uses a grid search approach to identify the membrane binding configurations of amphipathic helices

Two types of grid searches were run to assess the success of AmphiScan in accurately sampling helix rotation and tilt angles. In the first set of calculations (zα-scans), helix β angles from the reference structures were conserved and only rotation along the helix screw axis and movement along the z-axis was allowed (Figure 1A). In the second set of calculations (zαβ-scans), helix screw axis rotations and helix membrane depth changes were sampled at different β angles, and the pose that gave the best score through all the β angles was kept as the representative pose (Figure 1B) . The reason behind the use of two different kind of grid searches was that the β-angle is often assumed to be to be 0º, i.e. perpendicular to the membrane normal and parallel to the membrane plane. The zα-scans give information about whether the RosettaMembrane score function can accurately identify the ideal depth and rotation of the α-helix independent of sampling different tilt angles. The zαβ-scans go a step beyond and give information about whether Rosetta can sample the tilt angle of the helix with respect to the membrane.

AmphiScan accurately predicts the membrane depth and orientation of the LK peptides

The zα-scan calculations took 2 – 5 minutes depending on the helix size with the selected parameters on a single CPU. The difference between the best poses from the zα- and zαβ-scans were marginal. For the zαβ-scans of all six helices, the best-scoring pose was obtained at tilt angles of 2º and 3º, in accordance with the results of fluorescence spectroscopy and IRRAS experiments that showed near-parallel arrangement of the LK peptides with respect to the membrane surface. Also, the best-scoring structures obtained for all LK peptides were centered 3 – 4Å below the membrane surface within the mid-polar region (21,22). All low-scoring poses had hydrophobic leucine residues embedded in the non-polar region and the charged ammonium groups of the lysine side chains extending outside the membrane.

Heatmaps of AmphiScan calculations illustrate the broad range and symmetric distribution of the scores through the membrane

Next, heatmaps color-coded with the scores calculated at different membrane depth and helix rotation angles were plotted for the AmphiScan calculations at a helix arrangement parallel to the membrane surface (e.g. β=0°) to visualize the energy surface of the helix through the membrane (**Error! Reference source not found.**). The heatmaps illustrated that the most favorable region for the LK peptides was the mid-polar region of the membrane with additional low energy poses at the polar region. There were low energy regions within the hydrophobic core at the boundary of the mid-polar region as well, but these regions were shallow. The distribution of low-energy scores were dispersed at a depth range of ~1.5Å and a rotation angle range of ~50º.

Another expected feature of the heatmaps was the near-symmetric nature of the scores on both sides of the membrane. The regions across the membrane that correspond to a full 180º rotation around the helix screw axis of the lowest-scoring structure gave the lowest scores on the opposite side at the corresponding membrane depth. In addition, the α value corresponding to the lowest score yielded the highest scores on the other side of the membrane near the equivalent of the lowest-scoring depth. This is logical since such an arrangement would place the hydrophilic residues inside the hydrophobic core and the hydrophobic residues inside the polar region.

RosettaMembrane score function is suitable for scoring of amphipathic helices

Overall, the AmphiScan protocol was able to successfully predict the placement of the LK peptides at the mid-polar region of the membrane in a near-parallel arrangement with the membrane plane, and the resulting energy surfaces had a nearly-symmetrical distribution of the scores across the membrane. Since these results indicate that the RosettaMembrane score function can be used to accurately score amphipathic helices at regions both inside and outside the membrane, we moved on with the real-life case of predicting membrane interactions of experimental amphipathic helix structures.

166 **AmphiScan is benchmarked against membrane coordinates from OPM, PDBTM,**
167 **OREMPRO, and molecular dynamics simulations**

168 Prediction of the membrane coordinates of naturally occurring amphipathic helices with
169 experimentally determined structures is a more challenging task when compared to engineered
170 systems due to the potentially non-ideal helix configurations within the membrane caused by
171 interactions with the other domains of the protein or distorted structure of these helices when
172 using artificial membrane mimetics. There exist a myriad experimental methods including NMR,
173 EPR, and fluorescence-quenching to measure membrane immersion depth and orientation with
174 fair accuracy (23). However, systematic analyses of amphipathic helices are rather scant, and the
175 results from these experiments usually depend on environmental conditions and lipid composition
176 of the membrane and thus cannot be directly compared with our results.

177 Instead, membrane coordinates predicted by existing computational algorithms trained on
178 experimental data were used as reference. AmphiScan calculations were run with the 44
179 amphipathic helices in the benchmark set (Supporting Table 1) using the membrane coordinates
180 from the OPM (24–26) and PDBTM (27,28) databases, plus the OREMPRO server (29,30) as
181 references to prevent any biases caused by a single prediction algorithm. Because only the OPM
182 database had membrane coordinates for peripheral membrane proteins and some of the integral
183 membrane proteins in the benchmark set, membrane coordinates approximated from our 1μs MD
184 simulations were used as an additional reference point.

185 **Success of AmphiScan is measured by helix root mean square deviation and membrane-**
186 **embedded residue Matthews Correlation Coefficient values**

187 Because the depth and the identity of the membrane-embedded residues of amphipathic
188 helices play the most important role in helix – membrane interactions, these parameters were
189 used to quantify the error of the AmphiScan calculations. To measure the differences in the
190 membrane depth and orientation, root mean square deviation (RMSD) values were calculated as
191 the difference between the coordinates of the Cα of each amphipathic helix reference structure
192 from the membrane protein databases (i.e. membrane coordinates) and the best-scoring pose

193 predicted by the AmphiScan protocol. In order to avoid RMSD differences caused by differences
194 between the default Rosetta membrane thickness (30Å) and database-calculated values, the
195 membrane thicknesses calculated by the databases were used as the membrane thickness for
196 each Rosetta calculation. To measure the differences between the identity of the predicted and
197 reference membrane-embedded residue accuracies, Matthews Correlation Constants (MCC)
198 were calculated with respect to the membrane-embedded residues in the reference structures
199 (Methods, Equation 1).

200 In the following sections, the accuracy of the AmphiScan calculations with respect to the
201 reference coordinates from the OPM, PDBTM and OREMPRO algorithms, and MD simulations
202 are discussed.

203 **AmphiScan calculations with OPM reference structures demonstrate moderate root mean**
204 **square deviations and good membrane-embedded residue correlations for a diverse set of**
205 **amphipathic helices**

206 We started our analysis using membrane coordinates from the OPM database as the
207 reference for all amphipathic helices. 44 amphipathic helices belonging to 25 peripheral
208 membrane proteins and 19 integral membrane proteins were selected as the benchmark set for
209 the analyses as described in the Materials and Methods section. The benchmark set consisted of
210 amphipathic helices with diverse structural and conformational properties. The helix lengths
211 varied from 8 to 29 amino acids, the tilt angles with the membrane normal were between 0º and
212 45º, and the strength of helix – membrane interactions varied from interacting only with periphery
213 of the membrane surface to being mostly submerged into the non-polar region of the membrane.

214 For the zα-scans, the average RMSD value was 2.9Å and the average membrane-
215 embedded residue MCC was 0.51. For the zαβ-scans, the average RMSD value was 3.6Å and
216 the average membrane embedded-residue MCC was 0.44 (Supporting Table 2). The predictive
217 value of the AmphiScan calculations varied among the helices whereby *1h0a_h1, 3j5p_h1,*
218 *3jw8_h1, 4hhr_h2, 4zwn_h1, 5ek8_1, 5f19_h4, 5w7b_h1, 5w7l_h1*, and *6d26_h1* had RMSD

219 values below 1Å, but the helices *1q4g_h4, 3tij_h1, 4qnd_h1, 4ymk_h2, 4ymk_h3, 5ahv_h1,*
220 *5dqq_h1, 5w7l_h3*, and *6an7_h1* had RMSD values larger than 5Å. Examples to poses with
221 RMSD values smaller than 2Å, between 2 and 4Å, and larger than 4Å can be seen in the top,
222 middle, and bottom rows of the Figure 3 correspondingly.

223 **Good recovery of membrane-embedded residues can be obtained at RMSD values as large**
224 **as 3Å**

225 To understand the relation between the RMSD values and the membrane-embedded
226 residue accuracies, and to determine acceptable RMSD values based on the membrane-
227 embedded residue accuracies, we compared the two values (Supporting Figure 2, left panel). The
228 membrane-embedded residue accuracies were larger than 80% on average below an RMSD of
229 3Å for both zα- and zαβ-scans with MCC values of 0.50 and 0.57 correspondingly. At RMSD
230 values larger than 3Å, zα- and zαβ- scans showed different RMSD patterns such that the
231 accuracies decreased steadily at larger RMSD values for the zα- scans, but decreased only
232 slightly for the zαβ-scans. RMSD values lower than 3Å were considered acceptable based on
233 these results since the membrane-embedded residue profiles of the amphipathic helices were
234 accurately represented at RMSD values below this cutoff.

235 **Side chain repacking has no significant effect on prediction accuracy**

236 In order to assess whether side chain repacking increases the performance of our
237 algorithm, we ran zα calculations with side chain repacking at each rotation step. Helix RMSD
238 and membrane-embedded residue MCC values, and rotamer recovery rates for χ1 and χ2 angles
239 were calculated to measure the success of the algorithm under these conditions. The calculated
240 RMSD (3.0 Å) and membrane-embedded residue MCC values (0.49) were comparable to that of
241 the calculations with no side chain repacking. The rotamer recovery rates measured for the χ1
242 and χ2 angles were 68% and 41% respectively (Supporting Table 3). Examples to repacked
243 structures with χ1 rotamer recovery rates over 0.85 can be seen in Supporting Figure 1. Based
244 on lack of decrease in RMSD values or increase in MCC values, we concluded that side chain

245 repacking has no significant effect on the accuracy of the calculations and restricted our analyses
246 to calculations with no side chain repacking.

247 **AmphiScan performs better with charged helices compared to neutral helices**

248 Following the RMSD and membrane-embedded residue accuracy analyses, we looked for
249 a correlation between the RMSD values and helix properties such as charge, size, number of
250 prolines, experimentally determined tilt angle, native distance from the membrane center,
251 structure quality, and whether the helix is a terminal one to identify biases of the
252 RosettaMembrane score function. Of these parameters, helix net charge showed an inverse
253 correlation with the RMSD values, whereby neutral helices on average had the largest RMSD
254 values (Supporting Figure 2, right panel). This correlation was more evident in the zα-scans
255 compared to the zαβ-scans, possibly since fixed helix tilt angles disallow rearrangement of the
256 residues to diminish the unfavorable interactions between the helix and the membrane. It might
257 also be that less-charged amphipathic helices have broader energetic minima when interacting
258 with the membrane, thus allowing a wider range of physically realistic arrangements. Inspection
259 of the heatmaps showed no clear correlation in support of this idea, but the distribution of the
260 scores was heavily affected by the geometry of the helix, so a reliable heatmap comparison
261 between helices that have different β angles is not possible. The amphipathic helices that were
262 closer to the membrane core in the database structures had a larger ratio of high-RMSD helices
263 compared to peripheral helices, but this effect did not apply to all helices close to the membrane
264 core.

265 Overall, the average RMSD value calculated for the zα-scans was below our cutoff of 3Å,
266 and the value calculated for the zαβ-scans was 3.6Å. Next, we investigated the effect of database
267 selection on the accuracy of the AmphiScan results.

Proteins from OPM, PDBTM, and OREMPRO have similar membrane orientations but can have largely different membrane thicknesses

Since amphipathic helices have only tenuous interactions with the membrane surface in general, slight variations in the calculated helix tilt angles or membrane thicknesses predicted by computational algorithms can result in large differences in the identity of the membrane-embedded residues or the immersion depth of amphipathic helices deposited to the databases. In fact, a recent analysis comparing the OPM and PDBTM databases showed that the identity of the predicted membrane-embedded residues in OPM and PDBTM databases match only 60-80% for most proteins (31). Therefore, we used reference structures from the PDBTM database and the OREMPRO server for comparison with the OPM results.

Only 16 of the 44 helices in the benchmark set were available in the PDBTM database. The protein orientations calculated for the proteins hosting these helices were similar in OPM, PDBTM, and OREMPRO with changes in the protein angles less than 5° with respect to the membrane normal. On the other hand, the calculated hydrophobic monolayer thicknesses had differences up to 5Å, which caused drastic differences in the membrane-embedded residue profiles of the amphipathic helices (Supporting Table 4).

Selection of the reference database significantly affects the root mean square deviation values of individual helices

The functional form of the Rosetta membrane model is independent of membrane thickness, but the RMSD values in this study are calculated based on coordinate differences between the reference structures and AmphiScan poses, which are affected by the differences in membrane thicknesses. To prevent such deviations, membrane thicknesses calculated for each protein by the OPM, PDBTM, and OREMPRO algorithms were used as the hydrophobic thickness of the Rosetta membrane model for the corresponding AmphiScan calculations.

The differences in the calculated hydrophobic thicknesses failed to significantly affect the average RMSD and membrane-embedded residue values, but they significantly affected the outcome of the AmphiScan calculations at the helix level. For the zα-scans, the average RMSD

295 values calculated with respect to the OPM, OREMPRO, and PDBTM references were 3.4Å, 3.2Å,
296 and 3.2Å respectively with average embedded-residue MCC values of 0.42, 0.42, and 0.38 (Table
297 2). The helices *4ymk_h1*, *4ymk_h2*, and *6an7_h1* had consistently high RMSD values (> 4Å) and
298 *1rhz_h1*, *6igk_h1*, and *3j5p_h1* had moderate to low RMSD values (2.0Å to 4.0Å) in the zα-scans
299 regardless of the source of the membrane parameters. For the remaining 10 helices, the selected
300 source of the membrane parameters affected the RMSD values by 1.0Å to 7.4Å. Despite the large
301 variations, no single method gave consistently low RMSD values, suggesting that the prediction
302 accuracies differ on a case-by-case basis. For example, the RMSD value of *5dqq_h1* dropped
303 from 7.2Å to 1.6Å when the PDBTM membrane coordinates were used, the RMSD value of
304 *4ymk_h3* dropped from 5.0Å to 1.7Å when OREMPRO parameters were used, but the RMSD
305 value of *4umw_h1* increased to 5.1Å from 1.6Å when PDBTM parameters were used.

306 **Selection of the lowest root mean square deviation value for each helix improves**
307 **AmphiScan performance**

308 Because none of the methods used as a reference has a proven accuracy over the other
309 methods, the upper limit of AmphiScan's success was calculated by considering only the best
310 values for each helix. Specifically, the "best possible" RMSD values were calculated by using the
311 RMSD values and membrane-embedded residue accuracies from the method that gave the best
312 RMSD value for each helix. With this approach, the average RMSD value dropped by around
313 1.0Å to 2.2Å and the average membrane-embedded residue accuracy MCC value increased from
314 0.38-0.42 to 0.61.

315 The average RMSD values calculated for the zαβ-scans were slightly higher compared to
316 the zα-scans with zαβ-scan RMSD values of 3.8Å, 3.6Å, and 3.9Å calculated for OPM,
317 OREMPRO, and PDBTM respectively (Table 2). However, the membrane-embedded residue
318 accuracies were not significantly different than the zα-scans and the corresponding membrane-
319 embedded residue accuracy MCC values for the three methods were 0.37 for OPM, 0.39 for
320 OREMPRO, and 0.38 for PDBTM.

321 Similar to the zα-scans, significant RMSD differences were observed for some helices
322 despite the similar average RMSD values. *1rhz_h1*, *4ymk_h2*, *5mlz_h2*, and *6an7_h1* had similar
323 RMSD values for all three calculations, but the remaining helices showed variation depending on
324 the reference membrane coordinates compared to the zα-scans (Table 2). When only the method
325 that gave the best RMSD was considered for each helix as done for the zα-scans, the differences
326 between the two calculations also diminished. The average RMSD value dropped by 1.1Å – 1.4Å
327 to 2.5Å and the average embedded-residue accuracy MCC value increased 0.61, equal to that of
328 the zα-scans. To see whether this improvement also held true for the amphipathic helices with no
329 PDBTM entries, we compared the OPM results with the membrane coordinates calculated by MD
330 simulations.
331

332 **Root mean square deviation values for the molecular dynamics simulations are calculated**
333 **for an ensemble of helix coordinates**

334 As mentioned previously, OPM is the only database that includes membrane coordinates
335 of peripheral membrane proteins. We ran 1μs MD simulations in an explicit dioleoyl-
336 phosphatidylcholine (DOPC) membrane model for the 44 amphipathic helices to serve as a
337 second reference point for the AmphiScan calculations. Membrane thicknesses and RMSD
338 values were calculated for each 1ns snapshot over the last 800ns of the simulations, and the
339 averages calculated from these values were reported as the RMSD value of each helix. Average
340 membrane coordinates of the helices and the average membrane thicknesses calculated through
341 the MD simulations were used to calculate the membrane-embedded residue accuracy MCC
342 values.

343 **Molecular dynamics simulations show helix unfolding in the absence of backbone**
344 **restraints**

345 First, MD simulations with no restraints on the helices at the production stage were run to
346 sample helix dynamics. However, the trajectories of these calculations showed that the helices

347 were unfolding over 1μs, perhaps as part of their folding equilibrium, facilitated in some cases by
348 the interactions between the charged helix termini and the charged side-chains. Next, to address
349 this problem, restraints were applied to the backbone hydrogen bonds of the helices to conserve
350 the starting backbone structure and thus prevent unfolding during the MD simulations. In addition
351 to the backbone restraints, N- and C-termini of each helix were capped by an acetyl and
352 methylamide group respectively to prevent unwanted interactions between the residue side
353 chains and the charged terminal groups of the helices. The analysis of the φ- and ψ- angles of
354 the helix residues throughout the MD trajectories demonstrated that application of restraints and
355 terminal capping were effective in maintaining the helical structure of the peptides (data not
356 shown).

357 **Presence of amphipathic helices affects membrane thickness and curvature**

358 The restrained trajectories were used for the RMSD and membrane-embedded residue
359 accuracy MCC calculations. The RosettaMembrane model does not have an exact physical
360 representation, so there is no objective method to extrapolate the MD hydrophobic thicknesses
361 to Rosetta hydrophobic thicknesses. Therefore, the hydrophobic monolayer thickness of the
362 membrane was approximated as the distance between the lipid center of mass and the average
363 z-coordinates of the lipid carbonyl group oxygens (atom types O21 and O11) for practical
364 purposes.

365 The effect of helix binding on the membrane structure was evident in the MD trajectories.
366 The membrane monolayer thicknesses on the helix-bound side of the membrane was 1.0Å – 1.5Å
367 thinner compared to the other side (Table 3). To include these thinning effects in our calculations,
368 the hydrophobic thicknesses calculated for the helix-bound side of the membrane were used as
369 the monolayer thickness in the AmphiScans rather than half bilayer thicknesses. In addition to
370 membrane thinning, the curvature of the membrane visibly increased around the amphipathic
371 helices, consistent with the distortion analyses on the proteins hosting these helices in the
372 MemProtMD database based on coarse-grain and fully atomistic MD simulations (32).

373 **AmphiScan root mean square deviation values from molecular dynamics simulations are**
374 **larger than database references though with similar Matthews Correlation Coefficient**
375 **values**

376 Contrary to the RMSD values calculated using OPM, PDBTM, and OREMPRO membrane
377 coordinates as the reference, the average RMSD values calculated for the zα-scans (4.4Å) were
378 slightly higher than the values calculated for the zαβ-scans (4.2Å). The standard deviations of the
379 RMSD values over 800 frames were between 0.8Å and 2.0Å, depicting the mobile nature of the
380 helices on the membrane surface. Despite the large RMSD difference, the membrane-embedded
381 residue accuracy MCC values were comparable for the zα-scans (0.49) and the zαβ-scans (0.50),
382 both of which were higher compared to the calculations using membrane coordinates from the
383 databases as reference.

384 Choosing the best RMSD for each helix from the OPM and MD-simulation reference
385 structures yielded an average RMSD of 2.4Å and a membrane-embedded residue accuracy MCC
386 of 0.58 for the zα-scans, comparable to the values calculated for the 16 amphipathic helices from
387 integral membrane proteins through comparison of the OPM, PDBTM, and OREMPRO results.
388 For the zαβ-scans, the membrane-embedded residue accuracy MCC was 0.5. Interestingly,
389 selecting the best-RMSD case between the OPM and MD simulation results did not result in an
390 improvement of the MCC values, which remained at 0.5 despite the 1.3Å improvement in the
391 RMSD values (2.9Å).

392 In summary, the AmphiScan protocol was able to predict the coordinates of 44
393 amphipathic helices within less than 2.5Å with a membrane-embedded residue accuracy MCC
394 value of 0.61 in the best-case scenario.

395

396 ## Discussion

397 Many integral and peripheral membrane proteins possess amphipathic motifs that may
398 play an important role in protein – membrane interactions. However, analyses of these helices on

399 the membrane interface is not trivial. Although a variety of tools exist to model membrane proteins,
400 few methods apply to amphipathic helices that interact with the membrane surface. In this work,
401 we utilized PyRosetta to develop the AmphiScan protocol to predict the ideal depth, orientation,
402 and tilt angle of LK peptides and amphipathic helices belonging to peripheral and integral
403 membrane proteins in membranes. The results were analyzed using qualitative experimental data
404 for the LK peptides and membrane coordinates from OPM, PDBTM, OREMPRO servers and MD
405 simulations for the experimental amphipathic helix structures.

406 **AmphiScan accurately reproduces membrane coordinates of LK peptides**

407 In the scenario with the engineered LK peptides, the AmphiScan calculations correctly
408 predicted the near-parallel placement of the LK peptides with respect to the membrane surface
409 and the embedding of the helix center of masses to the mid-polar transition region of the
410 membrane. Consistent with the amphipathicity of these peptides, the leucine residues were partly
411 embedded in the non-polar region of the membrane and the lysine residues were outside the
412 membrane model in contact with the polar/solvent region. The heatmaps of the scores at different
413 helix screw axis rotation angles and membrane depths showed similar scores within 1.5Å of the
414 best-scoring membrane depth and helix rotations up to 50º. These variations may be a reflection
415 of the flexible nature of the amphipathic helix – membrane, or the resolution limit of the AmphiScan
416 method.

417 **AmphiScan's success on naturally occurring helices depend on the source of membrane**
418 **coordinates and inclusion of tilt angles**

419 For the amphipathic helices from experimental structures, our calculations with the
420 AmphiScan algorithm yielded average RMSD values of 2.9Å (zα-scan) and 3.6Å (zαβ-scan) and
421 membrane-embedded residue accuracy MCC values of 0.51 (zα-scans) and 0.44 (zαβ-scans) for
422 the 44 amphipathic helices in the benchmark set when membrane coordinates from OPM was
423 used as the only source. Use of PDBTM and OREMPRO membrane coordinates for 16
424 amphipathic helices from integral membrane proteins showed average RMSD values within 0.3Å

425 of the OPM results and similar average membrane-embedded residue accuracies. However,
426 selection of database had a significant effect on the values calculated for individual helices due
427 to the drastic differences in the hydrophobic membrane thicknesses calculated by the different
428 methods. Although membrane thickness might not significantly affect the membrane interactions
429 of a standalone amphipathic helix on the surface, the helices in our benchmark set were taken
430 from larger protein structures for which the membrane coordinates are determined by large
431 hydrophobic sections of the structure. As a result, differences in membrane thicknesses had a
432 direct effect on the calculated depth and membrane-embedded residues of the helices.

433 **Selection of the reference coordinates from the best root mean square deviation source**
434 **improves AmphiScan's performance**

435 The upper limit calculated for the accuracy of the AmphiScan protocol by only considering
436 the membrane coordinates that gave the best RMSD and membrane-embedded residue
437 accuracies for each helix was 2.2Å-2.5Å with a membrane-embedded residue accuracy MCC of
438 0.61. Part of the reason for the observed variations is that the membrane models used by OPM,
439 PDBTM, OREMPRO, and RosettaMembrane are not identical and treat the non-polar, mid-polar,
440 and polar regions of the membrane differently. While these differences may not matter for larger
441 membrane proteins, they cause large ambiguities regarding the identity of the residues at the
442 membrane surface.

443 **AmphiScan's success is limited by imperfect helix structures and assumptions intrinsic**
444 **to implicit membrane models**

445 Two main limitations of the AmphiScan protocol are the "non-ideal" structure of the
446 experimental amphipathic helices and simplistic assumptions behind the RosettaMembrane
447 model. In theory, an amphipathic helix would be situated on the membrane interface interacting
448 with both the polar and non-polar portions of the membrane, but it may behave differently as part
449 of a larger protein assembly in reality. This is observed for several helices that are buried inside
450 the hydrophobic core of the membrane despite being identified as amphipathic. Because

451 AmphiScan calculations only consider standalone helices, the ideal conformation for the
452 standalone helix may deviate from the conformation observed in a protein. Also, most
453 experimental amphipathic helix structures in the benchmark had distorted or kinked structures,
454 which may cause inaccuracies in calculation of helix screw axes, thus preventing sampling of all
455 the conformational space.

456 Other errors are associated with the simplistic nature of the RosettaMembrane model.
457 Properties such as membrane composition, membrane curvature, or disruption of the membrane
458 surface in response to amphipathic helix binding that are known to play a major role in
459 amphipathic helix – membrane interactions (33–36) cannot be fully represented by implicit
460 membrane models. The importance of these interactions was also evident in the MD simulations
461 whereby the helix-bound side of the membrane thinned compared to the other side, and the
462 membrane curvature changed in response to helix binding. Interestingly, there was an inverse
463 correlation between helix net charge and the corresponding RMSD values for the OPM-based
464 calculations, especially for the zα-scans. Diminished accuracy observed for the neutral helices
465 may indicate that the Rosetta membrane score function relies more on electrostatic terms to
466 accurately predict helix immersion depths or that neutral helix – membrane interactions are more
467 tenuous than charged helices.

468 **Limitations of molecular dynamics simulations make quantification of errors challenging**

469 Although MD simulations are not affected by the limitations mentioned for implicit
470 membrane calculations, helix – membrane interactions in MD simulations may be affected by
471 force field (37,38) or sampling biases (39) leading to helix disruption or unfolding in extreme
472 cases. This was indeed the case for the unrestrained MD simulations in this study where the
473 interactions between the lipid head groups and helix side-chains or helix termini led to unfolding
474 of most helices. However, it should be noted that the unfolding may be part of the folding
475 equilibrium of these helices. Lastly, comparison of membrane coordinates from implicit and

476 explicit membrane calculations may be inaccurate at the membrane boundary since implicit
477 models have no exact physical representation contrary to explicit lipid molecules.

478 Comparison of the results based on OPM reference structures with the results based on
479 MD simulations for 44 amphipathic helices gave larger average RMSD values for the zα-scans
480 (4.4Å) and zαβ-scans (4.2Å) compared to the results using OPM, PDBTM, and OREMPRO
481 membrane coordinates as reference. On the other hand, the average membrane-embedded
482 residue accuracy MCC of ~0.5 was higher than the database-based calculations. Of the two
483 calculations, the upper limits calculated for the zα-scans (0.58) were similar to that of the
484 database-based calculations (0.61). However, this approach failed to improve the results of the
485 zαβ-scans, whereby the membrane-embedded residue accuracy MCC value remained ~0.5 for
486 both cases. This discrepancy can be explained with a large difference between the binding
487 orientations of the MD-predicted and Rosetta-predicted geometries.

488 **Future directions**

489 The current algorithm was developed to do a grid search to identify low-scoring regions in
490 the membrane. Although the method was successful in its task, grid searches are computationally
491 expensive. The reason why a grid search was preferred for this algorithm despite the performance
492 limitation was that a full grid search enabled us to extract more information regarding the ideal
493 parameters in a quantitative manner. However, users of the algorithm need not be interested in
494 such detailed information on the system. Therefore, for the future applications of the protocol, a
495 scan algorithm based on gradient minimization will be considered rather than a grid search to
496 reduce the time necessary for the calculations.

497 Conclusion

498 Despite the limitations associated with calculation of helix membrane coordinates and
499 validation of the calculated results, our calculations overall demonstrated that the AmphiScan
500 protocol can predict the embedding depth of amphipathic helices within ~2.5Å and with a
501 membrane-embedded residue accuracy MCC of ~0.6. Further, the AmphiScan protocol can be

502 coupled with other PyRosetta applications to model ideal α-helices from sequence, which makes
503 it an excellent tool to screen for amphipathic helices based on sequence or structure information.

504 Materials and Methods

505 Structural modeling of the LK peptides

506 The LK peptides were selected based on their published amphipathicity and helicity under
507 experimental conditions and the sequences of the highly-helical LK peptides were taken from (21)
508 for the AmphiScan calculations. Each peptide was modeled with PyMOL as an ideal α-helix with
509 φ- and ψ-angles of -57° and -47° respectively.

510 The *seaborn* module of python was used to plot the heatmaps. The distance and rotation
511 parameters were defined as discrete values and the color coding was done based on the score
512 distribution for the calculation whereby blue color indicates favorable scores and red color
513 indicates unfavorable scores. Color coding was done separately for the score range of each plot.
514 The boundary of the membrane was shown with a black bar at 15Å on both faces.

515 Selection of the amphipathic helix structures

516 Because there is no database for amphipathic helices to select from, ~3000 peripheral
517 and integral membrane proteins in the OPM database bearing at least one α-helix motif were
518 visually inspected to identify α-helices in close contact with the membrane surface to a total of
519 ~100 amphipathic helix structures. Only the helices that were explicitly identified as amphipathic
520 by the authors of the primary citation were kept to ascertain the amphipathicity of the selected
521 helices. Amphipathic helices with disulfide bonds to other protein domains and unsolved side
522 chain coordinates were removed from the candidate list to prevent biases caused by
523 conformational restraints in the native structure and side-chain modeling by Rosetta. In addition,
524 the amphipathic helices belonging to protein regions that are known to go through large
525 conformational changes based on literature reports or structures for which differences in the
526 amphipathic helix coordinates were observed in other PDB entries were eliminated. For the
527 remaining 33 protein structures, only the amphipathic helices that are in close contact with the

membrane mid-polar region were used for the calculations. When multiple helices were present in a protein structure, the helices were labeled as "*_hn*" where n stands for the nth amphipathic helix within the same protein structure starting from the N-terminus (Supporting Table 1). The final benchmark set consisted of 44 amphipathic helices in total.

Transformation of the helix coordinates prior to AmphiScan calculations

PyRosetta (40) was used to create all the Rosetta protocols and the RosettaMP framework (41) was used to carry all the membrane protein operations. A new membrane mover was created based on the RosettaMP *AddMembraneMover* by removing the requirement to load span files since the amphipathic helices usually do not span the membrane. The membrane was kept fixed centered at (0, 0, 0) and only the protein was allowed to move for the scan calculations. The helices were moved to the scan starting point such that the helix center of mass defined as the centroid of all Cα s was 10Å from the membrane surface along the membrane normal (z-axis).

For the 44 amphipathic helices in the benchmark set, membrane thickness calculated for the whole protein hosting the amphipathic helix was used for each calculation with different reference structures to prevent RMSD variations caused by the differences between the default hydrophobic thickness of RosettaMembrane (30Å) and the hydrophobic thicknesses calculated by the databases used to obtain the reference structures. For the LK peptides, the default bilayer thickness was used. All thickness values were rounded to the nearest integer.

Adjustment of the helix membrane orientations

Following the placement of the helices to the starting z-values, the helices were aligned with the xy-plane for sake of consistency. The helix screw axis was calculated as the difference vector between the average Cα coordinates of the four residues on each end skipping the first and the last residues. The angle between the helix screw axis and x-, y-, and z-axes of the membrane was back-calculated using the definition of the dot product with the following equation:

$$\theta = \cos^{-1}\left(\frac{(\vec{u}\cdot\vec{v})}{(\|\vec{u}\|\cdot\|\vec{v}\|)}\right)$$

where u stands for the helix screw axis, v stands for the corresponding x-, y-, or z-axis, and θ stands for the angle between the helix screw axis and the coordinate system axis.

For the zα-scans, the tilt angle of the helices from the database coordinates were kept as the angle with the membrane normal, and the helices were only rotated to be orthogonal to the y-axis to retain the original tilt angle. For the zαβ-scans, the helices were first rotated to be orthogonal to the z-axis and then orthogonal to the y-axis to keep the helix screw axis parallel to the x-axis (Figure 1C). All the geometric operations were done using the *WholeBodyTranslationMover* and *WholeBodyRotationMover* movers of the RosettaMP framework.

AmphiScan search protocol

Grid scans were run following the alignment operations as summarized in Figure 1. Preliminary calculations with different helix rotation, tilt angle, and depth intervals showed the best resolution of the scores within the shortest amount of time was achieved with a depth interval of 0.1Å, helix rotation angle of 5˚, and tilt angle of 1˚, so these values were used as the step sizes for the AmphiScan calculations. Scoring was done with the *mpframework2012* score function, which is the implementation of the RosettaMembrane score function to the RosettaMP framework. Helices were scored within 10Å from the membrane surface on both ends of the membrane. At each depth value, helices were rotated along their screw axis from 0˚ to 180˚. For the zαβ-scans, the helix angle with the y-axis was varied between 0˚ to 90˚, and independent AmphiScan calculations were run at each tilt angle. Of the resulting energy profiles, the tilt angle that gave the best-score among all the tilt angles was considered to be the best, and the depth and helix rotation angle values belonging to this tilt angle were considered to be the best parameters for each helix.

Root mean square deviation and membrane-embedded residue accuracy calculations

RMSD values were calculated as the average root mean squared distance between the coordinates of the helix Cα of the best-scoring Rosetta structure and the corresponding reference structure from OPM, PDBTM, OREMPRO, or MD simulations. The center of mass of each

581 reference structure was subjected to the same translation and rotation operations as the helices
582 to eliminate translational differences between the reference and Rosetta coordinates.
583 A membrane-embedded residue was defined as a residue with an atom at least 1Å inside
584 the membrane surface. Membrane-embedded residue accuracies were calculated according to
585 the following equation:

586 $$MCC = \left(\frac{TP\ x\ TN - FP\ x\ FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}\right) \text{ x } 100$$

587 where TP stands for the number of correctly predicted membrane-embedded residues,
588 TN stands for the number of correctly predicted non-membrane-embedded residues, and the # of
589 residues stands for the total number of amino acids in the given amphipathic helix.

590 **Rotamer recovery rate calculations**

591 Rotamer recovery rate calculations were done using an in-house script utilizing BioPython.
592 Specifically, the input structures used for the AmphiScan calculations were used as the native
593 structures for the comparisons and the lowest-scoring repacked structure was used as the decoy
594 structure. The χ1 and χ2 angles of each residue of the decoy structures were compared with that
595 of the native structures, and residues with rotamer matches within 25° were considered to be
596 recovered.

597 **Unrestrained molecular dynamics simulations**

598 The OPM membrane coordinates were used as the reference point for the MD simulations.
599 Explicit DOPC molecules, neutralizing Na^+ and Cl^- ions, and TIP3P water molecules were placed
600 using the CHARMM-GUI server (42). Each system had a total of ~100-110 DOPC molecules. The
601 structures from CHARMM-GUI were converted to Amber structure format using a combination of
602 the *charmmlipid2amber.py* script and in-house bash scripts. The DOPC molecules were
603 parametrized with the Amber lipid14 force field (43) and the helix residues were parametrized with
604 the Amber ff14SB force field (44).

605 The system was minimized in three steps. A restraint weight of 10 kcal/mol $Å^2$ was applied
606 to all non-hydrogen atoms in the first step, a restraint weight of 5 kcal/mol $Å^2$ was applied to

607 protein backbone (CA, C, O, N) atoms and lipid molecules in the second step, and an unrestrained
608 minimization was run in the third step. The final structure from the third minimization step was
609 heated in four steps, each 5 ns (50K, 100K, 200K, 300K) to a final temperature of 300K in an NVT
610 ensemble with Langevin thermostat. A restraint weight of 5 kcal/mol $Å^2$ was applied to the protein
611 backbone CA, C, O, N atoms, and lipid molecules. For each simulation, the system was heated
612 for 2.5 ns and equilibrated at the target temperature for 2.5 ns. The heated system was
613 equilibrated in an NPT ensemble in two 5 ns steps, first with a restraint of 5 kcal/mol $Å^2$ on
614 backbone CA, C, O, N atoms and lipid molecules, then with a restraint of 2 kcal/mol $Å^2$. Lastly, a
615 1 μs production run was run in an NPT environment with no restraints applied on the helix or lipid
616 molecules.

617 **Restrained molecular dynamics simulations**

618 The steps to obtain the starting coordinates were the same as the unrestrained simulations
619 up to the 1 μs production MD simulations with the exception that N-terminus acylation (ACE) and
620 C-terminus amidation (NME) was applied at the parametrization stage. At the 1 μs production
621 step of the MD simulations, restraints were applied to the backbone hydrogen bonds to keep the
622 structural integrity of the helices in restrained MD simulations. Specifically, the distance between
623 the carbonyl oxygen atom and the amide hydrogen atoms were calculated for the native structures
624 and the corresponding distances were restrained with a force constant of 10 kcal/mol $Å^2$.

625 **Molecular dynamics trajectory analysis**

626 Trajectory analyses were done with the *pytraj* module of AmberTools (45,46). 800
627 snapshots from the 1 μs MD trajectories between 200 – 1000 ns were used for the analyses. The
628 distance between the average coordinates of phosphatidylcholine atoms O21 and O11 and the
629 lipid center of mass was used to calculate the monolayer thicknesses of the membrane. The
630 corresponding monolayer thickness of the helix-bound side of the membrane of was used as the
631 thickness in AmphiScan calculations for all helices. RMSD values for the best-scoring AmphiScan

632 poses were calculated with respect to 800 snapshots and the average RMSD values were
633 reported as the final RMSD values.

634 **Acknowledgements**

635 The authors would like to thank Dr. Julia Koehler Leman for scientific feedback on the
636 project and Dr. Cristina E. Martina for critical reading of the manuscript. Funding for this work was
637 provided by NIH R01 HL144131 and NIH NIGMS R01 GM080403.

638

639

References

1. Giménez-Andrés M, Čopič A, Antonny B. The many faces of amphipathic helices. Biomolecules. 2018;8(3):1–14.
2. Sharadadevi A, Sivakamasundari C, Nagaraj R. Amphipathic α-helices in proteins: Results from analysis of protein structures. Proteins Struct Funct Genet. 2005;59(4):791–801.
3. Khandelia H, Ipsen JH, Mouritsen OG. The impact of peptides on lipid membranes. Biochim Biophys Acta - Biomembr. 2008;1778(7–8):1528–36.
4. Segrest JP, De Loof H, Dohlman JG, Brouillette CG, Anantharamaiah GM. Amphipathic helix motif: Classes and properties. Proteins Struct Funct Bioinforma. 1990;8(2):103–17.
5. Drin G, Antonny B. Amphipathic helices and membrane curvature. FEBS Lett [Internet]. 2010;584(9):1840–7. Available from: http://dx.doi.org/10.1016/j.febslet.2009.10.022
6. Schibli DJ, Hwang PM, Vogel HJ. Structure of the antimicrobial peptide tritrpticin bound to micelles: A distinct membrane-bound peptide fold. Biochemistry. 1999;38(51):16749–55.
7. Hancock REW, Scott MG. The role of antimicrobial peptides in animal defenses. Proc Natl Acad Sci [Internet]. 2000;97(16):8856–61. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.97.16.8856
8. Reißer S, Strandberg E, Steinbrecher T, Ulrich AS. 3D hydrophobic moment vectors as a tool to characterize the surface polarity of amphiphilic peptides. Biophys J [Internet]. 2014;106(11):2385–94. Available from: http://dx.doi.org/10.1016/j.bpj.2014.04.020
9. Gautier R, Douguet D, Antonny B, Drin G. HELIQUEST: A web server to screen sequences with specific α-helical properties. Bioinformatics. 2008;24(18):2101–2.
10. Gofman Y, Haliloglu T, Ben-Tal N. Monte Carlo simulations of peptide-membrane interactions with the MCPep web server. Nucleic Acids Res. 2012;40(W1):358–63.
11. Shental-Bechor D, Haliloglu T, Ben-Tal N. Interactions of cationic-hydrophobic peptides with lipid bilayers: A Monte Carlo simulation method. Biophys J. 2007;93(6):1858–71.
12. Reid LM, Verma CS, Essex JW. The role of molecular simulations in understanding the mechanisms of cell-penetrating peptides. Drug Discov Today [Internet]. 2019;24(9):1821–35. Available from: https://doi.org/10.1016/j.drudis.2019.06.013
13. Muller MP, Jiang T, Sun C, Lihan M, Pant S, Mahinthichaichan P, et al. Characterization of Lipid-Protein Interactions and Lipid-Mediated Modulation of Membrane Protein Function through Molecular Simulation. Chem Rev. 2019;119(9):6086–161.
14. Yarov-Yarovoy V, Schonbrun J, Baker D. Multipass membrane protein structure prediction using Rosetta. Proteins Struct Funct Genet [Internet]. 2006 Mar 1;62(4):1010–25. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16372357
15. Barth P, Schonbrun J, Baker D. Toward high-resolution prediction and design of transmembrane helical protein structures. Proc Natl Acad Sci [Internet]. 2007;104(40):15682–7. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0702515104
16. Barth P, Wallner B, Baker D. Prediction of membrane protein structures with complex topologies using limited constraints. Proc Natl Acad Sci U S A. 2009;106(5):1409–14.
17. Koehler Leman J, Ulmschneider MB, Gray JJ. Computational modeling of membrane proteins. Proteins Struct Funct Bioinforma. 2015;83(1):1–24.
18. Lazaridis T. Effective energy function for proteins in lipid membranes. Proteins Struct Funct Bioinforma [Internet]. 2003;192(September 2002):176–92. Available from: http://onlinelibrary.wiley.com/doi/10.1002/prot.10410/full
19. Cornut I, Büttner K, Dasseux JL, Dufourcq J. The amphipathic α-helix concept. Application to the de novo design of ideally amphipathic Leu, Lys peptides with hemolytic activity higher than that of melittin. FEBS Lett. 1994;349(1):29–33.
20. Schrödinger L. The PyMOL molecular graphics system, version 1.8.

Https://Www.Pymol.Org/Citing. 2015.
21. Castano S, Cornut I, Büttner K, Dasseux JL, Dufourcq J. The amphipathic helix concept: length effects on ideally amphipathic LiKj(i=2j) peptides to acquire optimal hemolytic activity. Biochim Biophys Acta - Biomembr [Internet]. 1999;1416(1):161–75. Available from: http://www.sciencedirect.com/science/article/pii/S0005273698002193
22. Castano S, Desbat B, Laguerre M, Dufourcq J. Structure, orientation and affinity for interfaces and lipids of ideally amphipathic lytic L(i)K(j)(i=2j) peptides. Biochim Biophys Acta - Biomembr. 1999;
23. Hohlweg W, Kosol S, Zangger K. Determining the Orientation and Localization of Membrane-Bound Peptides. Curr Protein Pept Sci. 2012;13(3):267–79.
24. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: Resources for positioning of proteins in membranes. Nucleic Acids Res. 2012;40(D1):370–6.
25. Lomize AL, Pogozheva ID, Mosberg HI. 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(4):930–46.
26. Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI. The role of hydrophobic interactions in positioning of peripheral proteins in membranes. BMC Struct Biol. 2007;7:1–30.
27. Tusnády GE, Dosztányi Z, Simon I. PDB_TM: Selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res. 2005;33(DATABASE ISS.):275–8.
28. Tusnády GE, Dosztányi Z, Simon I. TMDET: Web server for detecting transmembrane regions of proteins by using their 3D coordinates. Bioinformatics. 2005;21(7):1276–7.
29. Postic, Guillaume; Ghouzam Y. OREMPRO web server: orientation and assessment of atomistic and coarse-grained structures of membrane proteins. Bioinformatics. 2016;32(16):2548–2550.
30. Postic G, Ghouzam Y, Guiraud V, Gelly JC. Membrane positioning for high- and low-resolution protein structures through a binary classification approach. Protein Eng Des Sel. 2015;29(3):87–91.
31. Shimizu K, Cao W, Saad G, Shoji M, Terada T. Comparative analysis of membrane protein structure databases. Biochim Biophys Acta - Biomembr [Internet]. 2018;1860(5):1077–91. Available from: https://doi.org/10.1016/j.bbamem.2018.01.005
32. Newport TD, Sansom MSP, Stansfeld PJ. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. Nucleic Acids Res. 2019;47(D1):D390–7.
33. Middleton ER, Rhoades E. Effects of curvature and composition on α-synuclein binding to lipid vesicles. Biophys J [Internet]. 2010;99(7):2279–88. Available from: http://dx.doi.org/10.1016/j.bpj.2010.07.056
34. Vanni S, Hirose H, Barelli H, Antonny B, Gautier R. A sub-nanometre view of how membrane curvature and composition modulate lipid packing and protein recruitment. Nat Commun. 2014;5.
35. Cornell RB. Membrane lipid compositional sensing by the inducible amphipathic helix of CCT. Biochim Biophys Acta - Mol Cell Biol Lipids [Internet]. 2016;1861(8):847–61. Available from: http://dx.doi.org/10.1016/j.bbalip.2015.12.022
36. Huang H, Ball JM, Billheimer JT, Schroeder F. Interaction of the N-terminus of sterol carrier protein 2 with membranes: role of membrane curvature. Biochem J. 1999;344(2):593–603.
37. Wang Y, Zhao T, Wei D, Strandberg E, Ulrich AS, Ulmschneider JP. How reliable are molecular dynamics simulations of membrane active antimicrobial peptides? Biochim Biophys Acta - Biomembr [Internet]. 2014;1838(9):2280–8. Available from: http://dx.doi.org/10.1016/j.bbamem.2014.04.009
38. Sandoval-Perez A, Pluhackova K, Böckmann RA. Critical Comparison of Biomembrane

Force Fields: Protein-Lipid Interactions at the Membrane Interface. J Chem Theory Comput. 2017;13(5):2310–21.

39. Wang Y, Schlamadinger DE, Kim JE, McCammon JA. Comparative molecular dynamics simulations of the antimicrobial peptide CM15 in model lipid bilayers. Biochim Biophys Acta - Biomembr [Internet]. 2012;1818(5):1402–9. Available from: http://dx.doi.org/10.1016/j.bbamem.2012.02.017

40. Chaudhury S, Lyskov S, Gray JJ. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics [Internet]. 2010/01/07. 2010 Mar 1;26(5):689–91. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20061306

41. Alford RF, Koehler Leman J, Weitzner BD, Duran AM, Tilley DC, Elazar A, et al. An Integrated Framework Advancing Membrane Protein Modeling and Design. PLoS Comput Biol. 2015;11(9).

42. Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, et al. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J Comput Chem [Internet]. 2014;35(27):1997–2004. Available from: http://dx.doi.org/10.1002/jcc.23702

43. Dickson CJ, Madej BD, Skjevik ÅA, Betz RM, Teigen K, Gould IR, et al. Lipid14: The Amber Lipid Force Field. J Chem Theory Comput [Internet]. 2014 Feb 11;10(2):865–79. Available from: https://doi.org/10.1021/ct4010307

44. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015;11(8):3696–713.

45. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput. 2013;9(7):3084–95.

46. Nguyen H, Roe DR, Swails JM, Case DA. PYTRAJ: Interactive data analysis for molecular dynamics simulations. 2016.

Figures

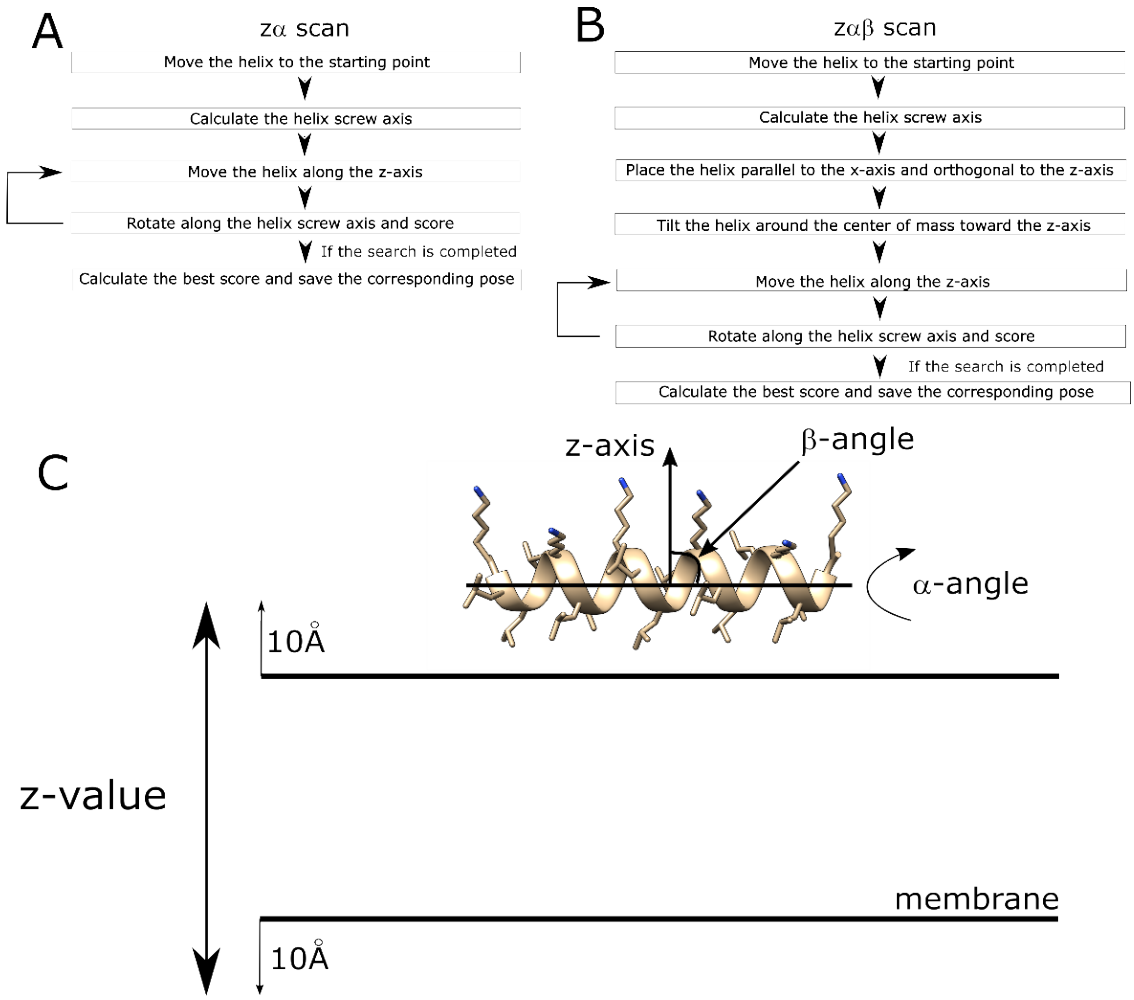


Figure 1: Flowcharts of the AmphiScan calculations: (A) zα-scan flowchart, (B) zαβ-scan flowchart, (C) visual representation of the rotation and tilt operations used to place the helices with respect to the membrane surface. Note that the plane of the membrane corresponds to the xy-plane and the membrane normal corresponds to the z-axis.

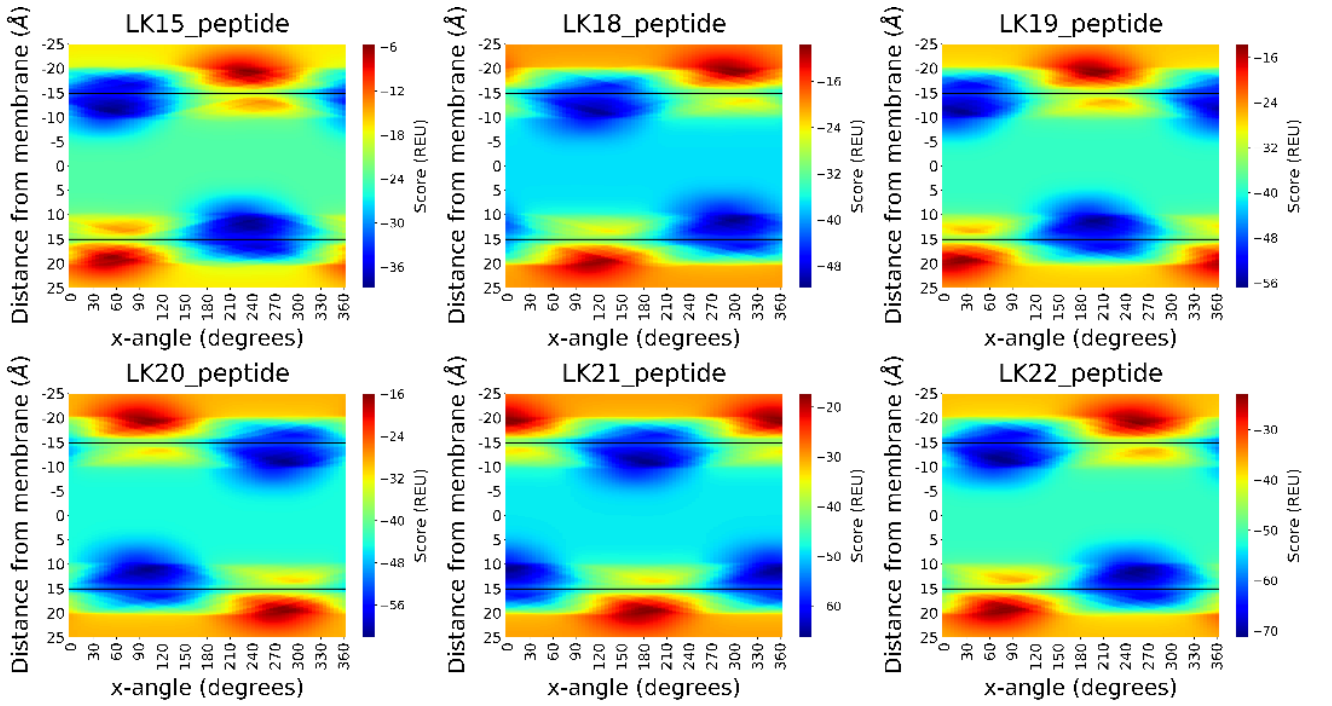


Figure 2: Heatmaps for the Rosetta scans at the best-scoring tilt angle for six LK peptides. The x-axis represents the rotation along screw axis of the helices and the y-axis stands for the depth of the membrane. Color coding was done separately for each helix where blue color represents the low-scoring regions and the red color represents the high-scoring regions. The black bars represent the boundary of the hydrophobic thickness of the implicit membrane model.

787

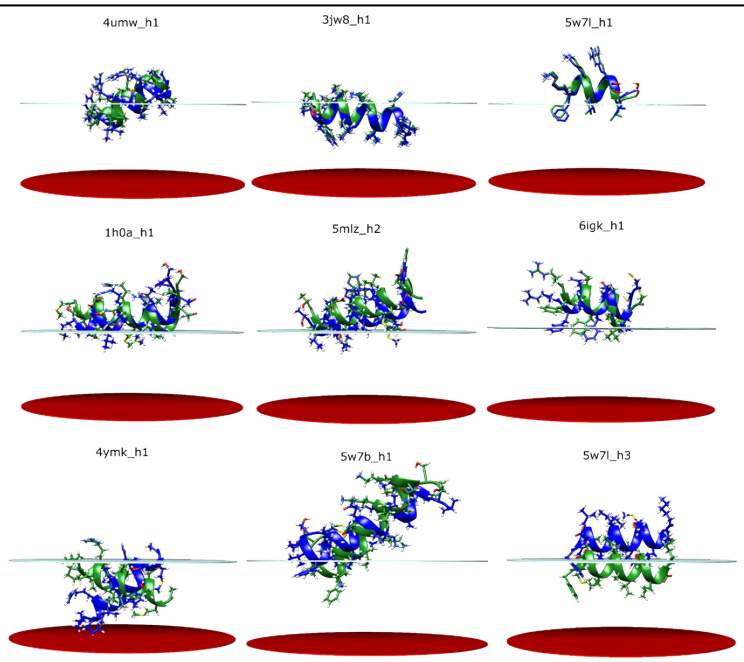


Figure 3: Example alignment poses for the native (green) and best pose structures (blue) from the OPM zαβ calculations. The structures in the top row had RMSD values smaller than 2Å, the middle row had RMSD values between 2 and 4Å, and the bottom row had RMSD values larger than 4Å.

Tables

Table 1: Sequences of the LK peptides used for the AmphiScan calculations.

LK15	K	L	L	K	L	L	L	K	L	L	L	K	L	L	K							
LK18	L	L	L	K	L	L	K	L	L	L	K	L	L	L	K	L	L	K				
LK19	K	L	L	L	K	L	L	K	L	L	L	K	L	L	L	K	L	L	K			
LK20	L	K	L	L	L	K	L	L	K	L	L	L	K	L	L	L	K	L	L	K		
LK21	L	L	K	L	L	L	K	L	L	K	L	L	L	K	L	L	L	K	L	L	K	
LK22	K	L	L	K	L	L	L	K	L	L	K	L	L	L	K	L	L	L	K	L	L	K

Table 2: The RMSD values calculated for the best Rosetta zα-scan poses where the depth and helix rotation around the screw axis were allowed, and the source of the starting structure used for the calculations. The standard deviations were calculated as the standard deviation among the RMSD values of the same helix with the starting structures from three different sources. All units are in Ångstroms (Å).

Name	zα OPM	zαβ OPM	zα OREMPRO	zαβ OREMPRO	zα PDBTM	zαβ PDBTM
1rhz_h1	2.8	2.2	2.8	2.0	2.7	1.8
2ziy_h1	2.8	2.7	2.8	2.6	1.5	1.4
3a7k_h1	1.4	1.7	2.4	2.7	3.9	4.0
3j5p_h1	1.0	1.9	0.8	0.7	1.0	2.3
4qnd_h1	4.7	4.1	1.2	2.2	2.4	2.0
4rp9_h3	3.8	3.7	1.6	1.5	2.8	3.2
4umw_h1	1.6	1.7	2.1	3.2	5.1	4.5
4ymk_h1	4.2	6.7	4.1	2.0	6.2	5.4
4ymk_h2	5.4	4.4	5.3	4.3	5.6	4.1
4ymk_h3	5.0	4.5	1.7	1.7	2.4	1.6
5dqq_h1	7.2	6.8	3.0	3.2	1.6	4.6
5lil_h1	3.5	6.8	1.1	7.7	1.7	4.3
5mlz_h2	2.3	2.2	5.0	2.2	1.7	2.0
5uz7_h1	2.0	1.8	9.4	9.4	4.2	8.6
6an7_h1	5.3	8.4	5.3	9.5	4.8	8.1
6igk_h1	2.1	2.0	2.0	2.0	2.8	4.4
Average	3.4	3.8	3.2	3.6	3.2	3.9

Table 3: The RMSD and membrane-embedded residue accuracies calculated from MD simulations.

Helix ID	MD thickness	zα RMSD	zα Stdev	zα MCC	zαβ RMSD	zαβ Stdev	zαβ MCC

1b4v_h1	14.6	5.3	1.5	0.50	4.7	1.4	0.41
1h0a_h1	14.4	4.6	1.6	0.86	4.6	1.6	0.45
1q4g_h1	14.9	3.7	1.4	0.69	3.7	1.4	0.69
1q4g_h2	14.8	3.8	1.4	0.69	4.0	1.3	0.83
1q4g_h3	14.7	3.5	1.3	0.57	3.7	1.3	0.46
1q4g_h4	14.8	6.5	1.2	0.02	2.8	1.1	0.63
1rhz_h1	14.2	2.6	0.8	0.59	2.8	0.8	0.59
2hih_h1	14.4	4.8	1.2	0.55	4.2	1.2	0.66
2ziy_h1	14.8	3.4	1.4	0.76	3.5	1.4	0.76
3a7k_h1	14.8	3.9	1.5	0.41	4.6	1.5	0.50
3hyw_h1	14.2	2.9	1.0	0.38	6.9	0.8	-0.15
3hyw_h2	14.6	2.6	1.0	1.00	2.6	0.9	0.82
3i9v_h1	14.6	4.1	1.4	0.60	3.7	1.3	0.77
3j5p_h1	14.6	2.6	1.0	0.70	2.7	1.0	0.54
3jw8_h1	14.7	2.0	0.9	0.35	2.0	0.9	0.35
3tij_h1	13.9	3.7	0.9	0.55	2.8	0.9	0.52
4hhr_h1	14.7	3.0	0.9	0.29	2.7	0.9	0.08
4hhr_h2	14.6	3.9	1.6	-0.16	2.7	1.1	-0.13
4hhr_h3	14.6	8.4	1.3	-0.24	2.4	0.8	0.64
4m5e_h1	14.6	4.2	1.4	0.57	4.2	1.4	0.57
4nwz_h1	14.2	2.5	1.0	0.73	6.4	1.6	0.24
4qnd_h1	14.6	3.6	1.3	0.57	2.8	1.2	0.82
4rp9_h3	14.8	4.2	1.9	0.50	4.1	1.9	0.50
4umw_h1	14.8	4.6	1.3	0.67	4.1	1.4	0.67
4ymk_h1	14.7	2.8	0.9	0.00	5.2	1.0	0.00

4ymk_h2	15.0	3.2	1.1	0.50	2.7	1.2	0.77
4ymk_h3	14.6	3.1	1.4	0.51	5.0	1.3	-0.19
4zwn_h1	14.3	4.2	1.3	0.71	4.1	1.3	0.81
5ahv_h1	14.6	5.7	1.4	0.23	6.0	1.7	0.18
5dqq_h1	14.6	3.1	1.3	0.76	2.9	1.1	0.49
5ek8_h1	14.2	3.2	1.0	0.89	2.8	1.0	0.68
5f19_h3	14.7	4.6	1.6	0.67	4.5	1.6	0.67
5f19_h4	14.7	7.0	2.8	0.57	3.9	1.3	0.77
5lil_h1	14.6	4.5	1.4	0.25	4.3	1.1	0.57
5mlz_h2	14.5	2.8	1.1	0.74	2.7	1.2	0.74
5uz7_h1	14.5	6.0	0.8	-0.26	6.0	0.8	-0.30
5w7b_h1	14.7	8.8	1.5	0.37	3.5	1.3	0.62
5w7l_h1	14.8	4.0	1.4	0.65	3.4	1.4	0.65
5w7l_h2	14.7	4.7	1.8	0.49	4.3	1.8	0.60
5w7l_h3	14.7	3.9	1.6	0.85	3.9	1.6	0.85
6an7_h1	13.8	11.1	1.5	0.00	14.4	1.4	0.00
6d26_h1	14.6	3.8	1.4	0.66	3.4	1.3	0.77
6dvy_h1	13.8	8.6	1.9	-0.05	7.9	2.0	0.35
6igk_h1	14.6	4.6	1.7	0.71	6.6	2.0	0.58
Average	14.6	4.4	1.3	0.49	4.2	1.3	0.50

Supporting Tables

Supporting Table 1: List of the amphipathic helices in the benchmark set and their corresponding residue numbers and chain IDs in the PDB files. The "_xn" denotation was used to distinguish multiple helices belonging to the same protein structure.

Helix name	PDB region
1b4v_h1	78-87.A
1h0a_h1	1-18.A
1q4g_h1	73-83.A
1q4g_h2	85-95.A
1q4g_h3	96-106.A
1q4g_h4	107-123.A
1rhz_h1	2-23.B
2hih_h1	190-206.A
2ziy_h1	320-329.A
3a7k_h1	20-29.A
3hyw_h1	380-395.A
3hyw_h2	400-412.A
3i9v_h1	18.6-33.6
3j5p_h1	559-575.A
3jw8_h1	167-181.A
3tij_h1	50-78.A
4hhr_h1	1-15.A
4hhr_h2	9-44.A
4hhr_h3	51-70.A
4m5e_h1	160-172.A
4nwz_h1	369-388.A
4qnd_h1	1-13.A
4rp9_h3	252-261.A
4umw_h1	165-176.A
4ymk_h1	128-141.A
4ymk_h2	281-288.A
4ymk_h3	315-326.A
4zwn_h1	156-176.A
5ahv_h1	1-16.E
5dqq_h1	198-213.A
5ek8_h1	1-23.A
5f19_h3	96-105.A
5f19_h4	106-122.A
5lil_h1	246-263.A
5mlz_h2	213-229.A
5uz7_h1	395-418.R
5w7b_h1	159-179.C

5w7l_h1	82-91.A
5w7l_h2	93-104.A
5w7l_h3	166-179.A
6an7_h1	3-19.D
6d26_h1	2309-2325.A
6dvy_h1	681-708.A
6igk_h1	302-403.A

Supporting Table 2: The RMSD values calculated for the best Rosetta zα- and zαβ-scan poses with OPM membrane thicknesses.

Helix ID	zα RMSD (Å)	zα MCC	zαβ RMSD (A)	zαβ MCC
1b4v_h1	2.1	0.00	3.9	0.00
1h0a_h1	0.5	1.00	2.2	0.27
1q4g_h1	3.6	0.69	3.6	0.69
1q4g_h2	3.2	0.39	3.7	0.26
1q4g_h3	4.2	0.36	4.2	0.29
1q4g_h4	13.3	-0.38	5.7	0.63
1rhz_h1	2.8	0.59	2.2	0.59
2hih_h1	1.9	0.83	1.8	1.00
2ziy_h1	2.8	0.67	2.7	0.67
3a7k_h1	1.5	0.36	1.7	0.22
3hyw_h1	3.8	0.38	5.6	-0.15
3hyw_h2	2.3	1.00	2.3	0.82
3i9v_h1	1.1	1.00	2.3	0.77
3j5p_h1	1.0	0.89	2.0	0.53
3jw8_h1	0.4	1.00	0.7	0.58
3tij_h1	6.3	0.14	5.0	0.17

4hhr_h1	1.5	0.78	3.0	0.08
4hhr_h2	0.4	0.46	2.9	0.51
4hhr_h3	2.9	0.14	6.3	0.55
4m5e_h1	1.6	0.82	1.8	0.82
4nwz_h1	3.2	0.40	8.9	0.19
4qnd_h1	4.7	0.37	4.1	0.44
4rp9_h3	3.8	0.41	3.7	0.41
4umw_h1	1.7	0.63	1.7	0.63
4ymk_h1	4.2	0.15	6.7	0.53
4ymk_h2	5.4	0.00	4.4	-0.15
4ymk_h3	5.0	0.49	4.6	0.56
4zwn_h1	0.9	0.41	1.7	0.34
5ahv_h1	4.8	0.37	3.5	0.29
5dqq_h1	7.2	0.17	6.8	0.17
5ek8_h1	0.3	1.00	0.7	0.80
5f19_h3	2.0	0.10	1.8	0.10
5f19_h4	0.6	1.00	6.3	0.44
5lil_h1	3.5	0.27	6.8	-0.24
5mlz_h2	2.3	0.63	2.2	0.63
5uz7_h1	2.0	0.67	1.8	0.59
5w7b_h1	0.5	1.00	7.3	0.50
5w7l_h1	0.9	1.00	0.8	1.00
5w7l_h2	2.1	0.68	2.0	0.84
5w7l_h3	5.5	0.33	5.5	0.33
6an7_h1	5.3	-0.36	8.4	-0.25

6d26_h1	0.4	1.00	1.2	0.86
6dvy_h1	2.9	-0.05	2.0	0.35
6igk_h1	2.1	0.82	2.0	0.56
Average	2.9	0.51	3.6	0.44

Supporting Table 3: The RMSD values, MCC values, and rotamer recovery rates for χ1 and χ2 belonging to the zα calculations with side chain repacking.

Helix name	RMSD (Å)	MCC	Rotamer recovery rate (χ1)	Rotamer recovery rate (χ2)
1b4v_h1_renum	2.1	0.00	0.70	0.50
1h0a_h1_renum	0.5	1.00	0.61	0.39
1q4g_h1_renum	3.6	0.69	0.64	0.46
1q4g_h2_renum	3.1	-0.04	0.91	0.55
1q4g_h3_renum	4.2	0.36	0.67	0.56
1q4g_h4_renum	10.2	-0.26	0.53	0.24
1rhz_h1_renum	2.8	1.00	0.55	0.14
2hih_h1_renum	2.0	0.83	0.50	0.19
2ziy_h1_renum	2.8	1.00	0.67	0.44
3a7k_h1_renum	1.3	0.65	0.60	0.30
3hyw_h1_renum	5.0	-0.17	0.73	0.33
3hyw_h2_renum	2.3	0.82	0.62	0.31
3i9v_h1_renum	1.3	1.00	0.39	0.15
3j5p_h1_renum	0.9	0.76	0.47	0.27
3jw8_h1_renum	0.4	1.00	0.69	0.39
3tij_h1_renum	6.1	0.14	0.63	0.38
4hhr_h1_renum	1.5	0.78	0.73	0.47

4hhr_h2_renum	1.2	0.67	0.67	0.47
4hhr_h3_renum	2.9	0.14	0.79	0.58
4m5e_h1_renum	1.9	0.82	0.73	0.36
4nwz_h1_renum	3.2	0.35	0.81	0.69
4qnd_h1_renum	4.6	0.37	0.80	0.60
4rp9_h3_renum	3.8	0.27	1.00	0.50
4umw_h1_renum	1.6	0.63	0.60	0.50
4ymk_h1_renum	4.2	0.15	0.75	0.67
4ymk_h2_renum	5.4	0.00	1.00	0.50
4ymk_h3_renum	5.0	0.41	1.00	0.70
4zwn_h1_renum	0.8	0.65	0.53	0.16
5ahv_h1_renum	2.2	0.52	0.86	0.57
5dqq_h1_renum	7.2	0.17	0.33	0.13
5ek8_h1_renum	1.3	1.00	0.71	0.38
5f19_h3_renum	1.9	0.10	0.89	0.56
5f19_h4_renum	0.6	1.00	0.63	0.31
5lil_h1_renum	3.5	0.34	0.80	0.33
5mlz_h2_renum	2.2	0.55	0.56	0.31
5uz7_h1_renum	1.8	0.34	0.68	0.27
5w7b_h1_renum	11.8	-0.45	0.74	0.47
5w7l_h1_renum	1.0	0.76	0.56	0.56
5w7l_h2_renum	2.2	0.84	0.55	0.27
5w7l_h3_renum	5.3	0.39	0.92	0.69
6an7_h1_renum	5.3	-0.31	0.38	0.06
6d26_h1_renum	0.4	1.00	0.59	0.41

6dvy_h1_renum	0.3	0.46	0.59	0.41
6igk_h1_renum	2.1	0.82	0.70	0.50
Average	3.0	0.49	0.68	0.41

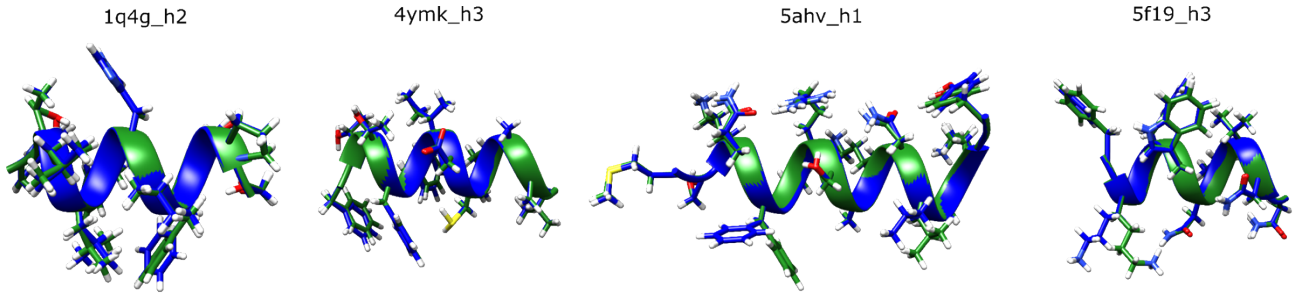
Supporting Table 4: Hydrophobic thicknesses calculated by different methods for each protein. The thickness calculations were done on full proteins, and the same thickness value was used for all the helices belonging to the same protein structure. Standard deviations were calculated for the different thicknesses given by different methods for the same protein structure. All units are in Ångstroms (Å).

PDB ID	OPM	OREMPRO	PDBTM	Stdev
1rhz	14.6	14.5	14.5	0.1
2ziy	16.5	17.0	15.3	0.9
3a7k	16.9	16.0	13.8	1.6
3j5p	14.2	16.6	16.5	1.4
4qnd	16.5	12.0	16.0	2.5
4rp9	14.9	10.5	13.0	2.2
4umw	15.5	11.5	12.8	2.0
4ymk	13.8	12.3	11.5	1.2
5dqq	14.9	11.1	12.5	1.9
5lil	15.5	19.5	18.0	2.0
5mlz	15.3	14.3	14.8	0.5
5uz7	15.1	20.0	18.0	2.5
6an7	14.9	15.5	14.3	0.6
6igk	17.1	18.0	14.8	1.7

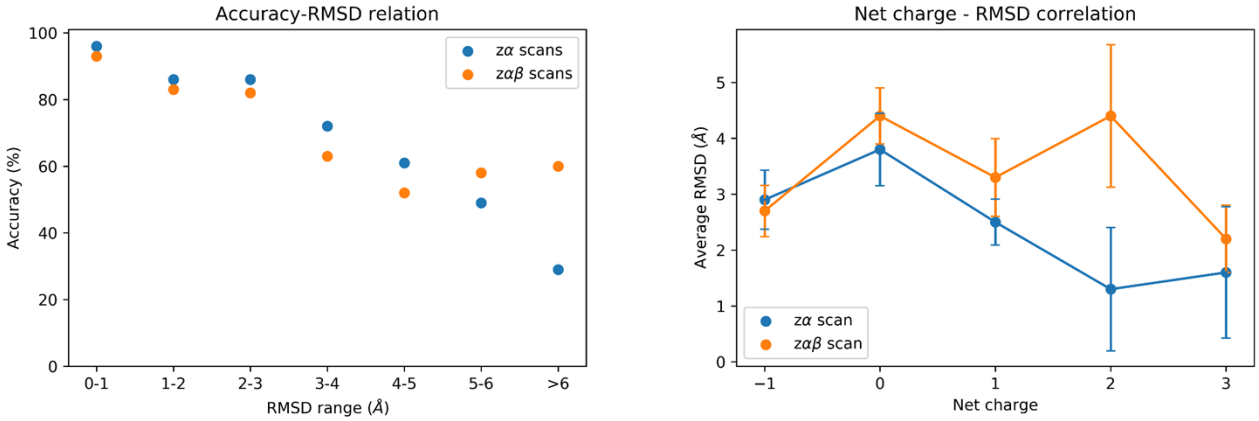
829

830

Supporting Figures



Supporting Figure 1: Example structures of the lowest-scoring poses from the calculations with side chain repacking with χ1 recovery rates over 0.85 aligned with the native structures. Green color indicates the native structure and blue color indicates the lowest-scoring Rosetta structure.



Supporting Figure 2: Left panel: The correlation between the calculated RMSD values and the membrane-embedded residue accuracies for the helices in the benchmark set. The accuracy values within 1Å RMSD range were averaged for a clearer representation. Right panel: The correlation between the net charge and the averaged RMSD values bearing the same net charge. Blue represents the result of the zα-scans and orange represents the results of the zαβ-scans.