A Learning Algorithm to Discover Soluble Vesicle-Binding Helical Peptides

Sharlene Denos,[†] Eric Gotkowski,[‡] and Martin Gruebele*[,†,‡,§]

Center for Biophysics and Computational Biology, Department of Chemistry, and Department of Physics, University of Illinois, Urbana−Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801

Received: December 24, 2009; Revised Manuscript Received: February 25, 2010

Membrane peptide folding studies require peptides that bind to lipid vesicles while remaining water-soluble. Currently available peptides are either artificial designs, or they have membrane-disrupting antimicrobial or venomous activity. As a first step to derive new soluble membrane-binding peptides from naturally occurring membrane proteins, we trained a learning algorithm on several water-soluble and insoluble helical peptides by comparing its predictions with experimental solubility and fluorescence vesicle binding assays. The algorithm yielded an easily computed score *S* to discover soluble peptides in databases of transmembrane helical proteins. To validate the algorithm, we selected four helices based on a good *S* score. Experiments showed that all four are soluble at >25 μM, and that three bind to vesicles. We illustrate with an example that the vesicle binding of such peptides can be temperature-tuned. Finally, we predict four additional peptides that should be water-soluble and able to bind to lipid vesicles.

Introduction

Membrane binding and insertion reactions of helical peptides have been studied thermodynamically and kinetically by perturbing the peptide's partitioning between lipid vesicles and the aqueous phase. For example, the pH can be tuned to shift the binding equilibrium, or sudden temperature jumps can switch the membrane−peptide system out of equilibrium so that relaxation rates can be measured.[1,2]

Successful peptides require a delicate balance between the ability of a peptide to bind to the membrane and its solubility. The peptide's hydrophobicity, amphipathicity, and charge must be tuned carefully. Success has been achieved with designed peptides,[3,4] or with venomous, antibiotic, or antimicrobial sequences.[1,5,6] No peptides derived from natural membrane proteins have been identified for use in binding kinetics or thermodynamics studies.

Our goal is to analyze helical membrane proteins to discover new peptides that are water-soluble, and that will bind to lipid vesicles with an easily detectable spectroscopic signature. We train a learning algorithm on a set of existing or newly characterized helical peptides, some of which are soluble, while others are not. We derive an easily computed score *S* that can be used to mine existing helical membrane protein databases.[7] Peptides predicted to bind vesicles, yet remain soluble have $S > 0$. Our screening process does not rule out peptides that oligomerize in solution but remain soluble, or those that bind to the surface of vesicles without inserting into the bilayer, but it greatly reduces the playing field for further investigation.

We validated *S* as a useful score by experimentally screening four new predicted peptides for aqueous solubility and vesicle binding. All four are soluble, and three (G3PT-11, TetA-4, and SecY-10) bind vesicles. We discovered an additional suitable peptide (CCOIII-4) during learning-algorithm training. Finally, we predicted four more peptides to be soluble yet membrane-binding for future validation.

Our combined computational and experimental approach involves four major steps. First, we characterized a set of membrane-binding peptides experimentally for solubility. In the second step, we calculated for these peptides sequence-based parameters such as hydrophobicity and amphipathicity and employed a learning algorithm to search for the optimal linear combination *S* of sequence-based parameters capable of separating this training set into soluble and insoluble peptides. In the third step, the predictive power of the single score *S* was tested by synthesis and characterization of four new trans-membrane helix sequences predicted to be soluble at the micromolar level. All four new peptides were soluble, and three out of four bound to membrane, as measured by fluorescence wavelength or intensity shifts in the presence of small unilamellar vesicles (SUVs). The same solubility test was used for all peptides discussed in our study to ensure a uniform solubility criterion. In the final step, we predicted four additional new soluble membrane-binding peptides and present an overall statistical analysis of a membrane protein database.[7]

Experimental and Computational Methods

Peptide Nomenclature and Synthesis. Peptides originating from helical membrane proteins are given the abbreviated protein name followed by a number that indicates the transmembrane helix (counting from the N-terminus) to which the peptide corresponds. *De novo* peptide names are kept from the original source. All syntheses were performed by Genscript Corp. (Piscataway, NJ), except for the TetA-4 and G3PT-11 syntheses, which were performed by the University of Illinois Biotechnology Center (Urbana, IL). All peptides discussed in this study are listed in Tables 1 through 3. Point mutations to introduce tryptophan as a fluorescence label and to improve solubility are also noted in the tables.

Solubility and Binding Measurements. Solubility and binding measurements were performed in 50 mM sodium phosphate buffer. Initial UV/visible measurements in pH 6, 7, and 8 buffer for each peptide revealed which pH was optimal for further studies. All peptides were studied at pH 7 except for CCOIII-4 (pH 6).

* Corresponding author. E-mail: gruebele@scs.uiuc.edu.
[†] Center for Biophysics and Computational Biology.
[‡] Department of Chemistry.
[§] Department of Physics.

10.1021/jp912157g
Published on Web 03/17/2010

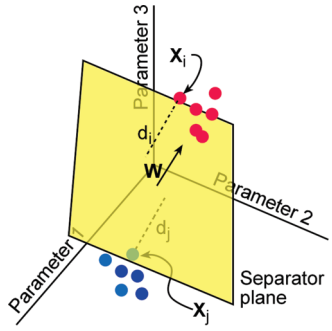

Figure 1. The optimization algorithm rotates and translates the yellow plane to maximize the distances d_i from each peptide point $\mathbf{X}_i$ to the plane, keeping soluble (red) and insoluble (blue) peptides on opposite sides of the plane in as small a parameter space (three shown here) as possible.

To determine solubility, peptides were centrifuged at 16 000 g for 20 min, and if no concentration decrease was observed in the supernatant, were spun at 125 000 g for an additional 2 h at 10 °C. The remaining supernatant concentration measured by UV absorption was taken as the solubility. Protein concentrations in the supernatant were determined by 280 nm absorbance. Extinction coefficients were calculated by summing all the tryptophan, tyrosine, and cysteine residues and assuming a spectral contribution at 280 nm of 5690, 1280, and 120 M^{-1} cm^{-1}, respectively, for each.[8] A peptide is defined as insoluble in the micromolar range if less than 1 μM remains in the supernatant after repeated centrifugation.

We studied peptide binding to neutral palmitoyl-oleoyl-phosphatidylcholine (POPC) vesicles and 19:1 POPC:POPG (POPG = palmitoyl-oleoyl-phosphatidylglycerol) vesicles with some anionic headgroups at the vesicle surface. Vesicle preparation is described in detail in the Supporting Information. Nearly uniform distributions of 50 nm SUVs were routinely obtained for binding experiments, as evidenced by electron microscopy, dynamic light scattering, and total phosphorus assay. SUV concentration was determined assuming an area of 64 $Å^2$ per lipid and a 3.9 nm bilayer thickness, which corresponds to ~21 000 lipids per 50 nm vesicle.[9]

Tryptophan or tyrosine fluorescence spectra were taken to test whether peptides were capable of SUV binding. All peptides were first tested with neutral POPC SUVs. If no binding was observed, 19:1 POPC:POPG SUVs were tested. Binding was determined by spectral blue shifts and large intensity changes in the presence of SUVs. The fluorescence excitation wavelength was 280 nm, (290 nm for Gly3P-11 to avoid tyrosine excitation). Emission spectra were obtained from 300 to 400 nm (290−350 nm for TetA-4 tyrosine emission).

We also tested the temperature-dependent binding of two peptides, to see how suitable they would be for thermal titration or temperature jump studies. Dry nitrogen was used to purge the sample chamber, and the temperature probe was placed directly in the buffer blank cuvette. Peptide concentrations were between 10 and 30 μM, and vesicle (not individual lipid) concentrations were between 0.15 and 0.4 μM. Singular value decomposition (SVD)[10] was used to detect temperature trends in the raw fluorescence spectra.

Learning Algorithm. Based on the solubility experiments, peptides are classified as soluble or insoluble at the micromolar level. Each peptide "i" is assigned to a point $\mathbf{X}_i$ in a multiparameter space (Figure 1). With a small input data set, a learning algorithm based on linear programming is superior to least-squares methods.[11] The goal of this algorithm is to find a hyperplane in the parameter space (yellow in Figure 1) that separates peptides into "good" (soluble red peptides) and "bad" (insoluble blue peptides), using the smallest possible number of parameters. The perpendicular vector $\mathbf{W}$ of the separator plane in Figure 1 is a linear combination of the parameters, and represents the best single score S to separate peptides once the separator plane has been optimized.

We tested the following five parameters to discover the best S: *Hydrophobicity* was calculated as the sum of transfer free energies from water to a low dielectric medium for all residues in the peptide, using the scale of Chen.[12] *Amphipathicity* was calculated as the magnitude of the hydrophobic moment for the entire length of the peptide, as described by Eisenberg et al.[13] *Chargepoints* were calculated by summing up the number of charged residues positioned within 3 residues of the N or C terminus, and subtracting the number of charged residues more than three residues from the termini. The maximum hydrophobic stretch length, *StretchMax*, is calculated as the maximum number of consecutive hydrophobic residues in the peptide (including Alanine). The maximum hydrophobicity of any stretch, *HPMax*, is calculated by summing the hydrophobicity values[12] for each stretch of consecutive hydrophobic residues, and then taking the most hydrophobic value among all of the stretches. We divide *Hydrophobicity* and *Amphipathicity* by 10 to bring their magnitude in line with other parameters. Details of the hydrophobicity scales used are described in the Supporting Information, which also has a table of the all the parameter values for all peptides used to train the learning algorithm.

The final score S was used to mine the known transmembrane helices obtained from the online MPtopo database.[7] This database is separated into three categories. The 3D_Helix category is for sequences from proteins with known X-ray crystallography or NMR structures, the 1D_Helix catergory is for sequences that are known to be trans-membrane because of indirect evidence obtained from gene fusion, proteolytic degradation, and amino acid deletion experiments, and the 3D_Other category (not used in this study) is for all sequences from monotopic or beta-barrel trans-membrane proteins.

Results

Initial Training Set. Table 1 lists the initial training set of nine peptides for the learning algorithm. The first five peptides in Table 1 were taken from the literature; the remaining four were synthesized *de novo*. The solubility of membrane binding peptides is sensitive to solvent conditions, so we determined the solubility for all peptides in this study, including previously described peptides, by the same centrifugation procedure outlined in the Experimental and Computational Methods section. Four of the nine peptides in the training set were soluble at the >1 μM level. Msba-1, PRC-1, and SRII-2 were too hydrophobic to survive aqueous purification after synthesis and are assumed to be soluble to less than 1 μM.

All of the soluble peptides in the training set show evidence of vesicle binding. Figure 2 shows fluorescence spectra of the four soluble peptides from the initial training set with and without SUVs. Helix-5 bound only to anionic vesicles, producing a 13 nm blue shift. CCOIII-4, TMX-3, and A_2IA_2 bound strongly to both neutral and anionic SUVs. They produced spectral blue shifts of 14 nm, 22 and 5 nm, respectively. A_2IA_2 and TMX-3 fluorescence intensity changed significantly upon binding. TMX-3 fluorescence more than doubles after binding. It has been shown that the pK_a of the histidine residues in TMX-3 is lowered from 6.8 in solution to 6.0 in the membrane,[2]

TABLE 1: Training Set Peptides and Measured Solubilities[a]

peptide	sequence	solubility
Vpu_{8-29}, from HIV-1 protease[17]	**KKG**$_{\text{IVALVVAIIIAIVVWSIVIIE}$**GKKK**	<1 μM
TMX-1 *de novo*[18]	Ac-WNALAAVAAALAAVAAALAAVAASKSKSKSK-NH_2	<1 μM
helix-5 *de novo*[3]	Ac-NELKKKLELCKAKWLEAKKKLEALK-NH_2	>13 mM
TMX-3 *de novo*[2]	GGWAALAAHLAPALAAALAHALASRSRSRSR-NH_2	14 μM
A_2IA_2 *de novo*[19]	KKAAAIAAAAAIAAWAAIAAAKKKK-NH_2	>13 mM
Msba-1, from transport ATP-binding protein Msba	Ac-$^{\mathbf{H}}_{\mathrm{K}}$AGLVVSTIALVINAAADT$^{\mathbf{W}}_{\mathrm{Y}}$MISLLK$^{\mathbf{S}}_{\mathrm{PL}}$	<1 μM
PRC-1, from photosynthetic reaction center	$^{\mathbf{RS}}_{\mathrm{G}}$AGIAAFAFGSTAILI$^{\mathbf{A}}_{\mathrm{I}}LFN^{\mathbf{W}}_{\mathrm{M}}$$AA_EVH - NH_2$	<1 μM
CCOIII-4, from cytochrome c oxidase polypeptide III	Ac − H$^{\mathbf{A}}_{\mathrm{L}}$PLINTLI$^{\mathbf{A}}_{\mathrm{L}}L^{\mathbf{A}}_{\mathrm{L}}$SGVAVTWAHHAFVKS	16 μM
SRII-2, from sensory rhodopsin II	$^{\mathbf{S}}_{E}$*RR*$^{\mathbf{H}}_{Y}$YVTLVGISGIAAVA$^{\mathbf{W}}_{\mathrm{Y}}$VVMAL^{H}-$NH_2$	<1 μM

[a] Changes or additions to the sequences are shown in bold above the sequence, wild type residues are below. Ac- = acetylation, -NH_2 = amidation. References are given for the five previously studied peptides.

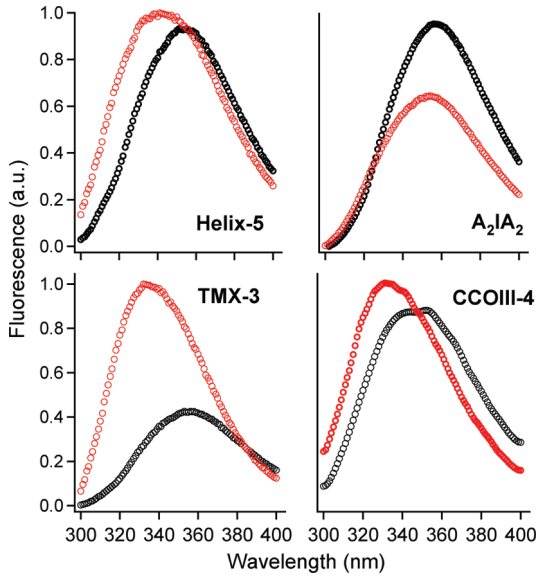

Figure 2. Fluorescence spectra of training set peptides with (red circles) and without (black circles) anionic SUV. Spectra are taken at 5 °C where binding is expected to be maximal. Blue spectral shifts are observed for all peptides in the presence of SUV. In addition, TMX-3 fluorescence is doubled, while A_2IA_2 binding is quenched upon binding. Peptide/vesicle ratios were 72 (Helix-5), 100 (CCOIII-4), 102 (TMX-3), and 36 (A_2IA_2).

and His quenching of tryptophan is more likely in the disordered coil in solution than in the helical membrane-inserted state.

Optimal Classification Score *S*. We were able to obtain complete separation of the initial training set with just two of the original five parameters. The trifluoroethanol hydrophobicity scale and ACFILMVYW hydrophobic residue list turned out to be best for computing hydrophobicity (see Supporting Information).

The only two-parameter solutions were *HPMax - Hydrophobicity* and *HPMax - Amphipathicity*, with the latter performing slightly better. The *Amphipathicity* and *Hydrophobicity* parameters turned out to be strongly correlated. Figure 3A shows the optimal *HPMax - Amphipathicity* solution. The yellow line is the optimal separator plane, and the black arrow indicates the optimal scoring function

$$S_{2D} = 35.4 - \text{Amphipathicity} - 1.13*\text{HPMax} > 0 \quad (1)$$

where $S_{2D} > 0$ indicates solubility above 1 μM. S_{2D} is the single parameter that best separates soluble and insoluble peptides.

Although a perfect solution was obtained with just two parameters, adding *StretchMax* as a third parameter improves the average distance from the separator plane by a factor of 2, more than any other added parameter. Figure 3B shows the three-dimensional (3D) separator plot. The optimal solution with three parameters is shown by the yellow plane, which predicts solubility based on the following inequality:

$$S_{3D} = -\text{Amphipathicity} - 0.845*\text{HPMax} - 0.0124*\text{StretchMax} + 33.8 > 0 \quad (2)$$

Solutions with other parameters (e.g., *Chargepoints*), or more than three parameters yielded no further significant improvement. The Supporting Information lists *S* and parameter values for all peptides discussed in our study.

Selection of the Validation Set. Next we tested the predictivity of eq 1. Figure 4 shows the distribution of *S* values for the complete set of helices from the 1D_Helix and 3D_Helix databases. All sequences that satisfy eq 2 also satisfy eq 1, and these are shown in red in Figure 4. Ninety-two out of 441 helices in the 3D_Helix database and 80 out of 261 helices in the 1D_Helix database are potential candidates for membrane-binding soluble helices.

We applied three additional criteria to narrow the choice. Helices shorter than 21 residues were eliminated, as were helices with $S < 31$ because only A_2IA_2 ($S = 31$) and Helix-5 ($S = 115$) in Table 1 were soluble at the highly desirable millimolar level. We also eliminated peptides lacking a tyrosine or tryptophan fluorophore. We chose for synthesis the two peptides from the 1D_Helix database with the largest *S* values, TetA-4 and ArsB-11.

Applying the same criteria to the 3D_Helix database yielded seven peptides that were further inspected structurally. Four helices lacked any salt bridges and appeared to require few contacts with the rest of the protein. The peptide that looked best from this visual inspection, G3PT-11, and the one that looked the worst, SecY-10, were chosen for synthesis to determine whether such visual inspection is worthwhile.

The validation set peptides are summarized in Table 2. Before synthesis, minor changes were made to the sequences to improve solubility and/or fluorescence detection. These are shown in bold in Table 2, above the replaced residues. Figure 3 plots the synthesized validation set as green triangles, while the actual synthesized sequences are shown as closed triangles.

Characterization of the Validation Set. All four of the validation set peptides proved to be soluble in the 25 μM to 2

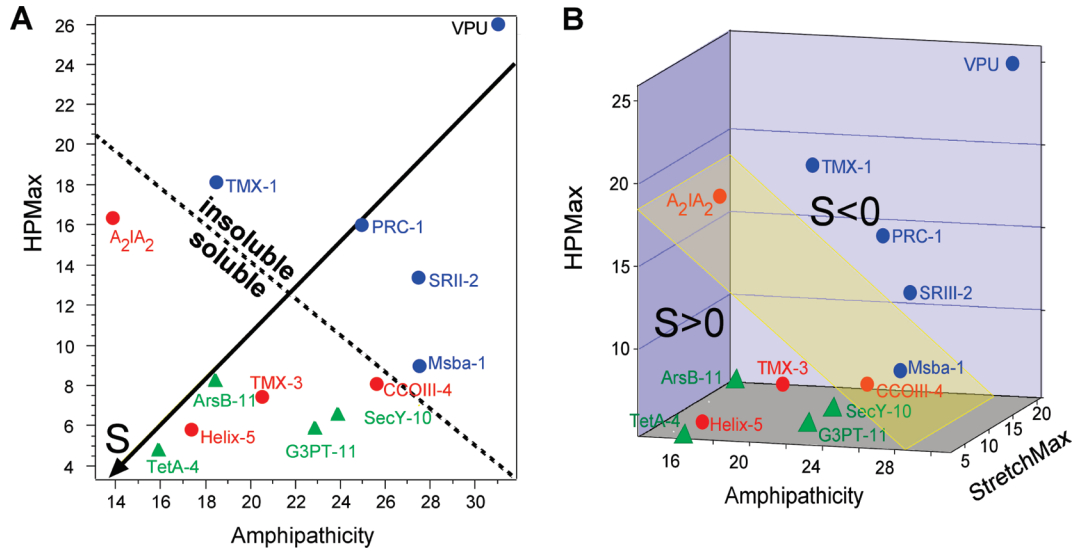

Figure 3. (A) Optimal two-dimensional (2D) solution using *Amphipathicity* and *HPMax*. Blue peptides: insoluble training set; red peptides: soluble training set; green peptides: validation set synthesized to be soluble. The dotted line is the optimal separator of soluble and insoluble peptides. The arrow is the optimal coordinate *S* (linear combination of *Amphipathicity* and *HPMax*) to classify peptides as soluble or insoluble (eq 1). (B) Optimal three-dimensional (3D) solution. Adding *StretchMax* as the 3rd parameter doubles the average distance to the separator plane (shown in yellow). All *S* values are listed in tables in the Supporting Information. All solutions were calculated using the trifluoroethanol hydrophobicity scale, and the hydrophobic residue list ACFILMVYW.

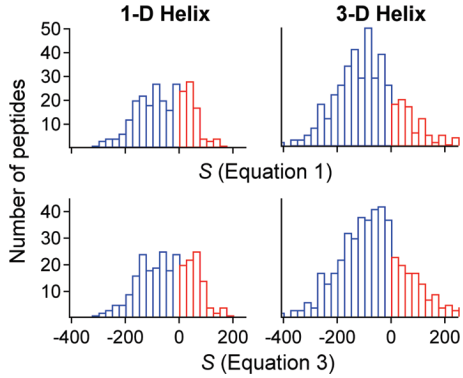

Figure 4. Distribution of scores, calculated according to the two-dimensional solutions given in eqs 1 and 3 for sequences in the 3D_Helix and 1D_Helix databases. Peptides with positive scores, shown in red, are expected to be soluble, while peptides with zero or negative scores, shown in blue, are expected to be insoluble.

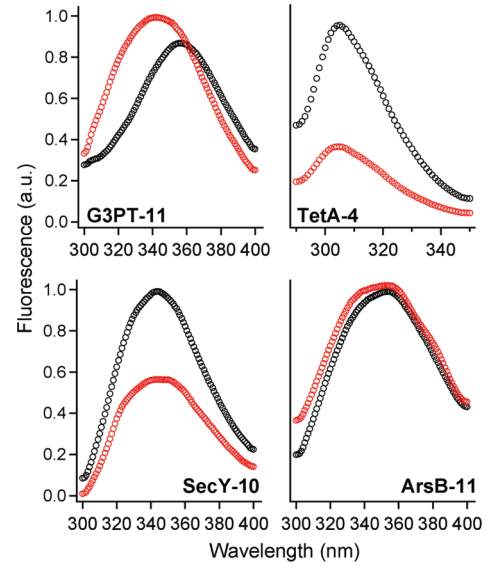

Figure 5. Fluorescence spectra of validation set peptides with (red circles) and without (black circles) anionic SUV. Spectra are taken at 5 °C. A large blue spectral shift is observed only for G3PT-11, although significant quenching is observed for both TetA-4 and SecY-10 in the presence of SUVs. TetA-4 contains only tyrosine as a reporter, hence the blue-shifted fluorescence compared to tryptophan-containing peptides. No spectral or intensity shifts are observed for ArsB-11 with SUVs at peptide−vesicle ratios as low as 12.

mM range. The observed solubility limits are listed in Table 2. Figure 5 shows fluorescence spectra for all peptides in the validation set before and after addition of anionic SUV. A large blue shift of 13 nm was observed for G3PT-11. Strong fluorescence quenching was observed for TetA-4 (62%) and SecY-10 (44%) upon the addition of anionic vesicle. TetA-4 contained only a tyrosine reporter. For ArsB-11, no evidence of binding was observed in either neutral or anionic SUVs. This

TABLE 2: Validation Set Peptides, along with Their Measured Solubilities[a]

name	sequence	solubility
ArsB-11, arsenical pump membrane protein	Ac-STATGVIKEAMI$^{W}_{Y}$ANVIGCDL	25 μM
SecY-10, preprotein translocase secy subunit	Ac-GTGVLLTVSIVYR$^{W}_{M}$YEQLLREKVSE$^{G}_{L}$-NH_2	25 μM
G3PT-11, glycerol 3 phosphate transporter	Ac-AAGTAAGFTGL$^{W}_{F}$GYLGGSVAASAIVGYTVD-NH_2	2 mM
TetA-4, tetracycline resistance protein	Ac-^{GT}IVAGITGATGAVAGAYIADIT-NH_2	600 μM

[a] Additions or mutations to the native sequences are shown in bold above the substituted residues. Ac- = acetylation, -NH_2 = amidation.

peptide came from the 1D_helix database, where no structural check could be made; it may not be sufficiently lipid exposed. Both of the peptides from the 3D_Helix database were water-soluble and showed fluorescence shifts indicative of vesicle binding.

Discussion

Application of the learning algorithm allows us to propose several trans-membrane helices from membrane proteins for future thermodynamic and kinetic studies of their membrane binding and helix−coil equilibria. We have shown that a simple score S based only on peptide sequence properties can be used to mine membrane protein databases and select soluble helical peptides suitable for membrane binding studies. Out of 702 helices, we applied S together with length and structural criteria (when available) to select a validation set of four sequences. Success, defined by solubility upon repeated centrifugation and vesicle binding detected by fluorescence, was achieved for three out of four peptides. In addition, CCOIII-4 from the training set is also a novel peptide. The resulting four successful peptides are CCOIII-4 (Table 1), SecY-10, G3PT-11, and TetA-4 (Table 2).

If we incorporate the four validation set peptides into the training set of our learning algorithm, we get an updated condition for solubility analogous to eq 1:

$$S = 37.8 - \text{Amphipathicity} - 1.21\text{HPMax} > 0 \quad (3)$$

This updated solution can be applied again to the 1D_Helix and 3D_Helix databases. Figure 4 shows the distribution of S values for both databases using the improved solution. Ninety-two and 103 peptides from the 1D_Helix and 3D_Helix databases, respectively, satisfy eq 3. As a final check, we tested four additional peptides from the literature that bind to membranes and are water-soluble, but were not included in our training set: three out of four were predicted to be soluble (see Supporting Information).

To be suitable for binding equilibrium studies, a peptide must satisfy two requirements. A detectable spectroscopic signature must change upon binding (verified above), and its binding affinity must be tunable by pH,[2] temperature,[14,15] or denaturing agent.[16] We illustrate the latter for one of the peptides, CCOIII-4, in Figure 6. The peptide was temperature-scanned in aqueous buffer and in the presence of vesicles. SVD[10] reveals the opposite wavelength shift in the presence versus absence of vesicles: at high temperatures, the fluorescence spectrum in vesicles approaches the spectrum in pure buffer, showing that the peptide unbinds from the membrane.

We conclude by predicting four additional soluble vesicle-binding peptides. The peptides in Table 3 are all from the 3D_Helix database and satisfy eq 3. They passed the structural inspection described in the results section. We have some evidence that structural inspection to eliminate salt bridges and favor lipid exposure is worthwhile. The aqueous solubility and fluorescence shift observed for G3PT-11, which passed the inspection, was significantly better than that for SecY-10, which did not pass. Furthermore, one of two peptides from the 1D_Helix database, where no structural inspection could be made, failed to bind to vesicles altogether. We believe that the additional sequences in Table 3, with minor modification to improve fluorescence detection (e.g., F,Y→W substitution), are likely to yield additional peptides for binding kinetics and thermodynamics experiments.

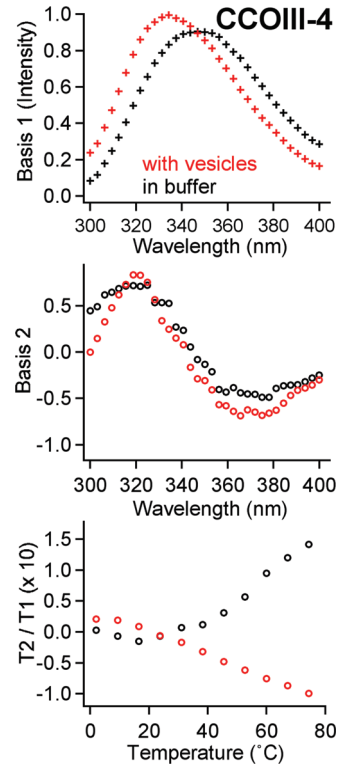

Figure 6. SVD analysis of the thermal titration of CCOIII-4. Top: First SVD basis function corresponds to average fluorescence spectra with (red) and without (black) anionic SUVs. Middle: Basis function 2 has a "derivative" shape that tracks wavelength shift and broadening with temperature. Bottom: The temperature dependence of basis function 2 reveals spectral shifts that go in opposite directions in the presence or absence of SUVs.

TABLE 3: List of Additional Sequences from the 3D_Helix Database, Which Satisfy All Scores Used in Selecting the Validation Set for This Study[a]

protein (PDB ID)	peptide sequence	S
lactose permease (1PV6)	GEQGTRVFGYVTTMGELLNASIM	13.1
cytochrome c oxidase polypeptide III (1OCC)	VPLLNTSVLLASGVSITWAHHSLM	4.4
glycerol-3-phosphate transporter (1PW4)	DLGFALSGISIAYGFSKFIMGSVSD	4.9
ADP, ATP carrier protein heart isoform T1 (1OKC)	FNVSVQGIIIYRAAYFGVYDTAKG	6.6

[a] The S scores were calculated using eq 3.

Conclusions

Four new water-soluble and membrane binding peptides have been identified by a learning algorithm optimized to mine helical transmembrane protein databases. These peptides are the first to be based on natural transmembrane proteins, rather than on designed or cytotoxic sequences. The parameter S characterizes the peptides as soluble if $S > 0$. S can be computed easily from the sequence of a peptide. We suggest four additional peptides for independent verification. Peptides discovered using S are not guaranteed to be monomeric in solution, or insert into the bilayer (rather than just bind), but the algorithm greatly reduces the number of sequences for characterization in greater detail.

Acknowledgment. This work was supported by a grant from the National Science Foundation, MCB 0613643. SD was a NSF

GK-12 fellow during part of the period when this work was carried out. MG thanks the Eiszner family for support through the James R. Eiszner chair.

Supporting Information Available: Transmission electron micrographs of vesicles before and after ultracentrifugation, as well as additional details regarding vesicle preparation, hydrophobicity scales, and algorithm implementation. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Tucker, M. J.; Tang, J.; Gai, F. *J. Phys. Chem. B* **2006**, *110*, 8105.
(2) Ladokhin, A. S.; White, S. H. *Biochemistry* **2004**, *43*, 5782.
(3) Russel, C. J.; King, D. S.; Thorgeirsson, T. E.; Shin, Y. K. *Protein Eng.* **1998**, *11*, 539.
(4) Tang, J.; Yin, H.; Qiu, J.; Tucker, M. J.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2009**, *131*, 3816.
(5) Constantinescu, I.; Lafleur, M. *Biochim. Biophys. Acta* **2004**, *1667*, 26.
(6) Tang, J.; Signarvic, R. S.; DeGrado, W. F.; Gai, F. *Biochemistry* **2007**, *46*, 13856.
(7) Jayasinghe, S.; Hristova, K.; White, S. H. *Protein Sci.* **2001**, *10*, 455.
(8) Gill, S. C.; von Hippel, P. H. *Anal. Biochem.* **1989**, *182*, 319.
(9) Gullingsrud, J.; Schulten, K. *Biophys. J.* **2004**, *86*, 3496.
(10) Henry, E. R.; Hofrichter, J. *Methods Enzymol.* **1992**, *210*, 129.
(11) Rojas, R. *Neural Networks, A Systematic Introduction*; Springer-Verlag: Berlin, 1996.
(12) Chen, Y.; Mant, C. T.; Farmer, S. W.; Hancock, R. E. W.; Vasil, M. L.; Hodges, R. S. *J. Biol. Chem.* **2005**, *280*, 12316.
(13) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. *J. Mol. Biol.* **1984**, *179*, 125.
(14) Meijberg, J.; Booth, P. J. *J. Mol. Biol.* **2002**, *319*, 839.
(15) Tang, J.; Gai, F. *Biochemistry* **2008**, *47*, 8250.
(16) Curnow, P.; Booth, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18970.
(17) Park, S. H.; Mrse, A. A.; Nevzorov, A. A.; Mesleh, M. F.; Oblatt-Montal, M.; Montal, M.; Opella, S. J. *J. Mol. Biol.* **2003**, *333*, 409.
(18) Wimley, W. C.; White, S. H. *Biochemistry* **2000**, *39*, 4432.
(19) Liu, L.-P.; Deber, C. M. *Biochemistry* **1997**, *36*, 5476.

JP912157G