Genome-wide Structural Analysis Reveals Novel Membrane Binding Properties of AP180 N-terminal Homology (ANTH) Domains

Antonina Silkov †, Youngdae Yoon ‡, Hunjoong Lee †, Nikhil Gokhale ‡, Emmanuel Adu-Gyamfi ‡, Robert V. Stahelin †, Wohna Cho ‡,* †, and Diana Murray ‡,* 2

From the † Department of Pharmacology, Columbia University, New York, New York 11032, the ‡ Department of Chemistry, University of Illinois, Chicago, Illinois 60607, the † Department of Biochemistry, University of Notre Dame, South Bend, Indiana 46617, the ‡ Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, South Bend, Indiana 46617, and the ** Division of Integrative Biosciences and Biotechnology, Pohang University of Science and Technology, Pohang 790-784, Korea

Background: There is a great need for a high throughput computational tool for predicting the function of lipid binding domains on a genomic scale.

Results: A newly developed computation protocol allows for genome-wide prediction of membrane binding properties of ANTH domains.

Conclusion: Membrane binding properties of proteins can be systematically and reliably predicted by our combinatorial approach.

Significance: A novel functionality of lipid binding domains can be computationally predicted.

An increasing number of cytosolic proteins are shown to interact with membrane lipids during diverse cellular processes, but computational prediction of these proteins and their membrane binding behaviors remains challenging. Here, we introduce a new combinatorial computation protocol for systematic and robust functional prediction of membrane-binding proteins through high throughput homology modeling and in-depth calculation of biophysical properties. The approach was applied to the genomic scale identification of the AP180 N-terminal homology (ANTH) domain, one of the modular lipid binding domains, and prediction of their membrane binding properties. Our analysis yielded comprehensive coverage of the ANTH domain family and allowed classification and functional annotation of proteins based on the differences in local structural and biophysical features. Our analysis also identified a group of plant ANTH domains with unique structural features that may confer novel functionalities. Experimental characterization of a representative member of this subfamily confirmed its unique membrane binding mechanism and unprecedented membrane deforming activity. Collectively, these studies suggest that our new computational approach can be applied to genome-wide functional prediction of other lipid binding domains.

Numerous proteins interact with cellular membranes during cell signaling, membrane tracking, and many other cellular processes. A majority of cellular proteins that interact with membrane lipids contain one or more of the specialized modular domains known as lipid binding domains (1, 2), such as pleckstrin homology, C1, C2, FYVE, and PX domains, or lipid-binding motifs, such as polybasic motif (3). Because of the prevalence and importance of lipid-mediated cellular processes and the key roles that lipid binding domains play in these processes, the structure and the function of representative members of these ubiquitous domain families have been extensively studied (1, 2, 4). In this post-genomic era, it is imperative to develop effective high throughput computational and experimental tools to predict and characterize the function of lipid binding domains on a genomic scale. Although computational prediction of lipid binding domains (and proteins) has been reported (5–10), robust genome-wide prediction of their membrane binding properties remains challenging because their membrane binding properties vary widely even within a family with high structural similarity. We therefore developed a new combinatorial computational approach by incorporating systematic analysis of biophysical properties of domains into an automated high throughput homology modeling pipeline tool, SkyLine, which allows rapid and reliable structure-based identification and classification of protein families (11). This combinatorial approach not only improves the functional classification of proteins but also facilitates the discovery of new and novel membrane binding activity. In this report, we describe the application of the approach to
the genome-wide identification and functional prediction of ANTH\(^3\) (AP180 N-terminal homology) domains.

The ANTH domain is a lipid binding domain found in proteins, such as AP180, CALM, and HIP1, that are involved in membrane tracking, clathrin-mediated endocytosis in particular (12–15). These proteins are also implicated in the development of a wide range of pathological processes, including Alzheimer disease, leukemia, and other human cancers (16–20). A prerequisite for their function in endocytosis is their interaction with phosphorylated derivatives of phosphatidylinositol collectively known as phosphoinositides, most notably phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P\(_2\)) enriched in the cytosolic leaflet of the plasma membrane (15, 21, 22). The ANTH domain shares function, sequence, and structural similarity with the ENTH (Epsin N-terminal homology) domain that is also found in proteins involved in endocytosis. Both domains interact directly with PtdIns(4,5)P\(_2\) and share a common helical substructure (Fig. 1, A and B). However, the ANTH domain has several additional C-terminal helices, whereas the ENTH domain contains an extra N-terminal helix (called \(H_0\); Fig. 1B). The amphiphilic \(H_0\) plays a critical role in the membrane penetration and deformation by the ENTH domain (23, 24), and it has not been found in the ANTH domain that does not exhibit these activities. In addition, co-crystal structures with the soluble headgroup of PtdIns(4,5)P\(_2\) reveal that although both ANTH and ENTH domains use a conserved cluster of basic residues to coordinate the headgroup of PtdIns(4,5)P\(_2\), the molecular locations of the lipid-binding sites are distinct (Fig. 1, C and D) (21, 23). Also, the ANTH domain binds PtdIns(4,5)P\(_2\) superficially, through the coordination of its headgroup by the side chains of basic residues (Fig. 1A), whereas the ENTH domain binds PtdIns(4,5)P\(_2\) within a cleft with both basic and hydrophobic residues contributing to membrane association (Fig. 1B). It is noteworthy that these comparisons are derived from the studies of a limited number (i.e. 1–3) of representatives and that other domains in each family may have different structural properties that confer variable functionalities, e.g. ANTH domains with ENTH domain-like activity. Collectively, ANTH and ENTH domains serve as an excellent model system to test the utility of our combinational computational approach in predicting functional differences among lipid binding domains based on differences in their structures and biophysical properties. We thus performed the genome-wide identification of ANTH domains and searched for novel functionality through structure-based functional classification and prediction, which was then experimentally tested. Our approach identified a novel group of plant ANTH domains with unique ENTH domain-like structure and function.

**EXPERIMENTAL PROCEDURES**

**Materials**—1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoserine (POPS), and inositol-1,4,5-trisphosphate (Ins(1,4,5)P\(_3\)) were from Avanti Polar Lipids, and a 1,2-dipalmitoyl derivative of PtdIns(4,5)P\(_2\), phosphatidylinositol 3,4-bisphosphate, and phosphatidylinositol 3,4,5-trisphosphate (PtdIns(3,4,5)P\(_3\)) were from Cayman Chemical.

**Pipeline Runs**—Automatically generated models of ANTH domains are available from the author. All ANTH structures available in PDB were used for high throughput homology modeling as follows: 1HF8, 1HFA, 1HG2, 1HG5, and 1HX8. The pipeline was run against the NCBI nonredundant peptide sequence data base (NRdb, www.ncbi.nlm.nih.gov, as of September, 2008). Each run implemented five rounds of Psi-blast. The threshold e-value for including a sequence into PSSM was set to 5 × 10\(^{-4}\) and for reporting a sequence to 10. Only sequences longer than 50 amino acids are retained. The redundant Psi-blast hits from the same species were reduced based on 100% sequence identity. The pipeline uses the program Modeller (25) for creating models of the retrieved sequences. The quality of the model was assessed in terms of pG score, the posterior probability that the model has a correct three-dimensional conformation, given its normalized z-score (obtained using the program Prosa II, (26)) and length. In our previous large scale analysis of the models produced by the pipeline, we used a pG threshold of 0.7 to consider a model “reliable”(11). Here, we opted for a more stringent value of pG = 0.85. The actual minimum pG score of our model set is 0.87. In the optimal case, the model should be aligned with the whole structure of the ANTH domain. However, taking into account that some sequences might be incomplete for technical reasons, we set to determine the minimal meaningful length of an ANTH model.
Because PtdIns(4,5)P$_2$-binding site is located at the N terminus of the ANTH domain, we set a requirement for the model to “cover” this site of the template structure (starting at residue 28 in 1HG2). To determine the ending point of created models in relation to the template structures, we truncated the PDB structures from C terminus by one amino acid residue and plotted the length of the remaining structures versus Verify3D score. In all structures tested, truncation beyond the 30th amino acid residue from the C terminus (residue 250 in 1HG2) is associated with a sharp drop of Verify3D score (supplemental Fig. S1). Overall, we considered the models with structural alignments extended from residue 28 to 250 of 1HG2 or to structurally equivalent residues of other templates. Typically, 30% sequence identity is used as the cutoff, above which a homology model is viewed as meaningful. This is because the template-target alignments at the lower sequence identity are often incorrect. However, our approach of evaluating the model quality allows us to filter out the unreliable models and to keep the reliable models based on the correct alignments, even if the sequence identity is lower than 30%.

Construction and Evaluation of the Composite Model of AAL11587—The composite model of AAL11587 was built with the Nest program (27), with the options -fast 1 -opt 2. The template for the N terminus, including $H_\sigma$ was the crystal structure of ENTH domain of rat epsin 1 (PDB entry 1HOA). The rest of the model was built using the crystal structure of ANTH domain of rat CALM protein (PDB accession number 1HFA). The templates were structurally aligned using Ska program (28). The side chains of the model were optimized using Plop program (29). The model was evaluated with statistical potential-based pG score (see Pipeline runs) and structure verification program Verify3D, which assesses the compatibility of the segments of the sequence with their three-dimensional structures by plotting the average statistical preference scores in a window of 21 residues (30), pG score for the model is 1.00. The minimum value of Verify3D profile window plot is 0.05.

Docking of the Lipid Headgroup—Docking of Ins(4,5)P$_2$ to the model of AAL11587 was performed with the package DOCK 6. The spheres representing the binding site were selected within 5 Å from the functional groups of the side chains participating in the formation of the basic patches. The ligand was allowed to be flexible during the orientation step.

Determining Orientation of the Model of AAL11587 Relative to the Membrane—Phosphatidylcholine (PC) bilayer was built as described previously (31). The principal moments of inertia of the atoms of the protein were calculated, and the protein was rotated so that the moments of inertia were aligned along the z axis of Cartesian axes system using Pdbinertia program (Dr. Arthur G. Palmer’s group, Columbia University). To search the most favorable protein orientation toward the lipid bilayers, each configuration was determined by rotating the protein using Euler angles. First, the protein was rotated by an angle $\theta$ by increments of 12° within a range of (0, 60°) about the x axis. Second, the protein was rotated by an angle $\psi$ by increments of 30° within a range of (0, 360°) about the z’ axis. The rotated protein was placed on the top of the membrane so that two objects with van der Waals radii included are 0.01 Å apart. The same procedure was repeated with the flipped protein. The atomic contact energy of each configuration is calculated using DOT 2.0 program (32), with the distance cutoff of 6 Å. Because the atomic contact energy had been originally derived from proteins, we used the backbone values of proteins for the PC bilayers. The most favorable orientation was chosen based on the atomic contact energy.

Protein Expression and Purification—The ANTH domain (residue 1–310) of AT2G01600 was subcloned into pGEX4T-1 vector. The ΔN mutant lacking residues 1–29 and the K38A/R50A/H51A mutant were generated by PCR mutagenesis. All plasmids were transformed into Escherichia coli BL21 (DE3) RIL cells for protein expression. The bacterially expressed protein was loaded onto a glutathione-agarose column, and after washing it was digested by thrombin and eluted from the column. Protein concentration was then determined using the bicinchoninic acid method (Pierce).

Surface Plasmon Resonance (SPR) Analysis—All SPR measurements were performed at 23 °C in 20 mM Tris-HCl, pH 7.4, containing 0.16 M NaCl using a lipid-coated L1 chip in the Bia-core X system as described previously (33). POPC/POPS/phosphoinositide (77:20:3) vesicles and POPC vesicles were coated onto the active surface and the control surface, respectively. Equilibrium SPR measurements were done at the flow rate of 5 μl/min to allow sufficient time for the response in resonance unit of the association phase to reach near-equilibrium values ($R_{eq}$). Assuming a Langmuir-type binding between the protein (P) and protein-binding sites (M) on vesicles (i.e. $P + M \leftrightarrow PM$) (34), $R_{eq}$ values were then plotted versus $P_{eq}$, and the $K_d$ value was determined by a nonlinear least squares analysis of the binding isotherm using the equation $R_{eq} = R_{max}/(1 + K_d/P_{eq})$ (34). Each data set was repeated three or more times to calculate average and standard deviation values. For kinetic SPR measurements, the flow rate was maintained at 15 μl/min for both association and dissociation phases.

Monolayer Penetration Assay—Surface pressure of solution was measured using a Wilhelmy plate attached to a computer-controlled Cahn electrobalance (model C-32) as described previously (35). Five to ten μl of lipid solution (POPC/PtdIns(3,4,5)P$_3$ = 97:3) in ethanol/hexane (1:9 (v/v)) was spread onto 10 ml of subphase (20 mM HEPE, pH 7.4, containing 0.16 M KCl) to form a monolayer with a given initial surface pressure. The protein solution (typically 40 μl) was injected into the subphase, and the change in surface pressure was measured as a function of time. The critical surface pressure was determined by extrapolating the change in surface pressure versus initial surface pressure plot to the x axis (34).

Ins(1,4,5)P$_3$ Binding Assay—Binding of AAL11587 ANTH domain WT and mutants to Ins(1,4,5)P$_3$ was performed using radiolabeled Ins(1,4,5)P$_3$ (PerkinElmer Life Sciences). For this assay, C-terminal His$_6$-tagged proteins were generated. Radiolabeled and unlabeled Ins(1,4,5)P$_3$ were mixed to a final concentration of 10 μM in 20 mM Tris-HCl, pH 7.4, with 160 mM NaCl. 0–10 μM of Ins(1,4,5)P$_3$ was mixed with 0.5 μM protein, and the mixture was incubated at 25 °C for 30 min. A His Spin-Trap column (GE Healthcare) was pre-equilibrated with 20 mM Tris-HCl, pH 7.4, with 160 mM NaCl, and the reaction mixture was added to the column and incubated for 5 min. The column was washed three times by 600 μl of the same buffer, and the
Proteomic Study of ANTH Domains

protein was eluted using 600 µl of 20 mM Tris-HCl, pH 7.4, with 160 mM NaCl and 300 mM imidazole. 200 µl of the eluted solution was taken for radioactivity measurement using the Beckman LS6500 liquid scintillation counter. Each binding isotherm was fit by nonlinear least squares analysis using the following equation: \[ [\text{Ins}(1,4,5)P_3] = n \cdot P_{\text{in}} / (1 + K_d / [\text{Ins}(1,4,5)P_3]) \] where \( n, P_{\text{in}} \), and \( K_d \) indicate binding stoichiometry, total protein concentration, protein-bound \( \text{Ins}(1,4,5)P_3 \) concentration, and total \( \text{Ins}(1,4,5)P_3 \) concentration, respectively.

Vesicle Tubulation Assay—Large unilamellar vesicle samples (1 mg/ml) prepared by extrusion through 100 nm membranes in POPC/POPE/PtdIns(3,4,5)P_3 (75:20:5) were incubated in the presence or absence of protein (0.5–2 µM) in low salt buffer for 15 min at 25 °C. The sample (8 µl) was applied to a carbon-Formvar-coated copper grid and incubated for 2 min. Excess liquid was carefully removed by using a wet absorbent tissue. The grids were incubated three times with 2% filtered uranyl acetate solution for 10 s and dried by air and then under a heat lamp. Negative staining was performed at 25 °C. Membrane morphologies were examined on an FEI Magellan scanning electron microscope with the electron energy set to 15,000 V. Representative images were taken with a direct magnification of ×42,000–202,000.

RESULTS

Genome-wide Identification of the ANTH Domain Family—To identify all members of the ANTH family, we used SkyLine (11) that allowed us to collect all sequences that have similarity to ANTH domains, to build the models of these sequences, and to assess the quality of the models. The high quality of the models corresponds to the high probability of the corresponding sequences to adopt an ANTH fold. We retrieved 246 models of nonredundant sequences (supplemental Table S2) that passed quality and length criteria (see under “Experimental Procedures” for model quality evaluation). We compared our set of annotated domains with the collection of ANTH domain family sequences retrieved by PFAM, the most comprehensive domain family resource based on sequence data (36). As seen from supplemental Fig. S2, results produced by both approaches largely overlap. Fifty five ANTH domains present in PFAM and absent from our collection were picked up by SkyLine sequence search but did not pass model quality and length criteria. Twelve proteins from our dataset, which are identified as “unique” proteins using UniProt, were not picked up by PFAM (supplemental Table S2 and supplemental Fig. S2). In this subset, seven sequences cannot be annotated as ANTH by any other methods tested, including CDD (37) and InterProScan (38). Forty five other sequences in our dataset were not identical to any of the PFAM hits, but because they are absent from UniProtKB, we could not use the latter resource to determine their uniqueness (supplemental Table S3 and supplemental Fig. S2). Two sequences from this subset are not annotated as ANTH by any publicly available resource tested. These two sequences, along with seven novel sequences from the subset of unique proteins, are shown in red in supplemental Fig. S3. Overall, our method added 13 species to 74 species represented in the PFAM ANTH dataset.

Structure-based Classification of Identified ANTH Domains—Membrane binding of lipid binding domains (and proteins) typically involves the combination of specific lipid recognition and the nonspecific contact, although the former is not essential for all membrane-protein interactions (1). Nonspecific interactions include electrostatic interactions between cationic patches on the protein and the anionic membrane surface (39) and partial membrane penetration of hydrophobic protein residues to the hydrophobic interior of the lipid bilayer (1). Specific phosphoinositide binding can take place in a well defined groove, as seen with pleckstrin homology domains (40), or at an unstructured polybasic motif serving as an inducible-binding site, as seen with small G proteins (3). Each of these interactions requires characteristic structural features of the protein. To systematically predict the membrane binding properties of identified ANTH domains, we therefore classified them according to key structural features important for membrane interactions. These features include the presence of a signature PtdIns(4,5)_2-binding motif (K(X)_nK(X)(R)(H)/Y) found in the crystal structure of the ANTH domain of CALM (21), the surface electrostatic properties, and the presence of an amphiphilic a-helix or exposed hydrophobic residues that may participate in partial membrane insertion.

Out of 246 sequences in our collection, 162 sequences bear the signature motif of the archetypal ANTH domain, including one previously nonannotated ANTH protein from Entamoeba histolytica (XP_651286, supplemental Fig. S3). As expected, all models with the “classic” PtdIns(4,5)_2-binding motif exhibit a basic patch similar to that in the crystal structure of CALM (“Classic” ANTH; Fig. 2A). Interestingly, in 44 models the basic patch is enlarged due to the presence of an extra basic amino acid residue adjacent to the “classic” patch (“Enhanced” classic ANTH; Fig. 2B). The “enhanced” ANTH domains are present only in HIP1/1HIP1R/Sla2 subfamily and in plant ANTH domains (supplemental Fig. S3). Furthermore, nine ANTH domains exhibit an even larger patch of strong positive electrostatic potential because of an additional basic amino acid residue (“Super-enhanced” classic ANTH; Fig. 2C). We predict that these three classes of ANTH domains would all bind phosphoinositide-containing membranes in similar orientations using surface basic residues. Also, super-enhanced and enhanced ANTH domains would bind phosphoinositide-containing membranes more tightly than do classic ANTH domains through enhanced electrostatic interactions.

All ANTH domains from the group of yeast proteins that includes Yap1801 and Yap1802 (supplemental Fig. S3) exhibit an area of strong positive potential on the surface that is predicted to face the membrane, although most of them lack the signature PtdIns(4,5)_2-binding motif. The positive patch containing five neighboring lysines (Fig. 2D and supplemental Fig. S4A) is “shifted” compared with the three classic patches. Interestingly, a group of plant ANTH domains (38 models, 37 proteins) exhibits a basic patch in entirely atypical location (“Atypical” ANTH; Fig. 2F). The “atypical” patch on the surface of the ANTH domain (Fig. 2F) lies in the region equivalent to the PtdIns(4,5)_2-binding site of ENTH domain (Fig. 2E).

The most unique group of ANTH domains is the one that possesses a combination of “enhanced” and “atypical” basic
The ENTH domain that are critical for membrane insertion (23) are conserved in these N-terminal sequences (supplemental Fig. S4B). Thus, they may have an unprecedented ENTH domain-like activity to penetrate and deform the membrane. In addition to dual-patch ANTH domains, 10 more plant models (nine Uniprot proteins) (supplemental Fig. S5) have N-terminal sequences with analogous properties, suggesting that ANTH domains with an N-terminal α-helix (“N-ANTH” domains hereafter) are relatively common in plants and may be involved in unique physiological functions of plants.

**Prediction of Membrane Binding Properties of the N-ANTH Domain**—To further investigate the membrane and lipid binding properties of the dual-patch N-ANTH domain, we created a high quality composite model for a representative member, AAL11587 (CAP8_ARATH). In our composite model of AAL11587, the N-terminal region forms a “H₀”-like amphipathic α-helix with conserved hydrophobic residues lying on the outer surface (Fig. 3), suggesting that they may penetrate the membrane and induce membrane curvature and deformation. The addition of H₀ to the model of AAL11587 transforms two basic patches into two basic clefts, ANTH-like (Fig. 3A) and ENTH-like (Fig. 3B) clefts, respectively. In silico docking of two molecules of Ins(1,4,5)P₃ into each of two sites resulted in the complex with numerous hydrogen bonds between the protein and the ligands (Fig. 3, C and D), suggesting that the ANTH domain may be able to bind two phosphoinositide molecules. The orientations of the docked Ins(1,4,5)P₃ in both ANTH-like and ENTH like sites of AAL11587 are different from the orientations of this ligand in the corresponding co-crystal structures of CALM (1hfa) and epsin (1h0a) (root mean square deviation is 6.8 Å for ANTH-like site and 6.0 Å for ENTH-like site; see supplemental Fig. S6). This difference is caused by the fact that AAL11587 has a slightly different arrangement of basic residues and hence a different pattern of hydrogen bonds with a ligand, as compared with CALM and epsin (supplemental Fig. S4B).

Mutation studies on hydrophobic residues on the outer surface of H₀ of the epsin ENTH domain support the model in which H₀ faces the membrane with the basic cleft adjacent to the H₀ interacting with a membrane-embedded phosphoinositide (23, 24, 42). The presence of an amphipathic α-helix and the location of two putative lipid binding clefts next to the H₀ in the composite model of the N-ANTH domain suggest that its membrane binding orientation may resemble that of the ENTH domain. However, the overall geometry of the model of AAL11587 differs significantly from that of the ENTH domain. During the protein-ligand interactions, the gross geometry of a protein affects the spatial distribution of collisions between the reactants (43) and, consequently, their mutual orientation in the complex. Also, the composition of the H₀ in AAL11587 and the epsin ENTH domain is different, meaning that the solvation free energies of interactions of these two proteins with the membrane might also differ. To see if the AAL11587 N-ANTH domain can have the same membrane binding orientation as the epsin ENTH domain despite different desolvation energy and molecular geometry, we determined its orientation relative to the PC membrane by calculating effective atomic contact energies, the desolvation free energies required to transfer atoms from water to a nonpolar interior, using the methods by
Miyazawa and Jernigan (44) and DeLisi and co-workers (45). As seen in Fig. 3, C and D, in the optimal orientation of the model of AAL11587, H₀ is positioned next to the PC bilayer, allowing the penetration of the two conserved hydrophobic residues on the outer surface into the membrane.

Experimental Verification of Functional Properties of the N-ANTH Domain—To test if the N-ANTH domain of AAL11587 has membrane/lipid binding properties predicted by our computational analysis, we performed a series of biophysical measurements on the wild type (WT), a mutant lacking the N-terminal region (/H9004), and a triple-site mutant, K38A/R50A/H51A, in the ANTH-like-binding site. We first measured their membrane binding by SPR analysis. Fig. 4 shows that the N-ANTH domain WT requires the presence of phosphoinositides for membrane binding. Intriguingly, it has modest selectivity for PtdIns(3,4,5)P₃ over PtdIns(4,5)P₂ and phosphatidylinositol 3,4-bisphosphate unlike the epsin ENTH domain that prefers PtdIns(4,5)P₂ to PtdIns(3,4,5)P₃ (24). This difference is consistent with our structure-based prediction that the N-ANTH domain has a different pattern of hydrogen bonds with a ligand, as compared with CALM-ANTH and epsin1-ENTH. Also, the N-ANTH domain has significantly higher membrane affinity than the AP180 ANTH domain (24) when assayed with the same vesicles (see Fig. 4, B and C, and Table 1). Under the same condition, both /H9004 and K38A/R50A/H51A show 10% of membrane affinity of the WT (Fig. 4 D). This underscores the critical role of the putative H₀ and the ANTH-like site in membrane binding. It also indicates that the ENTH-like site has much lower membrane lipid affinity than the ANTH-like site in the intact protein. For this reason, we focused on characterizing the ANTH-like site mutant in ensuing studies.

We then measured the binding of soluble Ins(1,4,5)P₃ to the N-ANTH domain WT and mutants. In agreement with the SPR data, K38A/R50A/H51A showed negligible binding to Ins(1,4,5)P₃ when compared with WT under our conditions, again indicating that the ENTH-like site has negligible Ins(1,4,5)P₃ affinity under this condition. The binding isotherm for WT was thus fit by assuming the presence of one active-binding site for Ins(1,4,5)P₃ with a $K_a$ of 174 μM (Fig. 5). Suc-
successful fitting of experimental data using the equation derived from the one-to-one binding model suggests the validity of our binding model.

We also measured the membrane penetration activity of N-ANTH domain WT and mutants by the lipid monolayer assay. We previously reported that the epsin ENTH domain can effectively penetrate in a PtdIns(4,5)P₂-dependent manner the lipid monolayer whose packing density is comparable with that of cell membranes (i.e. \( \pi \approx 31 \) dyne/cm), whereas the ANTH domain cannot (24). As shown in Fig. 6A, N-ANTH domain WT was able to penetrate the lipid monolayer with \( \pi > 31 \) dyne/cm in the presence of PtdIns(3,4,5)P₃ (or PtdIns(4,5)P₂) in the monolayer; however, \( \Delta N \) and K38A/R50A/H51A showed much lower activity under the same conditions. Thus, N-ANTH domain of AAL11587 has the ENTH domain-like membrane penetration activity because of the presence of the \( H₀ \)-like helix as predicted by our modeling. We then measured the membrane deforming activity of the N-ANTH domain by the vesicle tubulation assay. The electron microscopy images (Fig. 6B) clearly show that the N-ANTH domain WT has the vesicle tubulation activity, which is comparable with the reported activity of the epsin ENTH domain (23), whereas \( \Delta N \) and K38A/R50A/H51A do not. Taken together, these experimental data corroborate our computational modeling and prediction that the N-ANTH domain of AAL11587 has unique membrane binding properties and the membrane deforming activity due to the presence of the \( H₀ \)-like helix and novel phosphoinositide-binding sites. Most important, this example demonstrates that membrane binding properties of proteins can be systematically and reliably predicted by our computational analysis.

**DISCUSSION**

In this study, we combined the high throughput comparative modeling protocol, SkyLine (9), with systematic high throughput analysis of structural and biophysical properties of proteins for genome-wide identification of ANTH domains and explicit prediction of their membrane binding properties. We first compiled a collection of sequences that are likely to adopt the three-dimensional structure characteristic for ANTH domain through combined three-dimensional modeling and model evaluation. Using a stringent evaluation criterion, our approach yielded a collection of sequences displaying the largest extent of structural similarity with existing structures of ANTH domains, and yet it was able to detect novel sequences that
Proteomic Study of ANTH Domains

FIGURE 6. Membrane penetration and vesicle tubulation activities of the AAL11587 ANTH domain WT and mutants. A, monolayer penetration activity of WT ( ), K38A/R50A/H51A (Δ), and ΔN ( ). Epsin ENTH domain ( ) was shown for comparison. A POPC/PtdIns(3,4,5)P3 monolayer (97:3) spread on 20 mM Tris-HCl, pH 7.4, containing 0.16 m NaCl was used for all measurements. B, electron microscope images of POPC/POPE/PtdIns(3,4,5)P3 (75:20:5) vesicles incubated with 2 μM of each protein (or without protein) for 15 min at 25°C. Representative images were taken by scanning electron microscope. Scale bars, 0.5 μm.

could not be identified as ANTH domains by any other method, including PFAM. The collection of models of all available ANTH domains allowed us to systematically analyze their structural and biophysical properties and classify them into subgroups accordingly.

Many functional predictions of proteins rely on the search for the signature motif for a functional site, such as catalytic or ligand-binding site. However, this approach alone cannot successfully predict the membrane binding properties of proteins because nonspecific electrostatic and hydrophobic interactions contribute at least as much as specific lipid recognition to their overall membrane interactions. We thus incorporated other key structural aspects, such as spatial arrangement of basic patches and the presence of amphiphilic helices, which are essential for membrane-protein interactions into our functional prediction of ANTH domains. Indeed, we found that only 66% of identified ANTH domains contain a canonical PtdIns(4,5)P2-binding site. We identified other types of potential phosphoinositide-binding sites, each of which is conserved within a subgroup of ANTH domains. For example, a canonical PtdIns(4,5)P2-binding motif is present only in 2 of 33 proteins of fungi ANTH domains (Yap1801/Yap1802 subgroup). However, all proteins of this subgroup display an alternative “shifted” basic patch and are likely to perform the membrane-anchoring function through interaction with phosphoinositides. In a group of plant ANTH domains, an alternative phosphoinositide-binding site, which is similar to the phosphoinositide-binding site the ENTH domain family, is found. This migration of the ligand-binding site within the same structural frame, as well as revivals of ancient evolutionary pattern, as observed in ANTH family, can be captured only using structure-based approach. The presence of H0 in some ANTH proteins is another feature that has been previously attributed to the ENTH family. It is thus logical to assume that both atypical phosphoinositide-binding site and the ability to form H0 were inherent to the common precursor of ANTH and ENTH domains.

Our high quality composite model of the dual-patch N-ANTH domain of AAL11587 provides mechanistic insight into how its unique structure dictates membrane interaction. The presence of H0-like N-terminal helix transforms two basic patches on the domain into two potential phosphoinositide-binding sites. In silico docking of two Ins(1,4,5)P3 molecules to the model structure results in the extensive network of hydrophobic bonds in two separate lipid binding clefts, suggesting that the calculated ligand conformations are structurally feasible. Our experimental findings that the mutation of the ANTH-like site abrogates the Ins(1,4,5)P3 binding of the N-ANTH domain and that Ins(1,4,5)P3 binding of the N-ANTH domain WT can be successfully fitted assuming the one-to-one binding stoichiometry suggests that the ENTH-like site may be cryptic in the intact protein but may become exposed when the protein binds the membrane. Also, the N-ANTH domain shows considerably different phosphoinositide specificity (i.e. modest PtdIns(3,4,5)P3 specificity) from that of the epsin ENTH domain, which is consistent with our model suggesting that the N-ANTH domain has a different pattern of hydrogen bonds with a ligand, as compared with CALM-ANTH and epsin1-ENTH. PtdIns(4,5)P2 is generally the most abundant phosphoinositide at the plasma membrane (2), and PtdIns(3,4,5)P3 has not been detected in plants yet. Thus, PtdIns(4,5)P2 may still be a major ligand for AAL11587 under normal conditions. However, the protein would respond to PtdIns(3,4,5)P3 if produced locally under certain physiological conditions. Although we lack information about the lipid composition of the membrane to which AAL11587 binds, calculations with our detailed all-atomic models indicate that, even in the case of a neutral membrane, the desolvation effects favor the orientation that allows insertion of the hydrophobic residues of H0 into the membrane. It has been reported that membrane penetration of hydrophobic residues enhances the membrane affinity of peripheral proteins by slowing the membrane dissociation step (33). For the epsin ENTH domain, membrane penetration of H0 is also essential for its membrane deforming activity (24). Likewise, the H0-like N-terminal helix of AAL11587 drives membrane insertion, increases overall affinity, and induces vesicle deformation and tubulation. Thus,
AAL11587 and other subfamily members may play a key role in membrane remodeling processes in plant cells.

Undoubtedly, further studies are needed to fully understand the mechanism and the physiological significance of unique membrane binding properties of AAL11587 and its subfamily members, which is beyond the scope of this work. Most importantly, this study on ANTH domains demonstrates that membrane binding properties of proteins can be systematically and reliably predicted by our combinatorial approach incorporating high throughput homology modeling and in-depth calculation of biophysical properties. The same approach can be applied to the identification and functional prediction of other lipid binding domains.

REFERENCES

1. Cho, W., and Stahelin, R. V. (2005) Annu. Rev. Biophys. Biomol. Struct. 34, 119–151
2. Lemmon, M. A. (2008) Nat. Rev. Mol. Cell Biol. 9, 99–111
3. Heo, W. D., Inoue, T., Park, W. S., Kim, M. L., Park, B. O., Wandless, T. J., and Meyer, T. (2006) Science 314, 1458–1461
4. Cho, W. (2006) *Sci. STKE* 2006, pe7
5. Bhardwaj, N., Stahelin, R. V., Langlois, R. E., Cho, W., and Lu, H. (2006) *J. Mol. Biol.* 359, 486–495
6. Lin, H. H., Han, L. Y., Zhang, H. L., Zheng, C. J., Xie, B., and Chen, Y. Z. (2006) *J. Lipid Res.* 47, 824–831
7. Bhardwaj, N., Gerstein, M., and Lu, H. (2010) *BMC Bioinformatics* 11, Suppl. 1, S6
8. Xiong, W., Guo, Y., and Li, M. (2010) *Bioinformatics* 26, 427–431
9. Lee, H., Li, Z., Silkov, A., Fischer, M., Petrey, D., Honig, B., and Murray, D. (2006) *J. Struct. Funct. Genomics* 11, 51–59
10. Ford, M. G., Mills, I. G., Peter, B. J., Vallis, Y., Praefcke, G. J., Evans, P. R., and McMahon, H. T. (2002) *Nature* 419, 351–360