Origins of PDZ Domain Ligand Specificity

STRUCTURE DETERMINATION AND MUTAGENESIS OF THE ERBIN PDZ DOMAIN*[S]

Nicholas J. Skelton‡§, Michael F. T. Koehler‡, Kerry Zobel‡, Wai Lee Wong¶, Sherry Yeh¶, M. Theresa Pisabarro‡, Jian Ping Yin‡, Laurence A. Lasky‖, and Sachdev S. Sidhu‡**

From the Departments of ‡Protein Engineering, ¶Assay and Automation Technology, and ‖Molecular Oncology, Genentech, Inc., South San Francisco, California 94080

The LAP (<u>l</u>eucine-rich repeat <u>a</u>nd <u>P</u>DZ-containing) family of proteins play a role in maintaining epithelial and neuronal cell size, and mutation of these proteins can have oncogenic consequences. The LAP protein Erbin has been implicated previously in a number of cellular activities by virtue of its PDZ domain-dependent association with the C termini of both ERB-B2 and the p120-catenins. The present work describes the NMR structure of Erbin PDZ in complex with a high affinity peptide ligand and includes a comprehensive energetic analysis of both the ligand and PDZ domain side chains responsible for binding. C-terminal phage display has been used to identify preferred ligands, whereas binding affinity measurements provide precise details of the energetic importance of each ligand side chain to binding. Alanine and homolog scanning mutagenesis (in a combinatorial phage display format) identifies Erbin side chains that make energetically important contacts with the ligand. The structure of a phage-optimized peptide (Ac-TGW^{-4}ETW^{-1}V; IC$_{50}$ = ~0.15 μM) in complex with Erbin PDZ provides a structural context to understand the binding energetics. In particular, the very favorable interactions with Trp^{-1} are not Erbin side chain-mediated (and therefore may be generally applicable to many PDZ domains), whereas the β2-β3 loop provides a binding site for the Trp^{-4} side chain (specific to Erbin because it has an unusually long loop). These results contribute to a growing appreciation for the importance of at least five ligand C-terminal side chains in determining PDZ domain binding energy and highlight the mechanisms of ligand discrimination among the several hundred PDZ domains present in the human genome.

In the post-genomic era, identification of interactions between proteins has become a significant challenge to achieving a comprehensive understanding of cellular biology (1, 2). Although many such interactions have been described (3), many

more remain to be discovered (4). Research over the past decade has identified many types of protein-protein interactions that participate in intracellular signaling pathways (5). A common feature of such pathways is the involvement of several types of small protein domains (<100 residues) whose sole function is to recognize sequence motifs presented by other members of the pathway (6, 7). Single proteins often contain multiple copies of the same or different protein interaction modules, permitting the formation of the complex, multicomponent assemblies necessary to transmit a specific signal. Indeed, some proteins contain only protein interaction modules and may be considered adapters or scaffolds on which other "active" components of a signaling pathway are brought into proximity (8). Although these interaction domains are often readily identified on the basis of primary sequence, identification of the binding partner relevant to a particular signaling cascade is often difficult.

The PDZ domain, so-called because it was first recognized in the proteins post-synaptic density-95, <u>d</u>iscs large, and <u>z</u>onula occludens 1 (9–11), is a common component of such scaffold proteins. As many as 440 PDZ domains in 259 different proteins have been proposed to exist within the human genome (12). PDZ domains are ~90 residues in size and adopt a common fold consisting of a β-barrel capped by α-helices (13). The predominant function of PDZ domains is to recognize the extreme C termini of other proteins, thereby bringing signaling pathway components into proximity (14–16). Numerous structural and biochemical studies have demonstrated that C-terminal peptide ligands always bind in a groove between a β-strand (β2; see Figs. 1 and 2 below) and an α-helix (α2) (reviewed in Refs. 17 and 18). The ligand is arranged in an antiparallel fashion with respect to the PDZ domain strand, and the ligand carboxylate is hydrogen-bonded to backbone amide nitrogen groups in a conserved GLGF motif located prior to strand β2 (19).

Initial studies identified a C-terminal (S/T)XV motif as being necessary for PDZ binding (14–16). Study of ligands from synthetically or biochemically derived peptide libraries has revealed a more extensive and complex picture of selectivity that involves as many as six C-terminal residues (20–24). These methods also provide an alternative to protein ligand identification from yeast two-hybrid experiments; data base searches for proteins that have C termini that match the optimal ligand sequence are potential protein ligands *in vivo*. The optimized peptide ligands themselves may also be used to antagonize a particular PDZ domain and observe cellular phenotypes, giving further insight into function (24).

The LAP (<u>l</u>eucine-rich repeat <u>a</u>nd <u>P</u>DZ-containing) proteins are a recently described family of scaffold proteins that are involved in the formation of membrane complexes and the

FIG. 1. **Structure-based sequence alignment of Erbin PDZ domain (*Erbin*; Protein Data Bank code 1N7T) with the third PDZ domain of PSD-95 (*PSD95.3*; Protein Data Bank code 1BE9), α-syntrophin (*aSyn*; Protein Data Bank code 2PDZ), and H⁺/Na⁺ exchange regulatory factor PDZ domain 1 (*HNERF*; Protein Data Bank code 1I92).** Regular secondary structure elements (strands and helices denoted by *arrows* and *ellipses* above the sequence) were superimposed onto those of Erbin PDZ, and the corresponding sequence alignments were extracted. Additional sequences for Magi-3 PDZ2 (*Magi3.2*), Densin-180 (*Densin*), and human Scribble PDZ domains 1–4 (*Scrib1–Scrib4*) were manually aligned to the other sequences. The *numbers* in *bold type* correspond to the residue position within the Erbin PDZ domain, whereas the *numbers* in *regular type* correspond to the position within each element of secondary structure (see text). The *asterisks* under the sequence indicate those residues in Erbin PDZ that contact the ligand. *Underlined* Erbin PDZ residues correspond to sites that were varied in the combinatorial scanning mutagenesis; alanine mutations with $F > 10$ are in *bold type* and colored *red*.

maintenance of epithelial and neuronal cell shape and polarity (25). For example, in *Drosophila*, mutation of the Scribble LAP protein results in loss of epithelial cell polarity and morphology as well as uncontrolled, tumor-like growth (26). The LAP proteins have a domain structure comprising 16 N-terminal leucine-rich repeats and up to four C-terminal PDZ domains. On the basis of yeast two-hybrid experiments, a mammalian LAP protein that recognizes the C terminus of ERB-B2 (a member of the epidermal growth factor receptor family) has been identified and given the name Erbin, for ERB-interacting protein (27). More recently, we have used C-terminal phage display (23) to identify optimal ligands for the Erbin PDZ domain (24). These ligands are quite different in sequence from the C terminus of ERB-B2 and also bind ~1000-fold more tightly to Erbin PDZ than the C-terminal peptide from ERB-B2 (24). *In vivo* interactions with the p120-like catenin proteins have been proposed and tested on the basis of these results (24), whereas yeast two-hybrid screens have also identified these same interactions (28, 29). These data suggest that LAP proteins are targeted to p120-catenin-localized junctional regions via a PDZ-mediated interaction (24).

In the present work, we have explored in detail the interactions between Erbin PDZ and the optimal phage-derived peptide ligand. A preference for a penultimate tryptophan residue in the ligand has been confirmed by extensive phage library selections, a feature that is also present in the optimal ligand for the second PDZ domain of the membrane-associated guanylate kinase Magi-3 (23). Affinity measurements of synthetic peptide analogs of the optimal ligand have been made to quantitate the energetic contributions of the five C-terminal ligand residues; the penultimate tryptophan is indeed beneficial to binding (affinity decreases by >1000-fold when replaced by alanine). An efficient phage-based combinatorial scanning approach has also been utilized to identify residues within Erbin PDZ that contribute energetically to ligand binding (30). Although we have previously described a homology-based model of a domain of Magi-3 (23), the use of homology modeling to interpret the structure-activity data is of limited value because the primary sequence of Erbin PDZ differs in several key regions from that of other PDZ domains whose structures are known (Fig. 1). Thus, we have also used NMR spectroscopy to determine the structure of Erbin PDZ in complex with the p120-catenin-like phage-derived ligand.[1] This structure provides a clear view of the interactions made by the penultimate

tryptophan residue and more generally provides a framework within which to understand the affinity and selectivity of peptide binding to Erbin PDZ. This work provides one of the most extensive characterizations to date of the structural and energetic (ligand and protein) components of a PDZ domain interaction with a C-terminal peptide.

EXPERIMENTAL PROCEDURES

Materials—Enzymes were from New England Biolabs. Maxisorp immunoplates and 384-well assay plates were from Nalge NUNC International (Naperville, IL). *Escherichia coli* XL1-Blue, *E. coli* BL21, and M13-VCS were from Stratagene. Plasmid pET15b was from Novagen. Thrombin was from Calbiochem. Bovine serum albumin and Tween 20 were from Sigma. Horseradish peroxidase/anti-M13 antibody conjugate and Superdex-75 were from Amersham Biosciences. Nickel-nitrilotriacetic acid was from Qiagen. 3,3′, 5,5′-Tetramethyl-benzidine/H_2O_2 peroxidase substrate was from Kirkegaard and Perry Laboratories Inc. AlphaScreen™ reagents and a plate reader were from PerkinElmer Life Sciences.

Oligonucleotide Synthesis—Oligonucleotides for combinatorial scanning were designed as described previously using equimolar DNA degeneracies (30). The particular mutagenic oligonucleotides are listed in Supplementary Table I.

Synthetic Peptides—The peptides were synthesized using standard Fmoc (*N*-(9-fluorenyl)methoxycarbonyl) protocols, cleaved off the resin with 2.5% triisopropylsilane and 2.5% H_2O in trifluoroacetic acid, and purified by reversed-phase high performance liquid chromatography. The purity and mass of each peptide were verified by liquid chromatography/mass spectrometry.

Statistical Analysis of Erbin PDZ Binding Specificity—Previously described procedures were used to isolate peptides that bound to a GST-Erbin[2] PDZ fusion, using a library of random heptapeptides fused to the C terminus of the M13 gene-8 major coat protein (23, 24). After two rounds of selection, individual clones were grown in a 96-well format in 500 μl of 2YT broth supplemented with carbenicillin and M13-VCS, and the culture supernatants were used directly in phage enzyme-linked immunosorbent assays (31) to detect peptides that bound specifically to Erbin PDZ. A total of 148 peptide sequences derived from positive clones were aligned, and the occurrence of each natural amino acid at each position was tabulated. The occurrence of each amino acid was normalized by dividing by the number of times the amino acid was encoded by the NNS codon. The normalized data set was used to calculate the percentage of occurrence of each amino acid at each position (see Table I).

Construction of Libraries for Erbin PDZ Shotgun Scanning—Erbin PDZ was displayed on the surface of M13 bacteriophage by modifying a previously described phagemid (pS1602) (32). Standard molecular biology techniques were used to replace the fragment of pS1602 encoding

---

[1] Model 2 in this ensemble is closest to the geometric mean and has been used as a representative single structure for the figures in this manuscript.

[2] The abbreviations used are: GST, glutathione *S*-transferase; HNERF, H⁺/Na⁺ exchange regulatory factor; NOESY, nuclear Overhauser effect spectroscopy; wt, wild type; PBS, phosphate-buffered saline; TOCSY, total correlation spectroscopy; HSQC, heteronuclear single quantum coherence.

human growth hormone with a DNA fragment encoding Erbin PDZ. The resulting phagemid (pS2202d) contained an open reading frame that encoded the maltose-binding protein secretion signal, followed by an epitope tag (amino acid sequence: SMADPNRFRGKDLGS), followed by Erbin PDZ and ending with the C-terminal domain of the M13 gene-3 minor coat protein. *E. coli* harboring pS2202d were co-infected with M13-VCS helper phage and grown at 37 °C without isopropyl-1-thio-β-D-galactopyranoside induction, resulting in the production of phage particles that encapsulated pS2202d DNA and displayed Erbin PDZ in a monovalent format.

Libraries were constructed using previously described methods (31) with appropriately designed "stop template" versions of pS2202d. For each library, we used a stop template that contained TAA stop codons within each of the regions to be mutated. The stop template was used as the template for the Kunkel mutagenesis method (33) with mutagenic oligonucleotides (see above) designed to simultaneously repair the stop codons and introduce mutations at the desired sites.

For shotgun scanning, wild type codons were replaced with the corresponding degenerate codons shown in Table I of Vajdos *et al.* (34). Two separate libraries were constructed with each library designed to mutate 22 Erbin PDZ residues with no overlap between the two. Libraries A1 and A2 were constructed with mutagenic oligonucleotides *A1a* and *A1b* or *A2a* and *A2b*, respectively. Library A1 mutated residues in two continuous stretches of sequence between positions 16–28 and 46–55, whereas library A2 mutated residues between positions 31–43 and 78–95. For the shotgun homolog scan, libraries H1 and H2 were constructed in an analogous fashion with mutagenic oligonucleotides *H1a* and *H1b* or *H2a* and *H2b*, respectively. The library diversities were as follows: A1, 4.2×10^{10}; A2, 4.0×10^{10}; H1, 4.4×10^{10}; and H2, 4.2×10^{10}.

Library Sorting and Analysis—Phage from the libraries described above were propagated in *E. coli* XL1-blue with the addition of M13-VCS helper phage. After overnight growth at 37 °C, phage were concentrated by precipitation with polyethylene glycol/NaCl and resuspended in PBS, 0.5% bovine serum albumin, 0.1% Tween 20 as described previously (31). Phage solutions (10^{12} phage/ml) were added to 96-well Maxisorp immunoplates that had been coated with capture target and blocked with bovine serum albumin. Two different targets were used; for the display selection the target was an immobilized antibody that recognized the epitope tag fused to the N terminus of Erbin PDZ, whereas for the function selection a biotinylated peptide that binds to Erbin PDZ with high affinity (biotin-TGWETWV) (24) was immobilized on streptavidin-coated plates. Following a 2-h incubation to allow for phage binding, the plates were washed 10 times with PBS, 0.05% Tween 20. Bound phage were eluted with 0.1 M HCl for 10 min, and the eluent was neutralized with 1.0 M Tris base. Eluted phage were amplified in *E. coli* XL1-blue and used for further rounds of selection.

Individual clones from each round of selection were grown in a 96-well format in 500 μl of 2YT broth supplemented with carbenicillin and M13-VCS, and the culture supernatants were used directly in phage enzyme-linked immunosorbent assays (31) to detect phage-displayed Erbin PDZ variants that bound to either biotin-TGWETWV or anti-tag antibody. After two rounds of selection, greater than 50% of the clones exhibited positive phage enzyme-linked immunosorbent assay signals at least 2-fold greater than signals on control plates coated with bovine serum albumin. These positive clones were subjected to DNA sequence analysis (see below).

The sequences were analyzed with the program SGCOUNT as described previously (30). SGCOUNT aligned each DNA sequence against the wild type DNA sequence by using a Needleman-Wunch pairwise alignment algorithm, translated each aligned sequence of acceptable quality, and tabulated the occurrence of each natural amino acid at each position. For the function selection, the number of analyzed clones are indicated in parentheses following the name of each library: A1 (185 clones), A2 (180 clones), H1 (190 clones), and H2 (170 clones). For the display selection, the following numbers of clones were analyzed: A1 (83 clones), A2 (83 clones), H1 (94 clones), and H2 (96 clones).

DNA Sequencing—Culture supernatants containing phage particles were used as templates for PCRs that amplified DNA fragments containing the Erbin PDZ gene, and these fragments were sequenced as described previously (34).

Affinity Assays—An Erbin PDZ construct with GST fused to the N terminus was prepared as described (24). The binding affinities of peptides for Erbin PDZ were determined as IC_{50} values using the AlphaScreen™, a bead-based chemiluminescence assay with optimized concentration of reagents. The IC_{50} was defined as the concentration of peptide that blocked 50% of the chemiluminescence arising from the interaction of anti-GST acceptor beads coated with Erbin PDZ-GST and streptavidin donor beads coated with biotinylated peptide (biotin-TG-WETWV). The assays were performed at room temperature in white opaque 384-well plates (25 μl/well) under subdued lighting to reduce nonspecific chemiluminescence. The assay buffer was PBS, 0.5% Tween 20, 0.1% bovine γ-globulin, 1 ppm proclin. The reaction mixture contained fixed concentrations of anti-GST acceptor beads (16 μg/ml), biotin-TGWETWV (36 nM), and Erbin PDZ-GST (3 nM). Serial dilutions of peptide were added, followed by addition of streptavidin donor beads (20 μg/ml). The mixture was incubated at room temperature for 1 h and read on an AlphaQuest plate reader set at 1 s/well.

Purification of Erbin PDZ Protein for NMR Spectroscopy—A DNA fragment encoding residues 1273–1371 of Erbin was cloned into the *Nde*I/*Bam*HI sites of the pET15b expression vector, creating a fusion with an N-terminal His tag followed by a thrombin cleavage site. *E. coli* BL21 cultures harboring the expression plasmid were grown at 37 °C to mid-log phase ($A_{600} = 0.7$) in M9 medium with $^{15}NH_4Cl$ or $^{15}NH_4Cl/^{13}C$ glucose supplemented with 50 μg/ml carbenicillin. Protein expression was induced by the addition of 1.0 mM isopropyl-1-thio-β-D-galactopyranoside, and the cells were harvested after 2 h further growth by centrifugation at $4,000 \times g$ for 15 min and stored at −80 °C. The pellet was resuspended in 50 mM Tris, pH 8.0, 500 mM NaCl, 1 μM phenylmethylsulfonyl fluoride and sonicated for 3 min on ice. The suspension was centrifuged for 30 min at $10,000 \times g$, and the supernatant was loaded onto a nickel-nitrilotriacetic acid-agarose column. The column was washed with 50 mM Tris, pH 8.0, 500 mM NaCl, 10 mM imidazole, and then the protein was eluted with 250 mM imidazole in the same buffer. Fractions containing the protein of interest were pooled, thrombin was added (1 unit/mg of protein), and the sample was dialyzed overnight against PBS at 4 °C. The protein sample was then concentrated and further purified over a Superdex-75 column in PBS to remove thrombin and the cleaved His tag. The identity of the purified protein was verified by N-terminal sequencing and mass spectrometry. In addition to Erbin PDZ domain (residues 1273–1371), the construct also contains an N-terminal GSHM tail from the expression vector (*i.e.* Gly[5] of the present construct corresponds to Gly[1273] of full-length Erbin).

NMR Spectroscopy and Structure Determination—NMR samples typically contained 1.0–1.5 mM protein, 50 mM phosphate buffer, pH 6.5, 50 mM NaCl, and 0.1 mM d_{11}-2,2-dimethyl-2-silapentane-*s*-sulphonate (DSS) for chemical shift referencing in 93% H_2O, 7% D_2O. A "100% D_2O" sample was prepared by lyophilization and resuspension in 99.995% D_2O. NMR spectra were acquired at 25 °C on Bruker DRX600 and DRX800 spectrometers equipped with triple resonance, triple axis actively shielded gradient probes. Addition of the phage-optimized peptide (AcTGWETWV) to Erbin PDZ indicated that the free and bound resonances were in slow exchange. Aliquots of peptide were added until ^1H-^{15}N HSQC peaks for the free protein had disappeared and no further change of intensity or line width was seen for the bound peaks. Backbone resonance assignments for Erbin PDZ were obtained from the following double and triple resonance experiments acquired at 600 MHz in H_2O solution, as described (35, 36): ^1H-^{15}N three-dimensional NOESY-HSQC, ^1H-^{15}N three-dimensional TOCSY-HSQC, three-dimensional HNCA, three-dimensional CBCA(CO)NH, and three-dimensional HBHA(CBCACO)NH. Erbin PDZ side chain assignments were completed from a three-dimensional HCCH-TOCSY spectrum acquired at 800 MHz in D_2O solution. Peptide resonance assignments were obtained from ^{15}N-filtered two-dimensional ^1H NOESY and TOCSY spectra at 600 MHz in H_2O solution and ^{13}C-filtered ^1H NOESY and TOCSY spectra at 800 MHz in D_2O solution. Details of these experiments are presented in Supplementary Table II.

Distance restraints were obtained from analysis of the following NOESY spectra acquired at 800 MHz: three-dimensional ^1H-^{15}N NOESY-HSQC, two-dimensional ^{15}N-filtered ^1H NOESY, three-dimensional ^{13}C-edited NOESY, and two-dimensional ^{13}C-filtered ^1H NOESY. Intermolecular interactions were identified unambiguously in a three-dimensional ω_1-^{13}C-filtered, ω_2-^{13}C-edited NOESY spectrum. Initial NOESY peak assignments were made on the basis of the assigned resonance positions and a homology model of Erbin PDZ as described previously (37), followed by several rounds of structure calculation and manual restraint checking and peak assignment. The homology model was based on the second and third PDZ domain of PSD-95 (1QLC (38); 1BE9 (19)) and the location of secondary structure elements identified in a preliminary analysis of the NMR data. Dihedral angle restraints were obtained from analysis of three-dimensional ^{15}N-^1H HNHA (ϕ), three-dimensional ^{15}N-^1H HNHB, and three-dimensional ^{15}N-^1H TOCSY-HSQC (χ_1) spectra. Additional loose backbone dihedral angle restraints were obtained from analysis of backbone chemical shifts with the program TALOS (39). Restraints were applied for good fits to the

TABLE I

Statistical analysis of Erbin PDZ binding specificity

Erbin PDZ-binding peptides were isolated from a random heptapeptide library, 148 sequences were aligned, and the occurrence of each amino acid at each position was tabulated. The percentage of occurrence of each amino acid type at each position was calculated after normalization for codon bias. The preferred residue at each position is shown in bold type.

Position[a]	Occurrence of amino acid type																			
	Trp	Phe	Tyr	Met	Leu	Ile	Val	Cys	Ala	Gly	Pro	Gln	Asn	Thr	Ser	His	Arg	Lys	Glu	Asp
											%									
0					**2.7**	**1.4**	**96**													
−1	**100**																			
−2							2.9							**80**	**14**				1.4	1.4
−3										1.4					0.2				**36**	**62**
−4	**28**	8.6	7.8	5.2	2.6	5.2	1.7	0.9	1.7	4.3	2.6	2.6	1.7	0.9	1.7	3.4	5.2	5.2	3.4	6.9
−5	18	7.5	8.5		0.9	0.9	2.8	1.9	1.9	10	3.8	6.6	2.8	1.9	2.8	6.6	3.8	2.8	5.7	11
−6	11		2.8	2.8	2.8	1.4	4.2	6.9	4.2	11	1.4	5.6	2.8	4.2	8.3	2.8	13	11	1.4	2.8

[a] Positions in the peptide ligand, from the C terminus to the N terminus, are designated 0, −1, etc.

chemical shifts (as defined by the program) with the allowed range being the TALOS-defined mean ± the larger of 30° or three times the TALOS-calculated standard deviation. Finally, residual dipolar coupling restraints were measured from a solution of Erbin PDZ with peptide in the presence of 15 mg/ml of Pf1phage (ASLA Ltd.) (40) using the in-phase anti-phase method of Tjandra and Bax (41). The values for the axial component of the molecular alignment tensor (D_a) and the rhombicity (R) were obtained by fitting calculated to experimental residual dipolar coupling values using a simple Powell minimization procedure (41). Coordinates from initial structures calculated on the basis of nuclear Overhauser effect and dihedral angle restraints only were used for this purpose and yielded values of −19 Hz and 0.4 for D_a and R, respectively. The structures were calculated using the program CNX (v2000.1; Accelrys, San Diego, CA). 100 structures were calculated using torsion angle dynamics followed by Cartesian dynamics and minimization. The 20 structures of lowest restraint violation energy were chosen to represent the solution structure of the Erbin PDZ domain-peptide complex. Details of the input restraints and structural statistics are presented in Supplementary Table III.

RESULTS

Statistical Analysis of Erbin PDZ Binding Specificity—A phage-displayed peptide library was used to explore Erbin PDZ binding specificity. Instead of focusing on a small set of molecules with high affinity, we used a statistical analysis of a large number of potential ligands to accurately define the specificity of peptides able to bind to Erbin PDZ. After two rounds of binding selection, approximately half of the phage clones bound specifically to Erbin PDZ in a phage enzyme-linked immunosorbent assay; DNA sequencing revealed that most of these clones were unique. An alignment of 148 sequences defined the optimal binding consensus for Erbin PDZ, which agreed with earlier results obtained from sequencing a limited number of clones after three rounds of selection (24). In addition, the presence of rare clones among the large number of sequences revealed suboptimal but still significant preferences for certain amino acids at most binding positions (Table I). In the remainder of the manuscript, the standardized PDZ ligand nomenclature will be used to describe the peptide residues, wherein the C terminus is designated residue 0 (*e.g.* Val⁰) and the remaining residues are numbered with negative integers whose absolute value increases toward the N terminus (42). At the peptide C terminus, valine was overwhelmingly preferred along with a much rarer occurrence of leucine and isoleucine, whereas tryptophan was selected exclusively at the −1 position. Threonine was the predominant selection at the −2 position, although serine and, to a lesser extent, valine or acidic residues were also observed. Acidic residues were selected almost exclusively at the −3 site with a 2-fold preference for aspartate over glutamate. Although all amino acid types were observed at the −4 and −5 sites, aromatic residues, especially tryptophan, did predominant at −4, and there was a slight preference for Trp⁻⁵ (Table I).

Specificity of Peptide Binding—To more fully understand the

energetic contributions of different residues at each ligand position, the relative binding affinity was measured for a series of synthetic peptides (Table II). The heptapeptide identified in the preliminary phage analysis (24) bound with an IC₅₀ of 0.15 μM; the lack of consensus for the two N-terminal residues (Table I) suggests that they are not important for binding, and indeed removal of them actually leads to a slight improvement in affinity (7-fold; IC₅₀ = 0.02 μM). In the context of the pentapeptide, at most ligand sites the relative binding affinity of a given amino acid substitution correlates well with the preference for that residue in the phage selection. For example, Leu⁰ and Ile⁰ bind ~90-fold less tightly then Val⁰ and are selected in less than 3% of the clones. Similarly, Ser⁻² and Val⁻² are ~10-fold reduced in affinity compared with Thr⁻², and are also selected 6- and 25-fold less often than threonine, respectively. The one point of disagreement between the selection data and the binding data is in the relative affinity of Asp⁻³ *versus* Glu⁻³ because aspartate is selected more often yet binds 20-fold less tightly. Replacement of Trp⁻⁴ with other aromatic residues leads to an 8-fold loss in affinity, as expected from the presence of all three aromatic residues at this site; replacement with alanine was more detrimental to binding (~50-fold) as expected by selection of alanine in <2% of ligands sequenced. Tryptophan is the only amino acid selected at the −1 position, and even replacement with phenylalanine caused a dramatic loss of affinity (165-fold). Interestingly, replacing Trp⁻¹ with Pro, as found in the C-terminal peptide of ERB-B2, which has been implicated in Erbin binding (27), reduces the affinity of the pentapeptide by >3300-fold. The alanine scan data in Table II give an indication of the relative contribution of the peptide side chains to the binding energy. Generally, the closer a residue to the C terminus, the more important it is for binding, with Val⁰ → Ala being most deleterious (3700-fold). As found in previous studies, the C-terminal carboxylate is also critical for high affinity binding (2400-fold decrease in affinity when amidated). The importance of the C-terminal residues is also seen in the N-terminal truncation analogs in Table II; even the acetylated dipeptide Ac-WV binds with an affinity of ~300 μM.

Structure of Erbin PDZ domain—The Erbin PDZ contains six β-strands arranged into two antiparallel β-sheets (Fig. 2), as has been observed in other PDZ domain structures (*e.g.* see Ref. 17 and the references therein). The majority of the Erbin PDZ domain structure is well defined by and agrees well with the 1991 experimental NMR restraints (Fig. 3*A* and Supplementary Table III). Residues within Erbin PDZ domain will be referred to by their location within each of the secondary structure elements or loops (42) as defined in the canonical third PDZ domain of PSD-95 (19) (Figs. 1 and 2). Thus, β2, β3, and β4 form the "top" sheet, and β1, β6, β4, and β5 form the lower sheet, whereas helix α2 caps the β2–β5 edge of the sandwich (Fig. 2). The extreme termini of the domain (Gly¹–Lys¹⁰ and

TABLE II
IC$_{50}$ values for Erbin PDZ-binding synthetic peptides

The IC$_{50}$ values are the mean concentrations of peptide that blocked 50% of Erbin PDZ binding to a high affinity peptide ligand in a bead-based fluorescence assay performed in quadruplicate. The peptides were synthesized with acetylated N termini. Deviations from the highest affinity sequence are shown in bold type.

Position							IC$_{50}$	IC$_{50}$/IC$_{50}^{WETWV}$
−6	−5	−4	−3	−2	−1	0		
							μM	
Thr	Gly	Trp	Glu	Thr	Trp	Val	0.15 ± 0.01	7.5
	Gly	Trp	Glu	Thr	Trp	Val	0.13 ± 0.01	6.5
		Trp	Glu	Thr	Trp	Val	0.020 ± 0.002	1
		Phe	Glu	Thr	Trp	Val	0.16 ± 0.01	8.0
		Tyr	Glu	Thr	Trp	Val	0.16 ± 0.01	8.0
		Trp	**Asp**	Thr	Trp	Val	0.41 ± 0.06	21
		Trp	Glu	**Ser**	Trp	Val	0.16 ± 0.01	8.0
		Trp	Glu	**Val**	Trp	Val	0.26 ± 0.01	13
		Trp	Glu	Thr	**Phe**	Val	3.3 ± 0.2	170
		Trp	Glu	Thr	**Pro**	Val	66 ± 4	3300
		Trp	Glu	Thr	Trp	**Leu**	1.8 ± 0.1	90
		Trp	Glu	Thr	Trp	**Ile**	1.7 ± 0.1	85
		Ala	Glu	Thr	Trp	Val	1.0 ± 0.1	50
		Trp	**Ala**	Thr	Trp	Val	2.8 ± 0.7	140
		Trp	Glu	**Ala**	Trp	Val	2.0 ± 0.1	100
		Trp	Glu	Thr	**Ala**	Val	30 ± 7	1500
		Trp	Glu	Thr	Trp	**Ala**	74 ± 4	3700
		Trp	Glu	Thr	Trp	**Val-CONH$_2$**	47 ± 18	2400
			Glu	Thr	Trp	Val	0.19 ± 0.02	9.5
				Thr	Trp	Val	27 ± 1	1400
					Trp	Val	284 ± 31	14200
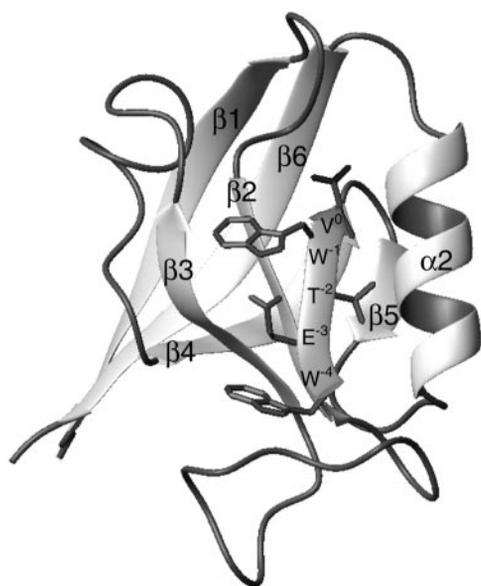
FIG. 2. **Ribbon view of representative Erbin PDZ structure bound to phage-derived peptide.** Elements of regular secondary structure are labeled, as are the side chains of the phage-derived peptide ligand.

Ser^{102}-Ser^{103}) and the N terminus of the ligand (Thr^{-6}–Gly^{-5}) are disordered (Fig. 3*A*); the longer loops (β2:β3 and β3:β4) are defined slightly less well than residues in regular elements of secondary structure (see Supplementary Table III). At the sequence level, several differences are apparent between Erbin PDZ and the canonical third domain from PSD-95, namely one residue fewer in the β1:β2 loop, an extra nine residues in the β2:β3 loop, and two fewer residues in the β3:β4 loop (Fig. 1). The structure of Erbin PDZ reveals that these sequence differences may be readily accommodated without distortion of the fold (Fig. 2). The β1:β2 loop is able to make a more direct connection between the strands without perturbing the important side chain- and backbone-mediated interactions with the ligand (see below). The β2:β3 loop is highly variable in length

and character among PDZ domains, and in the case of Erbin it initially turns toward the β5:α2 loop before a reverse turn at Val^{31}-Gly^{32} (type II) redirects the chain back across the top of strand β2 into two final reverse turns (Pro^{37}-Phe^{38}, distorted type I; Pro^{40}-Asp^{41}, type I) that lead the chain into the start of strand β3 (Fig. 2). The shorter β3:β4 loop in Erbin PDZ precludes the formation of the α1 helix that normally caps this edge of the β-sandwich; instead a series of nested reverse turns are present (Glu^{53}-Gly^{54}, Pro^{55}-Ala^{56}, Ser^{57}-Lys^{58}, and Lys^{58}-Leu^{59}).

Ligand Binding to Erbin PDZ Domain—The interaction between the five C-terminal residues of the phage-derived heptapeptide and Erbin PDZ is clearly defined by more than 200 intermolecular nuclear Overhauser effect restraints (Fig. 3*A* and Supplementary Table III). The peptide extends one edge of the β-sandwich via contacts with strand β2 (Figs. 2 and 3), with several backbone hydrogen bonds observed between Erbin PDZ (β2–1(Phe^{25}) and β2–3(Ile^{27})) and ligand (Val^0 and Thr^{-2}). The backbone amide protons of β1:β2–4 (Leu^{23}), β1:β2–5(Gly^{24}) and β2–1(Phe^{25}) are all directed toward the carboxylate oxygen atoms of Val^0 but at distances slightly longer than that usually considered for a hydrogen bond. The N^e amino groups of β1–7(Lys^{19}) and α2–9(Lys^{87}) are in the vicinity of the Val^0 carboxylate and may give rise to favorable Coulombic contacts. The poor definition of these interactions results from the absence of restraint-generating protons on the carboxylate and amino groups; in addition, we cannot rule out the presence of a bound water molecule in the carboxylate-binding pocket, as seen in other PDZ domain complexes (19). Turning to the ligand side chains, Val^0 pokes into the core of the protein and is surrounded by residues from β1:β2, β2, and α2, whereas Thr^{-2} abuts helix α2, is in van der Waals' contact with α2–5(Val^{83}), and participates in a hydrogen bond from its hydroxyl group to N^{ε2} of α2–1(His^{79}) (Fig. 3*B*). Although in a number of other studies the peptide side chain at position −1 does not make specific structural or energetically favorable intermolecular contacts (19, 20), this is not the case for Erbin PDZ domain: Trp^{-1} reaches across strand β2 and inserts between the side chains of β3–5(Arg^{49}) and β3:β4–1(Gln^{51}) (Fig. 3*B*). Glu^{-3} also reaches across strand β2 toward strand β3; although a lack of

Erbin PDZ Domain Structure and Function Analysis
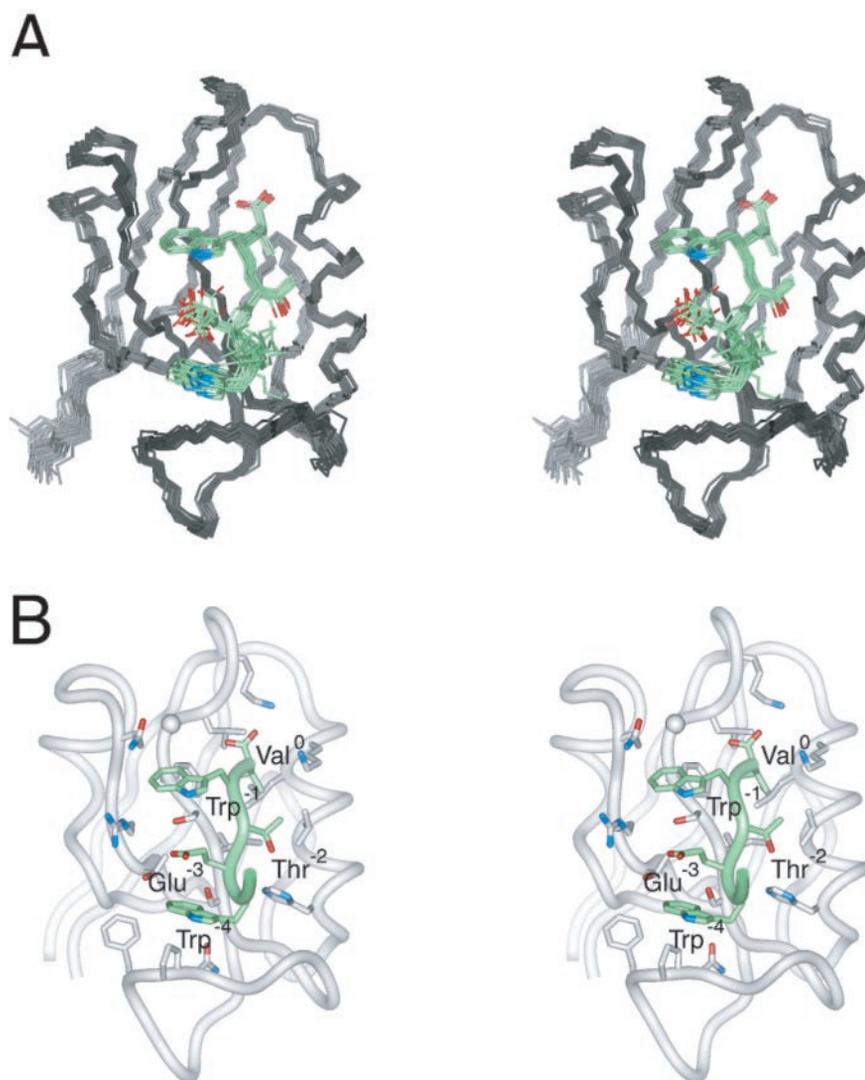
FIG. 3. *A*, ensemble of structures for Erbin PDZ domain (*gray*) bound to the phage optimized peptide (*green*). Only the backbone N, Cα, and C atoms are shown. Selected peptide side chain heavy atoms are included. Root mean square deviation to the mean structure = 0.40 ± 0.05 Å for backbone heavy atoms of residues 10–100; no distance or dihedral angle restraints violated by more than 0.17 Å or 2.5°, respectively. *B*, representative structure showing specific side chains around the peptide binding groove.

restraints precludes the definition of a precise orientation for this side chain, an ionic interaction with $\beta3$–5(Arg^{49}) is likely. Finally, Trp^{-4} also lies toward the $\beta2$ side of the binding cleft and has a number of interactions with Glu^{-3} and residues at the C terminus of the $\beta2$:$\beta3$ loop (Fig. 3*B*).

Shotgun Alanine Scanning and Homolog Scanning of Erbin PDZ Domain—The contribution to peptide binding of individual Erbin PDZ domain residues was assessed by combinatorial alanine scanning (30). A pair of libraries were constructed in which 44 residues in and around the peptide-binding site were represented by trinucleotides that encoded either the wild type Erbin amino acid or alanine (note that because of the particular codons used, some non-alanine mutants were also possible; see Ref. 30). Two additional libraries were constructed in which the same 44 residues were present as either the wild type or a homolog of the wild type residue (the so-called "homolog scan"; Table III). These libraries were then selected for binding to immobilized peptide (Ac-TGWETWV), and ~180 clones positive for binding were sequenced after two rounds of selection. The number of clones with the wild type residue at each position were compared with the number with each designed mutant (either alanine or homolog) and categorized as substitutions that reduce (ratio > 1), do not affect (ratio = ~1), or improve (ratio < 1) binding to peptide. To control for variation in expression or display level for different library members, the libraries were also selected for binding to an immobilized antibody capable of recognizing an epitope tag

that was displayed at the N terminus of all library members. The ratio of wild type to mutant in the peptide selection was then scaled by the ratio of wild type to mutant observed in the antibody selection to give a normalized frequency of occurrence (*F*; Table III).

The results of the alanine and homolog substitutions on peptide binding are mapped onto the structure of Erbin PDZ in Fig. 4. The majority of alanine mutations that have a significant effect on peptide binding (*F* > 20) are proximal to only three of the peptide side chains (Val^0, Thr^{-2}, and Trp^{-4}), emphasizing the importance of these interactions for peptide binding (Fig. 4*A*). Non-alanine mutations of many of these residues are also highly detrimental to peptide binding, even in the case of subtle substitutions of isoleucine or leucine for valine (*e.g.* $\beta1$:$\beta2$–4(L23V), $\beta2$–3(I27V), $\beta3$–1(I45V), and $\alpha2$–8(L86V)). Moreover, several of these mutations also cause a drop in display level, indicating that the wild type residue is necessary for efficient folding of the domain. Non-alanine mutations at a number of additional sites are also detrimental to peptide binding. Although some of these may indicate the loss of direct contacts with the peptide (*e.g.* $\beta1$–7(K19E)), many of the mutations are to proline and hence may decrease peptide binding by an indirect structural perturbation. Curiously, alanine substitutions of residues that contact Trp^{-1} and Glu^{-3} did not decrease peptide binding and in some cases actually improved it (*F*(S28A) = 0.44; *F*(E51A) = 0.25). However, alanine substitutions elsewhere in $\beta3$ were detrimental to binding, suggest-

TABLE III

Erbin PDZ shotgun scan

The wt/mutant ratios were determined from the sequences of binding clones isolated after selection for binding to either a high affinity peptide ligand (function selection) or an anti-tag antibody (display selection). A normalized frequency of occurrence (F) was derived by dividing the function selection wt/mutant ratio by the display selection wt/mutant ratio. In cases where a particular mutation was not observed amongst the function selection sequences, only a lower limit could be defined for the wt/mutant ratio and the F value (indicated by a greater than sign). The F values were determined for alanine (Ala) or homolog (Homo) substitutions and also for two additional substitutions (m2 and m3) in cases where the alanine scan required a tetranomial codon. The identities of non-alanine substitutions are shown in parantheses to the right of each F value. Bold numbers indicate mutations having more than a 10-fold effect on selection.

| Residue[a] | wt/mutant ratios | | | | | | | | F | | | |
| | Function selection | | | | Display selection | | | | | | | |
	wt/Ala	wt/Homo	wt/m2	wt/m3	wt/Ala	wt/Homo	wt/m2	wt/m3	Ala	Homo	m2	m3
Arg[16]	**14.4**	**10.9**	**>173**	**>173**	4.50	**14.5**	1.20	1.63	3.20	0.75 (Lys)	**>144 (Gly)**	**>106 (Pro)**
Glu[18]	1.73	4.94			1.76	7.55			0.98	0.65 (Gln)		
Lys[19]	4.61	**10.2**	**>152**	**>152**	0.60	2.48	2.00	1.50	7.7	4.1 (Arg)	**>76 (Glu)**	**>101 (Thr)**
Asp[20]	0.75	0.71			1.25	1.07			0.60	0.66 (Asn)		
Pro[21]	7.67	3.87			1.71	0.98			4.49	3.95 (Ala)		
Glu[22]	6.46	1.83			0.69	2.88			9.36	0.64 (Gln)		
Leu[23]	**>185**	**>190**	**>185**	**>185**	5.00	4.75	2.22	1.54	>37	**40 (Ile)**	**83 (Pro)**	**120 (Val)**
Gly[24]	**>185**	**>190**			1.65	**17.4**			**>110**	**>10.9 (Ala)**		
Phe[25]	**>183**	**>190**	**>185**	**>185**	4.60	**22.0**	7.67	2.09	**>40**	**>8.6 (Lys)**	**>24 (Ser)**	**>89 (Val)**
Ser[26]	1.48	0.33			1.65	1.97			0.90	0.17 (Ala)		
Ile[27]	**>179**	7.26	**>179**	**29.8**	2.50	1.85	3.33	1.54	**>72**	3.92 (Val)	**>54 (Thr)**	**19 (Val)**
Ser[28]	0.63	0.34			1.42	0.68			0.44	0.50 (Ala)		
Val[31]	9.44	**15.6**			3.61	**18.2**			2.6	0.86 (Ile)		
Arg[34]	4.20	1.68	5.72	**12.6**	1.45	2.39	1.00	0.96	2.9	0.70 (Lys)	5.7 (Gly)	**13 (Pro)**
Asn[36]	**>187**	**>182**	**>187**	**187**	0.30	1.23	0.35	0.50	**>620**	**148 (Asp)**	**>530 (Asp)**	**370 (Thr)**
Pro[37]	**>188**	**21.8**			0.69	1.41			**>270**	**15 (Ala)**		
Phe[38]	**>187**	**17.0**	**>187**	**>187**	1.18	1.79	1.63	1.24	**>160**	9.5 (Tyr)	**>110 (Ser)**	**>150 (Val)**
Arg[39]	2.81	1.12	**132**	**>132**	1.67	1.40	0.89	1.79	1.7	0.80 (Lys)	**>150 (Gly)**	**>74 (Pro)**
Pro[40]	4.50	1.23			0.76	1.09			5.9	1.1 (Ala)		
Asp[41]	7.36	1.81			2.19	1.44			3.4	1.3 (Asn)		
Asp[42]	**>187**	**43.8**			1.59	4.59			**>120**	9.5 (Asn)		
Asp[43]	7.13	1.46			1.68	1.74			4.2	0.84 (Asn)		
Ile[45]	**>183**	**12.6**	**>183**	**61**	2.13	2.03	4.25	0.85	**>86**	6.2 (Val)	**>43 (Thr)**	**72 (Val)**
Phe[46]	**>184**	2.71	**>184**	**>184**	3.57	2.44	6.25	2.27	**>52**	1.1 (Tyr)	**29 (Ser)**	**>81 (Val)**
Val[47]	**>189**	1.80			1.94	1.51			**>97**	1.2 (Ile)		
Thr[48]	2.90	3.97			1.50	1.28			1.93	3.1 (Ser)		
Arg[49]	3.44	1.95	**23.5**	**141**	1.42	1.52	1.89	2.43	2.42	1.3 (Lys)	**12 (Gly)**	**58 (Pro)**
Val[50]	**>189**	**12.6**			1.94	**10.4**			**>97**	1.2 (Ile)		
Gln[51]	0.11	1.31	0.70	**>16**	0.44	0.32	0.54	0.58	0.25	4.1 (Glu)	1.3 (Glu)	**>28 (Pro)**
Pro[52]	5.10	1.44			1.75	1.60			2.9	0.90 (Ala)		
Glu[53]	4.67	1.25			1.86	1.43			2.5	0.87 (Gln)		
Pro[55]	**46.3**	**62.3**			**8.60**	**17.8**			5.4	3.5 (Ala)		
Glu[78]	0.26	1.10			1.50	3.00			0.17	0.37 (Glu)		
His[79]	**>183**	**56.3**	**>183**	**>183**	0.63	4.00	1.33	0.29	**>290**	**14 (Asn)**	**>140 (Asp)**	**>630 (Pro)**
Gln[81]	1.62	1.36	0.63	**57**	0.37	0.68	0.32	1.38	4.4	2.0 (Gly)	2.0 (Gly)	**41 (Pro)**
Val[83]	5.60	**85.0**			0.50	2.76			11	**30.8 (Ile)**		
Ser[84]	0.39	1.34			0.39	1.26			1.0	1.1 (Ala)		
Leu[85]	7.38	7.19	**>96**	1.37	5.67	7.00	6.8	0.89	1.3	1.0 (Ile)	**>14 (Pro)**	1.5 (Val)
Leu[86]	**>181**	**171**	**>181**	**>181**	**23.0**	**95**	**17.3**	9.9	**>7.8**	1.8 (Ile)	**>10 (Pro)**	**18 (Val)**
Lys[87]	**24**	0.31	**>144**	4.64	0.35	1.23	0.15	0.88	**69**	0.25 (Arg)	**>960 (Glu)**	5.2 (Thr)
Thr[88]	0.12	0.17			0.15	0.22			0.80	0.77 (Ser)		
Phe[89]	0.74	**169**	**16.5**	3.47	0.86	**95**	2.00	1.41	0.86	1.8 (Tyr)	8.3 (Ser)	2.5 (Val)
Val[93]	**>178**	**84**			6.36	6.31			**28**	**13 (Ile)**		
Leu[95]	**>180**	**23.3**	**>180**	**>180**	**31**	**95**	3.88	**62**	5.8	0.25 (Ile)	**46 (Pro)**	2.9 (Val)

[a] Erbin PDZ residues are numbered as in Fig. 1.

ing that $\beta 3{-}1$(Ile[45]), $\beta 3{-}2$(Phe[46]), $\beta 3{-}4$(Val[47]), and $\beta 3{-}6$(Val[50]) are important for maintaining the $\beta 3$ conformation necessary for tight ligand binding.

The homolog scan data (Fig. 4*B* and Table III) reiterates the view seen from the alanine scan, with the high F values occurring proximal to Val[0], Thr[−1], and Trp[−4] with the caveat that the homologs are generally less disruptive to peptide binding, perhaps as expected from the less dramatic nature of many of the substitutions. One exception to this trend is for Val[83]; substitution with isoleucine ($F = 31$) disfavors binding much more than substitution with alanine ($F = 11$), indicating that the recognition of Thr[−2] is less tolerant of larger side chains at the $\alpha 2{:}4$ position. Despite the homolog mutations being conservative substitutions, a number of them do lead to a significant decrease in display level. The loss of display often involves substituting leucine or valine with isoleucine, implying that

even subtle changes of some hydrophobic core positions can perturb the ability of the domain to fold correctly.

DISCUSSION

The ligand binding and structural studies described herein for Erbin PDZ recapitulate earlier findings on the importance of a ligand C-terminal aliphatic and a −2 position hydroxyl-containing residue for binding to type I PDZ domains (14, 15, 20–22). However, in contrast to these earlier studies, we have also investigated the relative importance of PDZ domain residues in a systematic fashion. The current data show that the hydrophobic core residues surrounding the C-terminal side chain cannot be substituted even conservatively without loss of binding to the phage-derived peptide, thereby providing selectivity. High display levels on phage indicate that these mutants are well folded, suggesting that hydrophobic pockets of varying

Erbin PDZ Domain Structure and Function Analysis
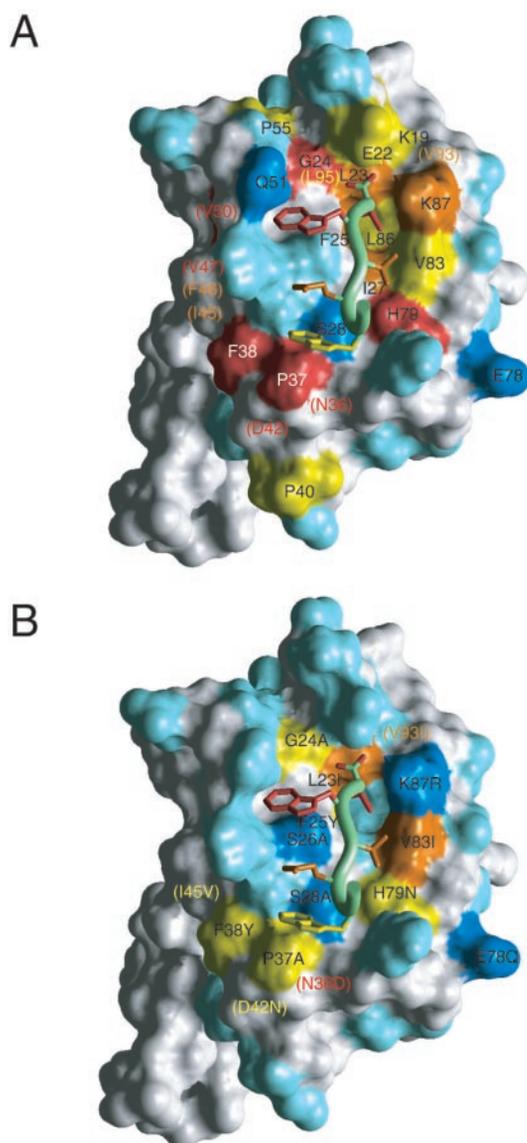
FIG. 4. **Results of the shotgun alanine scan (*A*) and homolog scan (*B*) of Erbin PDZ domain.** The PDZ domain is represented with a Connolly surface, and the peptide ligand is shown in tube form. The protein side chains varied in the libraries are colored *red* ($F > 100$), *orange* ($100 > F > 20$), *yellow* ($20 > F > 5$), *light blue* ($5 > F > 0.5$), and *dark blue* ($F < 0.5$); residues not varied in the libraries are left *white*. All residues with $F > 5$ or < 0.5 are labeled; in cases where the side chain is not visible, the labels are in *parentheses* and colored according to F value.

shape and size can be generated that recognize a variety of C-terminal ligand residues. Thus, libraries similar to those described herein may be used to select PDZ domain sequences that recognize ligands with particular sequences, including C-terminal residues other than valine. Indeed, a computational approach that achieves the same goal has recently been described (43).

Earlier studies of PDZ ligand interactions usually failed to find any preference for residues at the −1 site (20–22), a finding rationalized by early structural studies in which this side was often found to be oriented away from the PDZ domain (Fig. 5*A*) (19, 44, 45). More recently, structures have been published for PDZ domains from α-syntrophin (22) and the H^+/Na^+ exchange regulatory factor (HNERF PDZ1) (48, 49), in which specific hydrophobic or hydrogen bond interactions are observed to the −1 position side chain (Fig. 5). The energetic benefit of this contact in α-syntrophin is unclear because al-
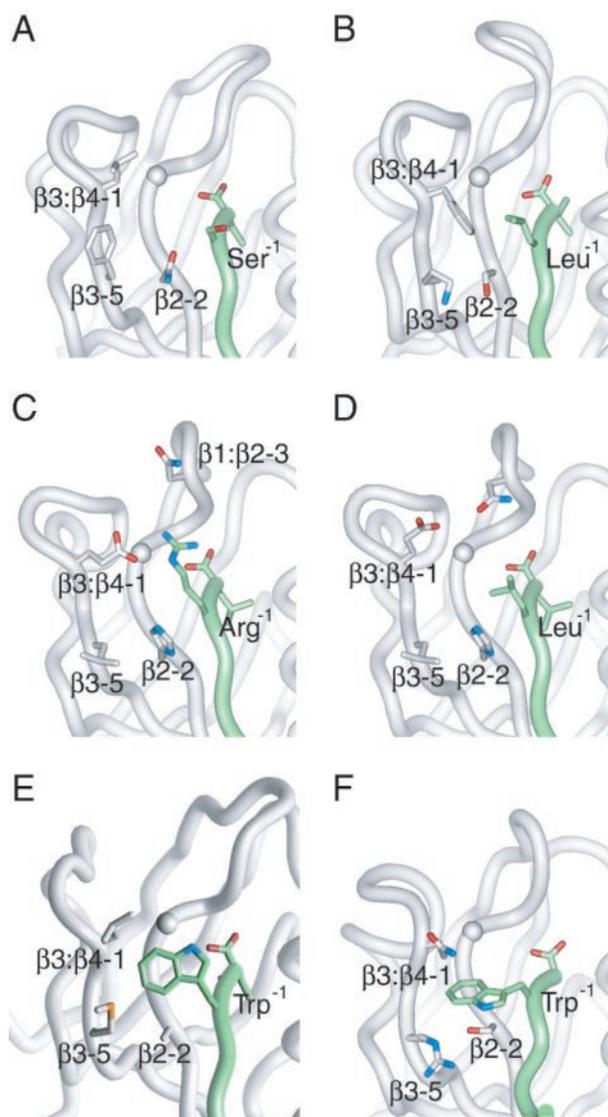
FIG. 5. **Comparison of intermolecular contacts observed with the peptide −1 side chain.** Note that for *B–D*, the exact energetic contribution of the interaction has not been evaluated experimentally. All of the structures are shown in the same orientation after superposition of the common elements of the regular secondary structure. *A*, third PDZ domain of PSD95 with the C-terminal peptide KQTSV from Crypt (Protein Data Bank code 1BE9). *B*, $\beta3{:}\beta4{-}1$(Phe) of α-syntrophin PDZ domain contacts Leu^{-1} of the peptide ligand GVKESLV (the C terminus of the vertebrate voltage-gated sodium channel; Protein Data Bank code 2PDZ). *C*, hydrogen bond formation from the Arg^{-1} side chain of the peptide QDTRL (C terminus of the cystic fibrosis transmembrane conductance regulator) to HNERF side chain $\beta3{:}\beta4{-}1$(Glu) and $\beta1{:}\beta2{-}3$(Asn) backbone carbonyl (via an intervening water molecule; not shown) (Protein Data Bank code 1I92). *D*, HNERF side chains $\beta1{:}\beta2{-}3$(Asn) and $\beta3{:}\beta4{-}1$(Glu) HNERF reorientate to accommodate the Leu^{-1} side chains of the peptide ligand NDSLL (C terminus of the β2-adenergic receptor; Protein Data Bank code 1GQ4). *E*, homology model of Magi3-PDZ2 in which Trp^{-1} of the phage-optimized ligand interacts with $\beta2{-}2$(Ala), $\beta3{-}5$(Met) and $\beta3{:}\beta4{-}1$(Leu) (23). *F*, Trp^{-1} of the phage-derived peptide TGWETWV reaches over $\beta2{-}2$(Ser26) and inserts between $\beta3{-}5$(Arg^{49}) and $\beta3{:}\beta4{-}1$(Gln^{51}) (Protein Data Bank code 1N7T).

most all amino acids are selected at the −1 site from libraries of potential ligands (22). Likewise, the effect of the −1 site of HNERF on affinity is also uncertain because peptide library selection experiments and Western blot overlay studies have shown that it has a preference for ligands with Arg^{-1}, Leu^{-1}, Phe^{-1}, or Tyr^{-1} (46, 47), and ligands with Ala^{-1} also appear to bind well (47). Thus, although these studies point to a compelling structural rationale for recognition of a −1 position ligand

Fig. 6. **Comparison of tryptophan orientation in bound Erbin PDZ ligand and stable β-hairpin peptide.** The peptide ligand and strand β2 of Erbin PDZ (*white*) are overlaid with the N- and C-terminal strands (*light blue*), respectively, of the hairpin peptide bhpW using backbone heavy atoms (root mean square deviation of 0.5 Å for residues Thr^{-2} to Val^0 and β2–2 to β2–4).
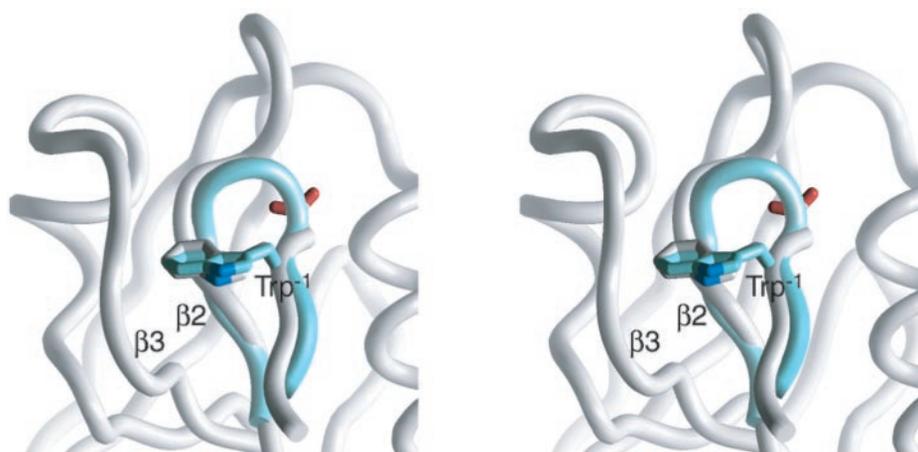
side chain, the absence of precise affinity measurements for mutant proteins or peptide analogs makes the absolute contribution of these interactions hard to gauge.

In contrast to these earlier studies, we have shown a distinct energetic preference for Trp^{-1} in ligands that bind to Magi-3 PDZ2 (23) and Erbin PDZ (Table I). An homology model constructed for Magi-3 PDZ2 suggested that interactions with residues in β2 and β3 might be the source of the favorable contribution to binding by Trp^{-1} (Fig. 5*E*) (23). Although these contacts have been confirmed in the present study of Erbin PDZ (Fig. 5*F*), shotgun alanine scan data indicate that these side chains do not confer any specific energetic contribution to binding (Fig. 4*A*). Thus, the benefit to binding conferred by Trp^{-1} derives from Erbin PDZ domain side chain-independent interactions with the backbone of strands β2 and β3. The orientation of the Trp^{-1} side chain with respect to strand β2 in the Erbin PDZ complex is reminiscent of the interstrand tryptophan contacts observed in recent studies of peptide β-hairpin stability (50–52). These studies indicated that tryptophan is best able to stabilize antiparallel interactions between two β-strands regardless of the residue type on the opposite strand (50). The similarity in backbone and tryptophan conformation in these two cases (Fig. 6) suggests a common mechanism for stabilization (either tight ligand binding in the PDZ case or β-hairpin stability in the peptide case) based on the side chain to backbone contacts. Thus, a Trp^{-1} residue may be a general and somewhat nonselective means to increase the affinity of C-terminal peptides for PDZ domains. A further degree of positive selection may be garnered by making the site around Trp^{-1} more hydrophobic, as observed in the case of Magi-3 PDZ2 (Fig. 5*E*) (23), in PDZ domains engineered to recognize hydrophobic peptides (43), and in the alanine scan results of Erbin PDZ (Fig. 4). Conversely, selection against Trp^{-1} ligands may be achieved by the inclusion of large bulky residues in the vicinity of the −1 site that might obstruct tryptophan-strand interactions.

The C-terminal phage display process selected only acidic residues at the −3 site (Table I). Structural analysis indicated the presence of a proximal basic residue at position β3–5(Arg49) (Fig. 3*B*), suggesting that a favorable Coulombic interaction is the cause of the 140-fold decrease in affinity for the Glu^{-3} to Ala^{-3} substitution. Curiously, mutagenesis of Erbin PDZ residues close to the Glu^{-3} location had very little effect on peptide binding (Fig. 4). One possible explanation is that given the interdigitation of side chains from the ligand and β3 (Fig. 3*B*) and the contacts between Glu^{-3} and Trp^{-4}, replacement of the Glu^{-3} side chain may decrease the ability of other ligand side chains to make optimal contacts with Erbin PDZ. In contrast, contacts between Erbin PDZ and Trp^{-4} (Fig. 3*B*) suggest

a structural rationale for the selection of aromatic residues at this site that is born out by the alanine scanning mutagenesis results (Fig. 4). The opportunity for these interactions to arise is made possible by the much longer, but still structured, β2:β3 loop in Erbin (15 residues) compared with other PDZ domains (usually 4–6 residues; Fig. 1). Utilization of an enlarged β2:β3 loop to provide additional ligand selectivity has been noted previously (38, 53). In addition to the β2:β3 residues that contact ligand, the alanine and homolog scan data reveal that several more are necessary for tight ligand binding (*e.g.* Asn^{36} and Asp^{42}; Fig. 4). The β2:β3 loop wraps around the side chain of the Asn^{36} (Fig. 3*B*) with potential hydrogen bonds to the backbone of residues Arg^{39}, Asp^{42}, and Gly^{44}, suggesting an important role for Asn^{36} in stabilizing the β2:β3 loop in a conformation competent for favorable interaction with the ligand.

On the basis of the optimal ligand identified by phage display, we have previously suggested that the interaction between Erbin and the family of p120-like catenins may be more physiologically relevant than the earlier postulated interactions with ERB-B2 (24, 27). The present, more expansive investigations yielded the same optimal ligand and substantiate our original hypothesis of relevant protein ligands for Erbin, and yeast two hybrid studies have also identified such an interaction (28, 29). The detailed structure-activity relationships discussed above allow us to make hypotheses about potential ligands for other mammalian LAP protein PDZ domains. Densin-180 is the most similar in primary sequence to Erbin PDZ (61% identity), with conservation of all of the side chains that contribute to ligand binding/selectivity described above. In accord with this, we have previously shown that the optimal ligand for Erbin PDZ is very similar to that for Densin-180 PDZ (24), and Densin-180 has been shown to co-localize with p120-catenins at neuronal synaptic junctions (54). The other mammalian LAP protein, Scribble, contains four PDZ domains and is also involved in maintenance of epithelial cell polarity, the formation of multi-protein membrane complexes, and the control of epithelial cell growth (26). All four domains have 30–40% sequence identity to Erbin PDZ (Fig. 1). The pocket at the 0 site is hydrophobic in all four cases, although β2–1(Phe^{25}) is conserved only in PDZ2 of Scribble, and the smaller aliphatic side chains present in Scribble PDZ1, PDZ3, and PDZ4 suggest that these domains may accommodate larger C-terminal ligand residues. Variability at β3:β4–1(Gln^{51}) suggests that the preference for Trp^{-1} may differ from that observed in Erbin PDZ. The residues at α2–1(His^{79}), α2–5(Val^{83}), and β3–5(Arg^{49}) are identical or very similar in all cases, suggesting a conserved preference for Thr^{-2} and either Glu^{-3} or Asp^{-3}. The β2-β3 loop is long in all cases, suggesting that

favorable contacts with the -4 ligand residue will be possible, although sequence variability in the loop makes it difficult to predict the preferred ligand residue. Thus, the consensus ligands for the human Scribble domains are likely to be X(D/ E)(T/S)XV for domain 2 and XDETS$X\phi$ for the other three domains (where X indicates specificity of an as yet undefined nature and ϕ is a large hydrophobic residue). Further studies are in progress with these and other PDZ domains to identify optimal ligands via C-terminal phage display so that potential binding partners may be ascertained.

In summary, we have used an expanded C-terminal phage display library to confirm and extend our earlier proposal of an optimal and biologically significant ligand for the Erbin PDZ domain (24). Importantly, the energetic contributions to binding of the side chains within this optimal ligand have been ascertained by binding affinity measurements with a large number of synthetic peptide analogs. All five C-terminal ligand residues are found to make a beneficial contribution to binding. Structural studies have identified the subset of Erbin PDZ residues that contact ligand side chains, and an efficient combinatorial scanning approach has been used to investigate the origins of affinity and selectivity within PDZ domains. A novel interaction with Trp^{-1} has been observed that is likely to be a generic method of stabilizing the interaction between C-terminal peptides and many PDZ domains. The particular conformation adopted by the long $\beta2$:$\beta3$ loop of Erbin PDZ also allows additional contacts with ligand residues at the -4 site. The combination of these investigations thus gives deep insight into the manner by which PDZ domains can recruit their particular cellular targets with both high affinity and selectivity. These results have been extended to make hypotheses about ligands that other LAP PDZ domains will recognize. Further applications of the experimental techniques described herein will be used to confirm or refute these hypotheses and to add to our growing knowledge of the manner by which PDZ domains participate in interactions of biological significance.

REFERENCES

1. Eisenberg, D., Marcotte, E. M., Xenarios, I., and Yeates, T. O. (2000) *Nature* **405,** 823–826
2. Hazbun, T. R., and Fields, S. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 4277–4278
3. Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S. M., and Eisenberg, D. (2002) *Nucleic Acids Res.* **30,** 303–305
4. Xenarios, I., and Eisenberg, D. (2001) *Curr. Opin. Biotechnol.* **12,** 334–339
5. Pawson, T., and Scott, J. D. (1987) *Science* **278,** 2075–2080
6. Kuriyan, J., and Cowburn, D. (1997) *Annu. Rev. Biophys. Biomol. Struct.* **26,** 259–288
7. Lee, C. H., Cowburn, D., and Kuriyan, J. (1998) *Methods Mol. Biol.* **84,** 3–31
8. Cowburn, D. (1996) *Structure* **4,** 1005–1008
9. Cho, K. O., Hunt, C. A., and Kennedy, M. B. (1992) *Neuron* **9,** 929–942
10. Woods, D. F., and Bryant, P. J. (1993) *Mech. Dev.* **44,** 889
11. Kim, E., Niethammer, M., Rothschild, A., Jan, Y. N., and Sheng, M. (1995) *Nature* **378,** 85–88
12. Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000) *Nucleic Acids Res.* **28,** 231–234
13. Morais Cabral, J. H., Petosa, C., Sutcliffe, M. J., Raza, S., Byron, O., Poy, F., Marfatia, S. M., Chishti, A. H., and Liddington, R. C. (1996) *Nature* **382,** 649–652
14. Kornau, H. C., Schenker, L. T., Kennedy, M. B., and Seeburg, P. H. (1995) *Science* **269,** 1737–1740
15. Matsumine, A., Ogai, A., Senda, T., Okumura, N., Satoh, K., Baeg, G. H., Kawahara, T., Kobayashi, S., Okada, M., Toyoshima, K., and Akiyama, T. (1996) *Science* **272,** 1020–1023
16. Niethammer, M., Kim, E., and Sheng, M. (1996) *J. Neurosci.* **16,** 2157–2163
17. Sheng, M., and Sala, C. (2001) *Annu. Rev. Neurosci.* **24,** 1–29
18. Harris, B. Z., and Lim, W. A. (2001) *J. Cell Sci.* **114,** 3219–3231
19. Doyle, D. A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. (1996) *Cell* **85,** 1067–1076
20. Songyang, Z., Fanning, A. S., Fu, C., Xu, J., Marfatia, S. M., Chishti, A. H., Crompton, A., Chan, A. C., Anderson, J. M., and Cantley, L. C. (1997) *Science* **275,** 73–77
21. Stricker, N. L., Christopherson, K. S., Yi, B. A., Schatz, P. J., Raab, R. W., Dawes, G., Bassett, D. E., Jr., Bredt, D. S., and Li, M. (1997) *Nat. Biotech.* **15,** 336–342
22. Schultz, J., Hoffmuller, U., Krause, G., Ashurst, J., Macias, M. J., Schmieder, P., Schneider-Mergener, J., and Oschkinat, H. (1998) *Nat. Struct. Biol.* **5,** 19–24
23. Fuh, G., Pisabarro, M. T., Li, Y., Quan, C., Lasky, L. A., and Sidhu, S. S. (2000) *J. Biol. Chem.* **275,** 21486–21491
24. Laura, R. P., Witt, A. S., Held, H. A., Gerstner, R., Deshayes, K., Koehler, M. F., Kosik, K. S., Sidhu, S. S., and Lasky, L. A. (2002) *J. Biol. Chem.* **277,** 12906–12914
25. Bryant, P. J., and Huwe, A. (2000) *Nat. Cell Biol.* **2,** E141–E143
26. Bilder, D., and Perrimon, N. (2000) *Nature* **403,** 676–680
27. Borg, J. P., Marchetto, S., Le Bivic, A., Ollendorff, V., Jaulin-Bastard, F., Saito, H., Fournier, E., Adelaide, J., Margolis, B., and Birnbaum, D. (2000) *Nat. Cell Biol.* **2,** 407–414
28. Jaulin-Bastard, F., Arsanto, J. P., Le Bivic, A., Navarro, C., Vely, F., Saito, H., Marchetto, S., Hatzfeld, M., Santoni, M. J., Birnbaum, D., and Borg, J. P. (2002) *J. Biol. Chem.* **277,** 2869–2875
29. Izawa, I., Nishizawa, M., Tomono, Y., Ohtakara, K., Takahashi, T., and Inagaki, M. (2002) *Genes Cells* **7,** 475–485
30. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A., and Sidhu, S. S. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 8950–8954
31. Sidhu, S. S., Lowman, H. B., Cunningham, B. C., and Wells, J. A. (2000) *Methods Enzymol.* **328,** 333–363
32. Sidhu, S. S., Weiss, G. A., and Wells, J. A. (2000) *J. Mol. Biol.* **296,** 487–495
33. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154,** 367–382
34. Vajdos, F. F., Adams, C. W., Breece, T. N., Presta, L. G., de Vos, A. M., and Sidhu, S. S. (2002) *J. Mol. Biol.* **302,** 415–428
35. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (1995) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, New York
36. Pan, B., Li, B., Russell, S. J., Tom, J. K., Cochran, A. G., and Fairbrother, W. J. (2002) *J. Mol. Biol.* **316,** 769–787
37. Starovasnik, M. A., Christinger, H. W., Wiesmann, C., Champe, M. A., de Vos, A. M., and Skelton, N. J. (1999) *J. Mol. Biol.* **293,** 531–544
38. Tochio, H., Hung, F., Li, M., Bredt, D. S., and Zhang, M. (2000) *J. Mol. Biol.* **295,** 225–237
39. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* **13,** 289–302
40. Hansen, M. R., Mueller, L., and Pardi, A. (1998) *Nat. Struct. Biol.* **5,** 1065–1074
41. Tjandra, N., and Bax, A. (1997) *Science* **278,** 1111–1114
42. Aasland, R., Abrams, C., Ampe, C., Ball, L. J., Bedford, M. T., Cesareni, G., Gimona, M., Hurley, J. H., Jarchau, T., Lehto, V. P., Lemmon, M. A., Linding, R., Mayer, B. J., Nagai, M., Sudol, M., Walter, U., and Winder, S. J. (2002) *FEBS Lett.* **513,** 141–144
43. Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S., Serrano, L., and Gonzalez, C. (2002) *Nat. Struct. Biol.* **9,** 621–627
44. Daniels, D. L., Cohen, A. R., Anderson, J. M., and Brunger, A. T. (1998) *Nat. Struct. Biol.* **5,** 317–325
45. Tochio, H., Zhang, Q., Mandal, P., Li, M., and Zhang, M. (1999) *Nat. Struct. Biol.* **6,** 417–421
46. Wang, S., Raab, R. W., Schatz, P. J., Guggino, W. B., and Li, M. (1998) *FEBS Lett.* **427,** 103–108
47. Hall, R. A., Ostedgaard, L. S., Premont, R. T., Blitzer, J. T., Rahman, N., Welsh, M. J., and Lefkowitz, R. J. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 8496–8501
48. Karthikeyan, S., Leung, T., and Ladias, J. A. (2001) *J. Biol. Chem.* **276,** 19683–19686
49. Karthikeyan, S., Leung, T., and Ladias, J. A. (2002) *J. Biol. Chem.* **277,** 18973–18978
50. Russell, S. J., and Cochran, A. G. (2000) *J. Am. Chem. Soc.* **122,** 12600–12601
51. Cochran, A. G., Skelton, N. J., and Starovasnik, M. A. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 9081
52. Cochran, A. G., Tong, R. T., Starovasnik, M. A., Park, E. J., McDowell, R. S., Theaker, J. E., and Skelton, N. J. (2001) *J. Am. Chem. Soc.* **123,** 625–632
53. Kozlov, G., Banville, D., Gehring, K., and Ekiel, I. (2002) *J. Mol. Biol.* **320,** 813–820
54. Izawa, I., Nishizawa, M., Kazuhiro, K., and Inagaki, M. (2002) *J. Biol. Chem.* **277,** 5345–5350