Computer-aided design of a PDZ domain to recognize new target sequences

Jose Reina[1], Emmanuel Lacroix[2], Scott D. Hobson[3], Gregorio Fernandez-Ballester[2], Vladimir Rybin[3], Markus S. Schwab[3], Luis Serrano[2] and Cayetano Gonzalez[1]

Published online: 24 June 2002, doi:10.1038/nsb815

PDZ domains are small globular domains that recognize the last 4–7 amino acids at the C-terminus of target proteins. The specificity of the PDZ–ligand recognition is due to side chain–side chain interactions, as well as the positioning of an α-helix involved in ligand binding. We have used computer-aided protein design to produce mutant versions of a Class I PDZ domain that bind to novel Class I and Class II target sequences both *in vitro* and *in vivo*, thus providing an alternative to primary antibodies in western blotting, affinity chromatography and pull-down experiments. Our results suggest that by combining different backbone templates with computer-aided protein design, PDZ domains could be engineered to specifically recognize a large number of proteins.

Protein identification and purification are essential processes in modern molecular and cell biology research, and require the use of lengthy procedures, such as protein tagging or raising antibodies. A possible alternative to such classical approaches would be to randomize the interaction surface of one or more protein-binding domains and then screen for those mutants that recognize each individual molecule of interest[1]. However, such an approach requires the screening of millions of randomized mutants. To accelerate this process, rational design could be introduced at an early step so that only a small number of likely solutions would have to be tested. However, computational methods require precise three-dimensional information on both the binding domain and the protein to be bound, thus limiting their applicability.

To circumvent this limitation, nonstructured stretches in the target protein sequence that adopt a particular conformation upon interacting with the binding domain could be exploited[2,3]. This condition is frequently met by C-termini, which are nonstructured and solvent accessible in most proteins[4]. Moreover, the C-termini of proteins from several organisms, including *Escherichia coli*, *Saccharomyces cerevisiae* and *Homo sapiens*, have complex sequences such that the last four amino acid residues are sufficient to uniquely specify ~74–97% of the proteins in these organisms[5]. To explore the possibility of using the C-termini as targets for antibody binding, we engineered a PDZ domain that recognizes a Class I C-terminal motif (X(T/S)X(L/V)) to recognize other Class I and Class II (X-Φ-X-Φ, where Φ represents a hydrophobic residue) motifs[6,7].

Selection of the PDZ domain

Superposition of three of the available PDZ structures shows good structural homology between PDZ domains that recognize an aliphatic (Class II) or a Thr/Ser (Class I) residue at position p(-2) (positions in the ligand peptide are referred to by p followed by a negative number starting from the C-terminal residue, p0) but less homology between Class III- and Class II-binding PDZ domains (Fig. 1). These structural differences

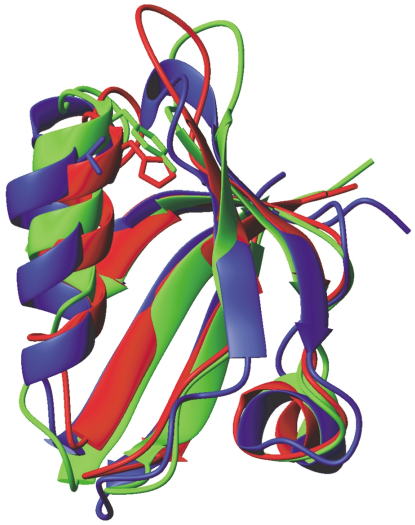


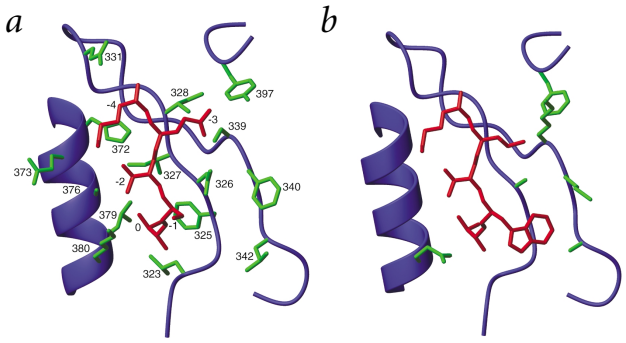
Fig. 1 Superposition of the ribbon diagrams of the third PDZ domain of PSD-95 (PDB entry 1BE9; red), CASK/Lim2 (1KWA; blue) and the PDZ of nNOS (1B8Q; green).

are mainly due to the displacement of the α-helix involved in peptide binding with respect to the β-sheet (Fig. 1). Such displacement is related to changes in specificity in cases similar to the PDZ domain of neuronal nitric oxide synthase (nNos), in which Tyr 77 at the N-terminus of the α-helix makes a hydrogen bond to the side chain of the Asp residue at p(-2) (ref. 8). In other more 'classical' PDZ domains with Ser/Thr at p(-2), such as PSD-95, a His at the equivalent position will not allow binding because of steric clashes. Likewise, PDZ domains similar to CASK/Lin[9], which recognizes an aromatic residue at position p(-2), also show important displacements of the α-helix to accommodate the bulky side chain.

Based on these observations we decided to perform our design exercise with a Class I PDZ that could also be Class II

[1]Cell Biology and Biophysics Program and [2]Structural Biology and Biocomputing Program, EMBL, 69117 Heidelberg, Germany. [3]CellZome A.G., Meyerhofstrasse 1, 69117 Heidelberg, Germany.

Correspondence should be addressed to C.G. *email: gonzalez@embl-heidelberg.de*



e

	P_{-4}	P_{-3}	P_{-2}	P_{-1}	P_0
Class I	X	X	-T/S-	X	Φ
Class II	X	X	Φ	X	Φ
Eg5	T	S	I	N	L
Hyd	K	I	T	W	V
Pol	K	R	T	E	V
Cript(wt)	K	Q	T	S	V
NL	S	T	T	R	V

Fig. 2 Ribbon diagrams of the PDZ–ligand complexes. ***a***,PSD-95 PDZ3 (PDZ-wt) and its natural ligand (KQTSV). ***b***, PDZ-hyd–hyd peptide (KITWV). ***c***, PDZ-pol–pol peptide (KRTWV). ***d***, PDZ-Eg5–Eg5 peptide (TSINL). The residues of the ligand (red), as well as those selected for mutagenesis (green), are numbered in (*a*). Only the ligand (red) and the mutations suggested by Perla (green) are shown in (*b–d*). ***e***, Alignment of the target sequences discussed in this work, including the Class I and Class II consensus, Eg5, hyd, pol and the two sequences known to bind the wild type domain, CRIPT (refered to in this work as wt peptide) and NL.

compatible. Therefore, we chose the PDZ/3 domain of the PSD-95 protein[10] (PDZ-wt), whose structure has been solved by X-ray crystallography at high resolution. This domain recognizes two class I ligands, NL (Ser-Thr-Thr-Arg-Val) and CRIPT (Lys-Gln-Thr-Ser-Val; 'wt peptide'), which contain Thr and Val at positions p(-2) and p0, respectively[11,12]. Mutation of the residue p(-1) of each of these peptides (Arg or Ser) into Asp abolishes binding to the PDZ domain[11].

Selection of the target

We have carried out two different design exercises that include changing the specificity of PDZ-wt to recognize a Class II and then a new Class I target peptide. The target sequence chosen in the first case was the C-terminal end of the kinesin-like molecule, Eg5 (Thr-Ser-Ile-Asn-Leu, Eg5 peptide)[13]. This sequence fits the consensus for a Class II target, and amino acids p(-1)–p(-4) are not present in the two known targets of PDZ-wt. The second exercise was to create new PDZ domains that could specifically recognize the amino acids located at positions p(-1) and p(-3). To test the efficiency of the design algorithm, we challenged it with amino acids that have different chemical properties: two hydrophobic residues at p(-1) and p(-3) in one target (Lys-Ile-Thr-Trp-Val; wt-hyd peptide) and two charged residues (Lys-Arg-Thr-Glu-Val; wt-pol peptide) in the other.

Computer-based design

The design was based on the automated computer algorithm Perla[14,15], which enables the identification and sorting of amino acid sequences that have optimal stability for a desired three-dimensional structure. The program works by evaluating a free-energy scoring function that uses the ECPP2 all-atom molecular mechanics force field and a combination of statistical terms, including entropy (derived from the protein database) and solvation (based on accessible surface approximations (ASA)). For PDZ domains and their bound peptides, the design process involves several steps. First, the residues of the PDZ domain that could establish specific interactions with the ligand must be selected. Second, the side chains that contact the positions to be mutagenized must be identified to allow for side chain movements that are necessary to accommodate the new residues introduced by the algorithm. Perla automatically selects these residues based on a geometrical approximation that takes into account Cα–Cα distances and the angle between Cα–Cβ vectors from the two residues being analyzed. Third, the algorithm places the natural amino acids at each position selected for mutagenesis and eliminates those side chain conformations and amino acids that are not compatible with the rest of the structure. Fourth, all possible pairwise interactions are explored to eliminate combinations that are less favorable. Finally, an output of sequences and PDB coordinates ranked in terms of free energy is produced. When working with a protein complex, the operator must take care to select sequences in which inter- rather than intramolecular interactions have been optimized by the algorithm.

When the number of amino acid combinations obtained is too high, a first round of design is done to select the best 10–20 sequences. A second round is then carried out, taking into account only the amino acids found in the first round so that the number of combinations is reduced.

Computer-based engineering of PDZ domain

The template used for this exercise was the structure of PDZ-wt in complex with wt peptide (KQTSV; PDB entry 1BE9) (Fig. 2). The following residues forming the binding interface (defined by visual inspection of the three-dimensional structure) were considered for mutation: Leu 323, Phe 325, Asn 326, Ile 327, Ile 328, Glu 331, Ser 339, Phe 340, Leu 342, His 372, Glu 373, Ala 376, Leu 379, Lys 380 and Tyr 397 (Fig. 2*a*). Positions 339, 340 and 397 were not mutated for binding the Eg5 target because the side chains at positions p(-1) and p(-3) of the ligand are short (Asn and Ser) and these residues are far away from the ligand (Fig. 2*d*). For recognizing the wt-pol and wt-hyd sequences, the number of positions explored was reduced because positions p0, p(-2) and p(-4) of the wild type sequence remained unchanged. However, because p(-1) (Trp in wt-hyd and Glu in wt-pol) and p(-3) (Ile in wt-hyd and Arg in wt-pol) are large residues, positions 339, 340 and 397 of the PDZ domain were included for mutagenesis. Thus, positions 326, 339, 340, 342, 380 and 397 were mutated to recognize

Table 1 Mutations considered and selected by Perla

Residue	Mutations considered	Best fit (1st round)	Best fit (2nd round)
Eg5 (TINSL)			
Leu 323	A,V,I,L,F	F, L	F,L
Phe 325	A,V,I,L,F	F,L	F,L
Asn 326	S,T,N,Q,D,E,K,R	K,T	K,T
Ile 327	A,V,I,L,F	I,L	I,F
Ile 328	S,T,N,Q,D,E,K,R	K,T	T,K
Glu 331	D,E	E	E
Leu 342	S,T,N,Q,D,E,K,R	K,R	K,R
His 372	A,V,I,L,F	A,L	A,L
Glu 373	N,Q,D,E,A,V,I,L,F	E,I	E,I
Ala 376	A,V,I,L,F	A,V	A,V
Leu 379	A,V,I,L,F	L,F,I	L,F,I
Lys 380	N,Q,D,E,K,R	K,R	K,R
hyd (KITWV)			
Asn 326	A,C,D,E,F,H,I,K,L,M,N,Q,R,S,T,V,W,Y	A	A
Ile 328	A,C,D,E,F,H,I,K,L,M,N,Q,R,S,T,V,W,Y	I	I
Glu 331	A,C,D,E,F,H,I,K,L,M,N,Q,R,S,T,V,W,Y	E	E
Ser 339	A,C,D,E,F,H,I,K,L,M,N,Q,R,S,T,V,W,Y	S,T,K,M,L,V,I,C	K,T
Phe 340	A,C,D,E,F,H,I,K,L,M,N,Q,R,S,T,V,W,Y	M	M
Leu 342	A,C,D,E,F,H,I,K,L,M,N,Q,R,S,T,V,W,Y	A	A
Glu 373	A,C,D,E,F,H,I,K,L,M,N,Q,R,S,T,V,W,Y	E	E
Lys 380	A,C,D,E,F,H,I,K,L,M,N,Q,R,S,T,V,W,Y	E	E
Tyr 397	A,C,D,E,F,H,I,K,L,M,N,Q,R,S,T,V,W,Y	I,F	F
pol (KRTEV)			
Asn 326	K,N,R,Q,S,T,E	E	
Ile 328	K,N,R,Q,S,T,E	T	
Ser 339	K,N,R,Q,S,T,E,A,V	V	
Phe 340	K,Q,R,N,D,E	R	
Leu 342	K,Q,R,E,	R,E	
Tyr 397	F	F	

wt-hyd (Fig. 2*b*); the same set of residues, except residue 380, were mutated to recognize wt-pol (Fig. 2*c*; Table 1). In the designs of the PDZ domains that bind the Eg5, hyd and pol peptides (PDZ-Eg5, PDZ-hyd and PDZ-pol), ~2.3×10^9, 3.9×10^{22} and 1.4×10^4 different sequences, respectively, were considered in the design (Table 1).

Description of the mutations

The following mutations were introduced to construct PDZ-Eg5: F325L, N326T, I328K, L342K, H372L, E373I and L379I. Phe 325 and Leu 379 are in the hydrophobic core of the PDZ domain and interact with the residue at position p0. By mutating them into a Leu and an Ile, respectively, we created a larger cavity to accommodate the Leu at p0. Asn 326 can make a hydrogen bond with Ser at p(-1) of the wild type peptide but cannot when Ser p(-1) is replaced by Asn. Perla placed a Thr at position 326 to make a hydrogen bond with the side chain of the Asn located at p(-1). The side chain of His 372 interacts *via* a hydrogen bond with Thr at position p(-2) of the wt-Eg5. Mutation of Thr p(-2) into Ile requires an aliphatic residue in PDZ to prevent a solvation penalty. This mutation is the main determinant in switching from a class I to a class II ligand. Ile 328 makes hydrophobic contacts with Gln at p(-3) of wt-Eg5. Replacement of Gln at p(-3) by Ser abolishes this interaction and desolvates the Ser side chain. Lys 328 in PDZ-Eg5 allows for better solvation of the Ser residue. Leu 342 in the PDZ-wt was mutated into Lys to allow this side chain to make a hydrogen bond to the side chain of Asn at p(-1) in the Eg5 ligand. Glu 373 makes a favorable electrostatic interaction with Lys at position p(-4). The program finds Ile at 373 as a favorable mutation because of the packing with the side chain Cγ2 methyl of Thr at position p(-4) in the Eg5 ligand.

The designed PDZ-hyd sequence contains the following mutations: N326A, S339K, F340M, L342A, K380E and Y397F. Asn 326 and Leu 342 were mutated to Ala and Phe 340 to Met to accommodate the bulky side chain of the Trp residue at p(-1) of the ligand. To accommodate Ile at p(-3) of the ligand, the algorithm suggested mutating Ser 339 and Tyr 397 into Lys and Phe, respectively. The side chain of Lys 339 makes favorable van der Waals and hydrophobic contacts with Ile at p(-3), as well as with the side chain of Phe 397. Phe is chosen at 397 because a Tyr would have been unfavorably desolvated by the aliphatic chain of Lys 339. Lys 380 was mutated to Glu because it improves the general electrostatics of the protein. This mutation is an example of residues introduced to optimize the stability of the protein rather than that of the complex.

PDZ-pol has the following mutations: N326E, I328T, S339V, F340R, L342R and Y397F. Some compensatory mutations, Arg 340 and 342, were introduced by Perla to interact with the charged residues at positions p(-1) (Glu) and p(-3) (Arg) of the ligand. As in the previous case, two mutations at positions 339 and 397 were introduced; Val 339 interacts favorably with the aliphatic side chain of Arg at position p(-3), whereas the Phe 397 side chain packs nicely against Val. The thermal stability and structure of such re-designed PDZs, as determined by CD, remains largely unchanged (data not shown).

Binding affinity and specificity of designed PDZ domains

To quantify the interactions between the different PDZ domains and their ligands, we determined the binding affinity (K_d) of these complexes by fluorescence polarization assays. The K_d values of PDZ-wt, PDZ-Eg5, PDZ-pol and PDZ-hyd for their corresponding ligands were found to be 20 ± 1, 96 ±11, 26 ± 3 and 0.1 ± 0.01 μM, respectively (Fig. 3). The estimated K_d of the PDZ-wt–wt peptide complex is similar to published values[11]. Thus, two of the engineered PDZ domains have affinities for their targets that are on the same order of magnitude as PDZ-wt for its natural ligand, whereas the affinity of the third, PDZ-hyd, is about two orders of magnitude higher.

To quantify the specificity of binding between each of the engineered PDZs and their ligands, we carried out fluorescence polarization assays in the presence of unlabeled competing peptides (Table 2). The best competitors for the binding between PDZ-wt and the wt peptide (wt-hyd and wt-pol) are about one order of magnitude less efficient than the wt peptide itself, and wt-Eg5 cannot compete at all. The efficiency of

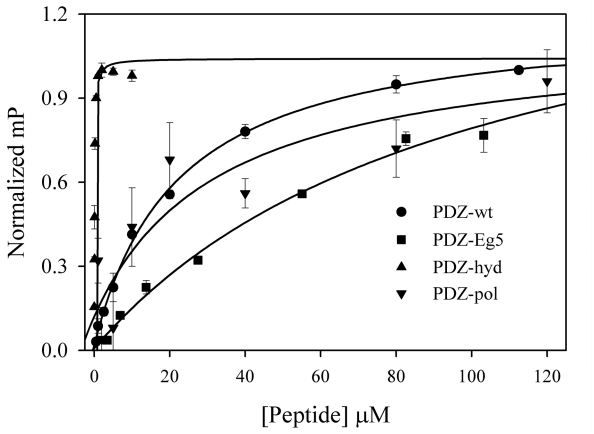


Fig. 3 Fluorescence polarization binding assays. For the saturation curves, each fluorescently labeled peptide at a concentration of 5 nM was incubated with increasing amounts of the corresponding PDZ domain, and the change in polarization was measured (mP). The data are normalized to the highest value for each experiment. The estimated K_d of each complex is listed in Table 2. The experiment was repeated twice, except for PDZ-pol, which was repeated four times because of poor signals. In this case, a higher concentration of 160 μM protein was also measured, which was consistent with the other measurements (not shown). Error bars correspond to the difference in mP between the average an individual values.

the wt peptide to compete with the binding of the designed PDZ domains to their ligands is also less than one order of magnitude in the case of PDZ-Eg5 and PDZ-pol, and several orders of magnitude in the case of PDZ-hyd. These results suggest that binding between the four PDZs presented in this work and their corresponding ligands is specific. This conclusion is further substantiated by the other binding assays described below.

In vivo binding of designed PDZ domains

To determine whether the designed PDZ domains were able to bind to their new targets *in vivo* and to further estimate the specificity of such interactions, we carried out a series of two-hybrid analyses[16]. To avoid protein–protein interactions with residues outside the C-termini, octapeptides carrying the last five C-terminal amino acids of each target peptide and a short spacer sequence were fused to a carrier protein (GFP). The resulting fusions, GFP–wt, GFP–Eg5, GFP–hyd and GFP–pol, as well as GFP as a control, were then fused C-terminal to the DNA-binding domain of Gal4. These constructs were tested in all pairwise combinations against constructs carrying PDZ-wt, PDZ-Eg5, PDZ-hyd and PDZ-pol fused to the activation domain of Gal4. Essentially, every designed PDZ domain recognized its target sequence with high affinity (Table 3), as shown by the number of viable colonies. PDZ-wt did not recognize any of the three new targets, Eg5, hyd and pol. Only a minor cross-reaction was observed between PDZ-hyd and GFP-wt.

Binding assays of engineered PDZ domains

We then tested the performance of the designed PDZ domains in overlay, affinity purification and pull-down assays. We found that the designed PDZ sequences were able to specifically recognize their target in an overlay assay using whole extracts from *E. coli* expressing the corresponding GFP–peptide fusion that were blotted on a nitrocellulose membrane (Fig. 4). In all cases, the background level was low and we could not detect any unspecific bands because of binding to *E. coli* proteins. Moreover, cross-reactivity between the PDZs and target sequences assayed was also found to be extremely low.

To determine whether the engineered PDZ could be used as a tool for affinity purification, we followed two approaches. First, we assayed the ability of PDZ-wt, PDZ-Eg5, PDZ-hyd and PDZ-pol that were immobilized on a solid phase to affinity purify their target peptides in extracts made from *E. coli* overexpressing GFP, GFP–wt, GFP–Eg5, GFP–hyd or GFP-pol (Fig. 5*a*). We found that each PDZ binds specifically to the corresponding GFP-fused ligand, thus providing further evidence substantiating our conclusion that the designed PDZ domains efficiently recognize their intended peptide. These results suggest that such engineered PDZs could be used to purify the target protein from complex cell extracts.

Finally, we tested the ability of PDZ-Eg5 to pull down Eg5 protein expressed in mammalian cells. We transiently expressed Eg5 kinesin and a mutant lacking the five C-terminal amino acids of Eg5, both tagged with Protein A-FLAG in HEK 293 cells. The cells were lysed and the tagged Eg5 protein was precipitated with purified recombinant GST–PDZ-Eg5 protein. PDZ-Eg5 precipitated the expressed Eg5 fusion protein (Fig. 5*b*, lanes 1–3), whereas it did not precipitate detectable amounts of the mutant form of Eg5 (Fig. 5*b*, lanes 4–6). Interestingly, the amount of Eg5 protein pulled down by PDZ-Eg5 is only slightly less than the amount pulled down by binding the Protein A-tag of the fusion protein with IgG–agarose beads (Fig. 5*b*, lanes 7–9). The interaction of PDZ-Eg5 with its target protein was resistant to extended washing even in the presence of 500 mM NaCl (data not shown). These results demonstrate a specific, stable interaction of PDZ-Eg5 with its target protein expressed in an established eukaryotic cell line.

Table 2 Affinity and specificity of the binding between the PDZ domains and their target peptides[1]

PDZ	K_d (μM)	Specificity (K_i wt / K_i target)			
		wt	Eg5	hyd	pol
pol	26 ± 3.20	14	N.D.	N.D.	1
hyd	1 ± 0.008	90	N.D.	1	N.D.
Eg5	96 ± 11.6	9	1	N.D.	N.D.
wt	20 ± 1.5	1	>100	5	9

[1]Binding of each of the other PDZ domains to their ligands was competed with the ligand itself (as a control) and the wt peptide. The best competitors are at least an order of magnitude less efficient than the corresponding peptide, suggesting that binding between the four PDZs presented in this work and their corresponding ligands is specific.

PDZ–ligand binding specificity

We have shown that an automated computer algorithm can be used to change the ligand specificity of a PDZ domain. These mutant PDZ domains can be used as protein recognition tools in assays that are normally carried out with antibodies, such as western blots, affinity chromatography and pull-down experiments. In principle, this strategy could be applied to target many proteins of known sequence in which the last 4–5 C-terminal amino acids are solvent exposed and unstructured, a condition met by the majority of proteins whose structures have been solved. Binding between such engineered PDZs and their targets is expected to be highly specific because 4–5 amino acids provide a sequence tag that is unique to most pro-

Table 3 Two-hybrid interactions between the PDZ domains and their target ligands. A full matrix of all four PDZs and ligands plus controls is shown.

	pGBK[1]	GFP	GFP–wt	GFP–Eg5	GFP–pol	GFP–hyd
pGAD[1]	–[2]	–	–	–	–	–
PDZ–wt	–	–	+++	–	–	–
PDZ-Eg5	–	–	–	+	–	–
PDZ-pol	–	–	–	–	+	–
PDZ-hyd	–	–	+	–	–	+++

[1]pGBK and pGAD are plasmids carrying the Gal4 binding and activation domains on their own.
[2]Negative interaction are shown as minus (**–**). Positive interactions are shown as plus (**+**). For details on quantitation see Methods.

teins Moreover, some PDZ domains, such as PTP1E/Fas[17], interact with the C-terminal hexapeptides, thus affording greater specificity. However, some limitations may apply. Because there is a certain degree of degeneracy, single PDZs may bind more than a single target sequence. In addition, all PDZ-binding peptides bear a hydrophobic amino acid at p0 (Ala, Val, Ile and Leu) in common. Recognizing charged or polar amino acids at p0 will likely be difficult because the side chain at this position is pointing towards a hydrophobic pocket. Because the percentage of human proteins carrying a hydrophobic residue at the C-terminus is no less than 25% (data not shown), the number of proteins that are, in principle, amenable to PDZ tagging is certainly high. Moreover, it could be possible to circumvent this limitation through a more extensive re-engineering of the domain that would involve some backbone rearrangements and/or placement of a polar side chain into the pocket.

Our computer design exercise shows that every position in the target peptide contributes to the specificity of binding to the PDZ domain. The specificity of the interactions with amino acids located at positions p(-1) and p(-3) of the ligand is mainly governed by residues located on the first and second β-strands of the PDZ pocket[11,18]. Polar residues at these positions require hydrogen bonding or electrostatic partners in the β-sheet, whereas apolar residues require amino acids with apolar side chains. (Even though Lys and Arg are charged residues, they have a long aliphatic side chain that can take part in stabilizing hydrophobic interactions.) Bulky amino acids such as Trp need an Ala or Gly in the nearby positions of the β-strands of the PDZ domains. Because these two positions are solvent exposed, they allow for more promiscuity than positions p0 and p(-2), as shown by the two ligands of PDZ-wt that have different residues at p(-1) and p(-3)[11].

The residue at position p(-2) of the ligand governs both the identity of the residue located at the beginning of the α-helix lining the binding pocket and the orientation of this α-helix. Ligands with Thr/Ser at p(-2) require a His residue in the helix to satisfy a hydrogen bond donor in the His side chain. In contrast, an aliphatic residue at p(-2) requires another aliphatic residue. Binding to Tyr/Phe at p(-2) requires a displacement of the α-helix to accommodate a bulky aromatic side chain[18]. Displacement of the α-helix is also required to bind Asp at p(-2) to allow the Tyr residue that will occupy the position of the canonical His to make a side chain–side chain hydrogen bond with Asp[8].

As far as position p(-4) of the ligand is concerned, the specificity requirements are not as restricted, at least in traditional PDZ domains. This side chain can make only specific interactions with the N-terminus of the α-helix and with the loop preceding the first β-strand, but it has a large conformational freedom. Therefore, as it is the case with any protein–protein interaction, increasing the affinity and specificity of a PDZ domain for a particular target sequence is possible, although there could always be some level of cross-reactivity.

In silico versus experimental screening

Experimental screens based on random mutagenesis have been applied to select modified PDZ domains with novel specificity[1], including new PDZ sequences capable of binding unique 'non-natural' ligands. Likewise, experimental screens have also been carried out to identify artificial ligands that can be recognized by a given PDZ domain[19]. Our hyd peptide corresponds to one of these: a 'non-natural' ligand that can bind to the PDZ2 of MAGI-3 (membrane-associated guanylate kinase with inverted orientation-3)[19]. To design the PDZ domain capable of binding this peptide, Perla explored 3.9×10^{22} sequences, a number of combinations that exceeds in many orders of magnitude the number that could be explored by phage display. Interestingly, a comparison between the sequences of PDZ-hyd (generated by introducing the mutations suggested by Perla in PDZ-wt) and PDZ2 of MAGI-3 reveals that the algorithm has, in fact, introduced some of the residues in the PDZ domain of MAGI-3. This is the case of the residues that interact with p(-1) and p(-3) of the ligand (Ala 26, Lys 37 and Met 38 are equivalent to Ala 326, Lys 339 and Met 340 in the PDZ-wt-hyd). At other positions, Perla has suggested keeping the PDZ-wt sequence unchanged (Ile 328 equivalent to Ala 28 in the PDZ2 of MAGI-3). Finally, at position 342, Perla made the L342A mutation, whereas in the PDZ2 of MAGI-3 it is a Leu.

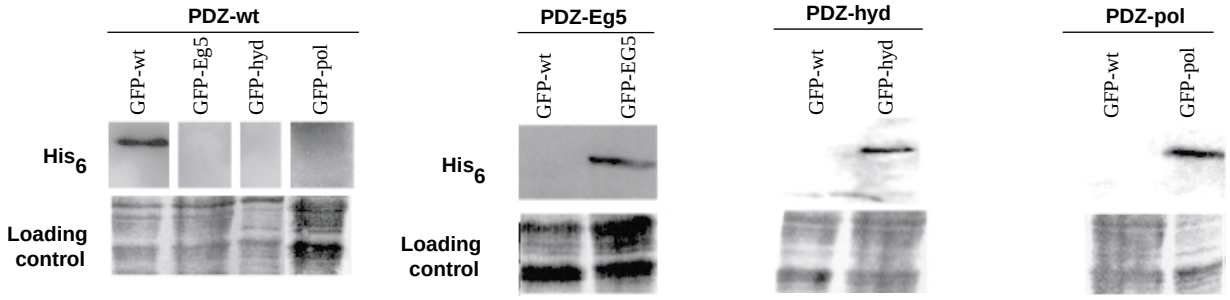


Fig. 4 Interactions between the PDZ domains and their target peptides *in vitro*. Membranes blotted with *E. coli* extracts containing GFP–wt, GFP–hyd, GFP–pol or GFP–Eg5 were incubated with the PDZ-wt, PDZ-Eg5, PDZ-hyd and PDZ-pol domains fused to His_6 as shown. Binding of the PDZs was revealed by an anti-His_6. Each PDZ domain was able to bind the blotted GFP fusion carrying the corresponding C-terminal peptide. Poinceau red-stained membranes following blotting of the input *E. coli* extracts are shown as loading controls.

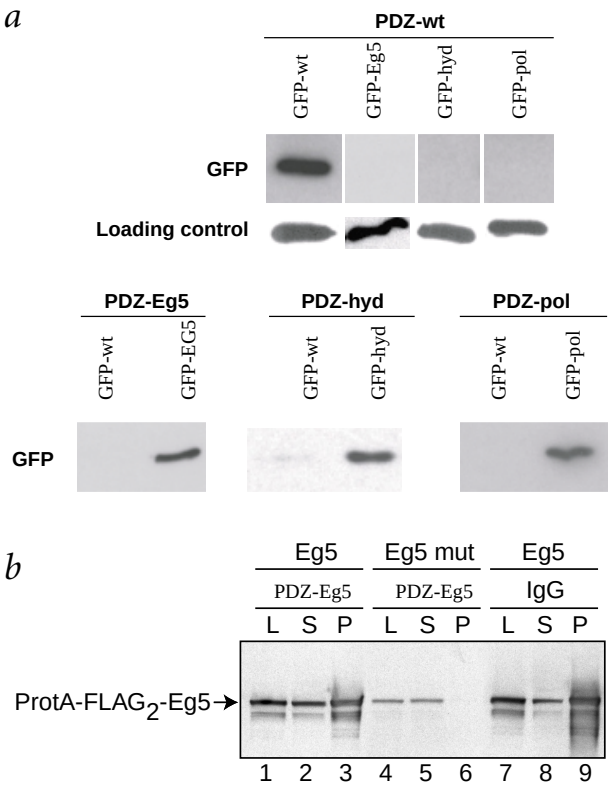


Fig. 5 Interactions between the PDZ domains and their target peptides *in vitro* and *in vivo*. ***a***, Western blot of affinity-purified protein from extracts of *E. coli* expressing GFP–wt, GFP–Eg5, GFP–hyd and GFP–pol. Binding was carried out with the PDZ-wt, PDZ-Eg5, PDZ-hyd and PDZ-pol domains that were immobilized in a solid phase. Each PDZ domain was able to retain the GFP-fusion carrying the corresponding C-terminal peptide. Input GFP is shown as a loading control. ***b***, GST pull-down experiment showing the ability of the designed PDZ-Eg5 domain to purify the native Eg5 protein from cell lysates. The PDZ-Eg5 domain was fused to GST. Input lysates (L) were loaded with homogenates from HEK 293 cells expressing ProtA-$FLAG_2$–Eg5 fusion protein (lanes 1 and 7) or ProtA-$FLAG_2$–Eg5 mutant that lacks the C-terminal five amino acids (lane 4). ProtA-$FLAG_2$–Eg5 and the ProtA-$FLAG_2$–Eg5 mutant proteins were pulled down (P) from 500 μg of lysates with glutathion (lanes 2 and 5) or IgG (lanes 8) beads. For comparison, equal amounts of protein were taken from the supernatants (S) after the pull down. The amount of Eg5 in each lane was revealed by an anti-FLAG M2 (Sigma). The PDZ-Eg5–Eg5 kinesin interaction is almost as efficient as the IgG–ProtA interaction in pulling down the native Eg5 protein.

These observations reveal that by exploring a vast number of sequences, computer-design algorithms can identify *in silico* sequences that are similar to those obtained by screening methods. Computer design can, therefore, provide an important additional step in custom designing recognition domains. Indeed, the two approaches — experimental and *in silico* screening — are complementary, and their combination is envisaged to significantly accelerate the pace of discovery of new interacting partners.

Reliability of the design algorithm

In all three design exercises we have carried out, we have found that the most favorable sequences predicted by the algorithm bind their targets with similar affinities and specificities to those described for other PDZ domains. Moreover, we constructed two additional PDZ variants for Eg5 (data not shown), one with Phe at position 325 and the other with Leu at 342, that exhibited similar affinity and specificity as PDZ-Eg5. Likewise, a variant of PDZ-hyd with Thr at 339, whose energy was predicted to be close to the PDZ-hyd that we analyzed, behaved in a similar manner in all binding assays (data not shown). However, although the rate of success using the best sequences generated by the algorithm seems to be high, predicted sequences need to be selected by the operator. One example is a third variant of PDZ-Eg5 suggested by Perla that had Val instead of Ala at position 376. Although the energy of the complex formed by this mutant was predicted to be good, it did not bind its target. The reason seems to be a small van der Waals clash that is not penalized enough by the program force field because of a softening of the van der Waals radius that is compensated by the burying of larger hydrophobic surface. Similarly, the operator must identify the most relevant positions so that the number of residues to be checked can be kept down to a minimum to optimize efficiency.

Methods

Computer program. The automated computer algorithm Perla was used as described[14,15,20] (see also *http://ProteinDesign.EMBL-Heidelberg.DE*).

Cloning of the PDZ domains and their corresponding targets. The third PDZ domain of the mouse PSD-95 protein (PDZ-wt), which spans amino acids 302–402 (ref. 10), was amplified from a mouse brain cDNA library by PCR. The fragment was then cloned into the pQE30 expression vector (Qiagen) using the *Bam*HI and *Kpn*I sites to obtain the plasmid pQEPDZp-3, which encodes a His_6–PDZ3 fusion. A pPDZ cassette was made by removing the *Sal*I site and creating an *Ngo*MIV site in pQEPDZp-3 (changing nucleotides 243 and 246 to C) without changing the ORF of the PDZ. The mutant domains were generated by PCR and cloned into the pPDZ cassette to generate PDZ-hyd, PDZ-pol and PDZ-Eg5. GFP fusions carrying the eight C-terminal amino acid residues of the target ligand peptides (GFP–wt, GFP–Eg5, GFP–pol and GFP–hyd) were made by PCR. GFP–wt contains the target of the third PDZ domain of PSD-95 (ref. 10). GFP–Eg5 contains the C-terminal octapeptide from the mouse Eg5 kinesin (PLHTSINL)[13]. GFP–pol and GFP–hyd contain residues p(-7), p(-6) and p(-5) of the third PDZ domain of PSD-95, followed by the corresponding pentapeptide target sequences.

Yeast two-hybrid analysis. GFP–wt, GFP–hyd, GFP–pol and GFP–Eg5 were cloned between the *Nco*I and *Bam*HI sites in the pGBKT7 plasmid (Clontech) to obtain a fusion with the binding domain of Gal4. PDZ-wt, PDZ-hyd PDZ-pol and PDZ-Eg5 were cloned between the *Eco*RI and *Xho*I sites of pGADT7 (Clontech) to obtain a fusion with the activation domain of Gal4. The resulting plasmids were introduced in the yeast strain J569a according to Agatep *et al.*[21] Yeast cells were plated in minimal media lacking tryptophan and leucine, and then struck in minimal media lacking tryptophan, leucine, histidine and adenine. Cells can grow only in this selective medium if there is an interaction (binding) between the expression products of the two plasmids. Interactions between the different PDZs and ligands were quantified as the rate of growth in selective media.

Affinity chromatography. PDZ-wt, PDZ-hyd, PDZ-pol and PDZ-Eg5 were expressed in *E. coli* XL1-Blue as fusions with a His_6-tag, bound to nickel resin (Ni-NTA agarose, Qiagen) and incubated with *E. coli* extracts containing GFP, GFP–wt or GFP–Eg5, GFP–hyd and GFP–pol. Following four washes with 50 mM NaH_2PO_4, 300 mM NaCl, 50 mM imidazole and 0.1% (v/v) of TritonX-100, the bound proteins were eluted using 50 mM NaH_2PO_4, 300 mM NaCl and 200 mM imidazole and analyzed by western blot using anti-GFP (Roche).

K_d measurements and competition studies. All K_d determinations and competition studies were performed using fluorescence polarization on an Analyst HT from Molecular Devices. All labeled

and unlabeled peptides (Diver Drugs S.L.) have three extra amino acids at the N-termini (Lys-Lys-Gly) added to favor solubility and to separate the fluorescent group from the binding sequence. PDZs and peptides were diluted in 50 mM Tris, pH 7.4, 100 mM NaCl and 0.01% Tween-20. K_d measurements were performed following 15 min incubations of 5 μM labeled peptide with increasing concentrations of the corresponding PDZ. For competition studies, each PDZ (at a concentration determined by their estimated K_d to have 50% binding of the corresponding 5 μM peptide) was incubated with the corresponding unlabeled peptide for 15 min, followed by an additional 20 min incubation with 5 μM labeled peptide. Table 2 shows the ratio between Ki of the wt and target peptides (K_i wt / K_i target).

Overlay assays. Extracts from *E. coli* expressing GFP–wt, GFP–hyd, GFP–pol or GFP–Eg5 were run in SDS-PAGE and blotted onto nitrocellulose membranes. These were then blocked overnight with 5% non-fatty milk in PBS, incubated for 1 h at room temperature with each PDZ and washed four times with PBST. A further incubation with a mixture of mouse anti-His_6 (1:2,000) and horseradish peroxidase-conjugated anti-mouse IgG (1:1000) was carried out for 1 h, the membranes were then washed four times with PBST and peroxidase was detected with ECL (Amersham).

GST pull-down assays. PDZ-Eg5 was cloned into pDEST15, and the full-length ORF of the mouse kinesin Eg5 and a mutant, lacking the five C-terminal amino acids, were cloned into a modified pSG5 (Stratagene) eukaryotic expression vector carrying a N-terminal ProtAp-$FLAG_2$-tag. Cloning was carried out by Gateway (Invitrogen) recombination, following the instructions from the manufacturer. GST–PDZ-Eg5 fusion protein was expressed from pDEST15-PDZ-Eg5 in *E. coli* strain BL21 and purified with Glutathione Sepharose (Amersham Pharmacia Biotech). In brief, cells were lysed by sonication in GST buffer (20 mM Tris, pH 8, 100 mM NaCl, 2 mM EDTA, 0.5% Triton X-100 and 1 mM dithiotreitol (DTT)) containing 1× Complete Protease Inhibitors (Roche) and centrifuged. The supernatant was incubated with Glutathione Sepharose for 1 h at 4 °C. Beads were washed two times with 20 bed volumes of lysis buffer and two times with 20 bed volumes lysis buffer containing 500 mM NaCl, followed by one wash with 20 mM Tris, pH 8. GST-fusion proteins were eluted with 25 mM glutathione in 50 mM Tris, pH 8. Proteins were dialyzed against 20 mM Tris, pH 7.4, 100 mM NaCl, 2 mM EDTA and 1 mM DTT overnight and stored in 50% (v/v) glycerol at –20 °C. HEK 293 cells were transiently transfected with Eg5 or a mutant form in the pSGN-ProtAp-$FLAG_2$-GW vector using a Ca^{2+}-phosphate protocol. For each 15 cm dish, 50 μg of each construct were co-transfected with 50 μg of a GFP expression vector. After 48 h the cells were harvested and lysed by shearing forces in a buffer containing 50 mM Tris, pH 7.4, 100 mM NaCl, 1.5 mM $MgCl_2$, 5% (v/v) glycerol, 0.2% NP-40, 1 mM DTT and 1× Complete Protease Inhibitors (Roche). The debris was removed by centrifugation (50,000*g* for 30 min at 4 °C) in a Beckman TL-100, and the protein concentration of the supernatant was determined by Bradford assay. For each precipitation, 500 μg of lysates from HEK 293 cells expressing the wild type or mutant fusion proteins were used. The purified GST–PDZ-Eg5 (20 μg) was added and incubated for 1 h on ice in a volume of 140 μl. Then 20 μl of Glutathione Sepharose or IgG-beads were added and incubated while rotating for 1 h at 4 °C. The beads were washed one time with 1 ml of wash buffer (50 mM Tris, pH 7.4, 100 mM NaCl, 1.5 mM $MgCl_2$, 5% (v/v) glycerol, 0.2% NP-40 and 1 mM DTT), and proteins were eluted from the beads by boiling with SDS-PAGE sample buffer. Eluted proteins were separated by SDS-PAGE and transferred onto nitrocellulose membrane. Precipitated proteins were revealed with M2 anti-FLAG (Sigma).

Acknowledgments

We thank M. Kress who kindly provided us with the plasmid containing the cDNA of the mouse Eg5 kinesin, the members of our laboratories for helpful discussions and the EMBL genomics and protein expression core facilities for their support. This work was financed by a grant from the 'Fundación Ramón Areces' and grants from the European Union.

Competing interests statement

The authors declare competing financial interests: see the Nature Structural Biology website (*http://structbio.nature.com*) for details.

Received 20 December, 2001; accepted 9 May, 2002.

1. Schneider, S. *et al.* Mutagenesis and selection of PDZ domains that bind new protein targets. *Nature Biotechnol.* **17**, 170–175 (1999).
2. Minor, D.L. Jr & Kim, P.S. Context-dependent secondary structure formation of a designed protein sequence. *Nature* **380**, 730–734 (1996).
3. Cregut, D., Civera, C., Macias, M.J., Wallon, G. & Serrano, L. A tale of two secondary structure elements: when a β-hairpin becomes an α-helix. *J. Mol. Biol.* **292**, 389–401 (1999).
4. Berezovsky, I.N., Kilosanidze, G.T., Tumanyan, V.G. & Kisselev, L. Amino acid composition of protein termini are biased in different manners. *Protein Eng.* **12**, 23–30 (1999).
5. Wilkins, M.R., Gasteiger, E., Sanchez, J.C., Appel, R.D. & Hochstrasser, D.F. Protein identification with sequence tags. *Curr. Biol.* **6**, 1543–1544 (1996).
6. Sheng, M. & Sala, C. PDZ domains and the organization of supramolecular complexes. *Annu. Rev. Neurosci.* **24**, 1–29 (2001).
7. Harris, B.Z. & Lim WA. Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.***114**, 3219–3231 (2001).
8. Tochio, H., Zhang, Q., Mandal, P., Li, M. & Zhang, M. Solution structure of the extended neuronal nitric oxide synthase PDZ domain complexed with an associated peptide. *Nature Struct. Biol.* **6**, 417–421 (1999).
9. Hata, Y., Butz, S. & Sudhof, T.C. CASK: a novel dlg/PSD95 homolog with an N-terminal calmodulin-dependent protein kinase domain identified by interaction with neurexins. *J. Neurosci.* **16**, 2488–2494 (1996).
10. Doyle, D.A. *et al.* Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* **85**,1067–1076 (1996).
11. Niethammer., M. *et al.* CRIPT, a novel postsynaptic protein that binds to the third PDZ domain of PSD-95/SAP90. *Neuron* **20**, 693–707 (1998).
12. Irie, M. *et al.* Binding of neuroligins to PSD-95. *Science* **277**, 1511–1515 (1997).
13. Ferhat, L. *et al.* Expression of the mitotic motor protein Eg5 in postmitotic neurons: implications for neuronal development. *J. Neurosci.* **18**, 7822–7835 (1998).
14. Fisinger, S., Serrano, L. & Lacroix, E. Computational estimation of specific side chain interaction energies in α-helices. *Protein Sci.* **10**, 809–818 (2001).
15. Lopez de la Paz, M., Lacroix, E., Ramirez-Alvarado, M. & Serrano, L. Computer-aided design of β-sheet peptides. *J. Mol. Biol.* **312**, 229–246 (2001).
16. Fields, S. & Sternglanz, R. The two-hybrid system: an assay for protein-protein interactions. *Trends Genet.* **10**, 286–292 (1994).
17. Kozlov, G., Gehring, K. & Ekiel, I. Solution structure of the PDZ2 domain from human phosphatase hPTP1E and its interactions with C-terminal peptides from the Fas receptor. *Biochemistry* **39**, 2572–2580 (2000).
18. Songyang, Z. *et al.* Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* **275**, 73–77 (1997).
19. Fuh, G. *et al.* Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display. *J. Biol. Chem.* **275**, 21486–21491 (2000).
20. Lacroix, E. *Protein design: a computer based approach.* PhD thesis, U. Libre de Bruxelles (1999).
21. Agatep, R., Kirkpatrick, R.D., Parchaliuk, D.L., Woods, R.A. & Gietz, R.D. Technical Tips Online (1998) *http://tto.trends.com.*