Probing the Architecture of a Multi-PDZ domain protein: Structure of PDZK1 in solution

Nelly R. Hajizadeh[1#], Joanna Pieprzyk[2#], Petr Skopintsev[2,a], Ali Flayhan[2], Dmitri I. Svergun*[1] and Christian Löw*[2,3,4]

[#]These authors contributed equally
[1] European Molecular Biology Laboratory, Hamburg Outstation, Building 25a, Notkestrasse 85, 22607 Hamburg, Germany
[2] Centre for Systems Structural Biology (CSSB), DESY and European Molecular Biology Laboratory, Notkestrasse 85, D-22607 Hamburg, Germany
[3] Department of Medical Biochemistry and Biophysics, Karolinska Insititutet, Scheels väg 2, SE-17177 Stockholm, Sweden
[4] Lead contact

[a] Present address: Laboratory of Biomolecular Research, Division of Biology and Chemistry, Paul Scherrer Institut, 5232, Villigen PSI, Switzerland

[*] Corresponding authors:
Dmitri I. Svergun
European Molecular Biology Laboratory, Hamburg Outstation, Building 25a, Notkestrasse 85, 22607 Hamburg, Germany
Tel: +49 4089902125
Email: svergun@embl-hamburg.de

Christian Löw
European Molecular Biology Laboratory, Hamburg Outstation, Building 25a, Notkestrasse 85, 22607 Hamburg, Germany
Tel: +49408988-87570
Correspondence: christian.loew@embl-hamburg.de

Summary

The scaffolding protein PDZK1 has been associated with the regulation of membrane transporters. It contains four conserved PDZ domains, which typically recognize a 3-5 residue long motif at the C-terminus of the binding partner. The atomic structures of the individual domains are available but their spatial arrangement in the full length context influencing the binding properties remained elusive. Here we report a systematic study of full-length PDZK1 and deletion constructs using small angle x-ray scattering, complemented with biochemical and functional studies on PDZK1 binding to known membrane protein partners. A hybrid modeling approach utilizing multiple scattering data sets yielded a well-defined, extended, asymmetric L-shaped domain organization of PDZK1 in contrast to a flexible ‘beads-on-string’ model predicted by bioinformatic analysis. The linker regions of PDZK1 appear to play a central role in the arrangement of the four domains underlying the importance of studying scaffolding proteins in their full-length context.

Keywords

scaffolding protein, SAXS, X-ray Crystallography, PDZ-domains, peptide transporter, sodium-proton exchanger

Abbreviations

DLS - dynamic light scattering

I(0) - Intensity at zero angle

R_g - radius of gyration

R_H - Hydrodynamic radius

K_D - Dissociation constant

d_{max} - maximum dimension

NSD - normalized spatial discrepancy

NHERF - Na^+/H^+ exchanger regulatory factor

NHE3 - sodium-hydrogen exchanger 3

PDZK1 - Na(+)/H(+) exchange regulatory cofactor NHE-RF3

PEPT2 - human peptide transporter 2

PEPT1 - human peptide transporter 1

SAXS - small angle X-ray scattering

SEC-SAXS - size-exclusion chromatography SAXS

Introduction

Despite being bound to membranes, membrane proteins often form multiple interactions with other proteins or protein complexes on the cytoplasmic and extracellular side, which are necessary for cellular processes such as biosynthesis, signal transduction and cell polarity (Lee and Zheng, 2010; Li et al., 2007). Interactions between membrane proteins at the cytoplasmic side are often mediated through molecular scaffolding proteins that anchor e.g. the membrane proteins to the cytoskeleton, thereby localizing the participants in sub-cellular space (Birrane et al., 2013). However, these scaffold proteins are not simple 'molecular glues' but contribute actively to the formation and stability of membrane protein complexes caused by different phosphorylation patterns and intramolecular interactions, thereby affecting the dynamics of membrane proteins and diversify their signaling (Bezprozvanny and Maximov, 2001; Pawson, 2007). The interest in the molecular scaffolding proteins has therefore increased significantly in recent years with the recognition of their pivotal and multivariate role for intracellular networks (Good et al., 2011; Manjunath et al., 2017).

Scaffolding proteins are uniquely suited for their role through a modular structure, consisting of multiple protein-protein interaction domains connected by linkers that confer their ability to associate with multiple partners. There is an incredible diversity in scaffolding proteins and their domain composition, but a common and highly evolutionary conserved domain is the PDZ domain (Chi et al., 2012; Dunn and Ferguson, 2015; Feng and Zhang, 2009; Gallardo et al., 2010; Ivarsson, 2012; Ye and Zhang, 2013). The high abundance of PDZ domains implicate their crucial role, and they have been associated with cellular

functions such as intracellular trafficking, ion channel signaling and cell-cell junctions (James and Roberts, 2016; Sobhy, 2016). The PDZ domains execute their scaffolding task by recognizing a 3-5 long residue motif, which can appear both on the cytoplasmic C-terminus of their protein partners (Cheng et al., 2009; Fanning and Anderson, 1996), or as a finger-like projection from an internal sequence (Dunn and Ferguson, 2015). The PDZ domains themselves are relatively small domains of 80-90 residues in length, with a highly conserved fold, which are well suited for high-resolution studies by X-ray crystallography and NMR. Much of the effort has been focused both on understanding the structural characteristics of the individual domains and on their interactions with the target proteins (Harris and Lim, 2001; Ivarsson, 2012; Teyra et al., 2012; Trave, 2011; Ye and Zhang, 2013).

The members of the sub-family Na^+/H^+ exchanger regulatory factors (NHERFs) are model PDZ proteins (Broadbent et al., 2017) and have been extensively studied. The NHERF family consists of four evolutionary related members, with the common functionality to regulate the sodium-hydrogen exchanger 3 (NHE3) (Zachos et al., 2009). One member, NHERF3 also known as PDZK1, has been associated with the regulation of membrane ion transporters in gastrointestinal tissue and kidney epithelia such as the NHE3 ion exchanger, the peptide transporter PEPT2 and the cystic fibrosis trans-membrane regulator CFTR (Kato et al., 2006; Noshiro et al., 2006; Park et al., 2014; Sugiura et al., 2006, 2008; Zachos et al., 2009).

PDZK1 contains four PDZ domains connected by three linkers and an extended C-terminal tail. It has been suggested that each PDZ domain of PDZK1 binds independently to the PDZ binding motif of various membrane

proteins (Park et al., 2014) to organize a macromolecular complex (Wang et al., 2000). Structurally all four PDZ domains of PDZK1 have been solved by high-resolution techniques, however most of their interactions and their role in cell physiology remains relatively unexplored (Park et al., 2014; Walther et al., 2015). Furthermore, the overall arrangement of the domains in PDZK1 in the full-length context is unknown, most likely due to the expected flexibility caused by the three internal loops and the C-terminal tail. Indeed, from a sequence perspective, the arrangement of the PDZ domains in PDZK1 appears to follow a simple ‘beads-on-a-string’ model as suggested by bioinformatics tools, an association which has most likely been further reinforced by the hitherto focus on the well-defined PDZ domains.

Here, we complement the picture of full-length PDZK1 through a systematic analysis of small angle x-ray scattering (SAXS) experiments on full-length and deletion constructs, thereby arriving at a representative model of the domain organization of full-length PDZK1 in solution. Our data indicate that PDZK1 has a relatively defined conformation, as opposed to a freely flexible ‘beads-on-string’ model and further propose that the linkers play an active structural role in supporting its conformation. The SAXS analyses are complemented with biochemical and functional studies of the binding of PDZK1 to known membrane protein partners.

Results & Discussion

Overall structural parameters for PDZK1 from SAXS

To understand the arrangement of the PDZ domains in the full length protein, single domain (D1, D2, D3, D4) and multiple domain constructs (D12, D23, D34, D13, D24, D14, and FL) were expressed and purified based on known crystal structures and bioinformatics analysis (see Methods). The naming convention for the constructs designates the first and the last domain in the linear sequence, where FL refers to the full-length PDZK1 construct containing also the C-terminal tail. In addition, a mutant of D14 was also studied (Figure 1), from here on referred to as 'MUT'. SAXS data were collected for these 12 PDZK1 constructs (Figure 1) with the aim to determine the overall domain organization of PDZK1.

The quality of all purified constructs was analyzed by SDS-PAGE and analytical gel filtration prior to SAXS data collection (Figure S1). For each sample, concentration series were collected and the data appropriately processed to remove aggregation or inter-particle effects, by extrapolation to infinite dilution or by merging the data at different concentrations. All constructs revealed linear Guinier regions supporting a homogenous sample preparation (Figure S2, Figure S3). The determined molecular weights (MW) after *I*(0) (forward scattering) analysis were within ~10% of the MWs expected for the monomeric form in solution (Table 1, Table S2).

The SAXS profiles (scattering intensities as a function of momentum transfer $s = 4\pi\sin(\theta)/\lambda$ where 2θ is the scattering angle and λ the X-ray wavelength) and atom-pair distance distributions (*P(r) versus r*) of the single domains are consistent with the expected globular shapes of these domains

(Figure S3). The two-domain fragments associate to form compact but elongated structures, as already reported in earlier studies for two-domain PDZ constructs (Delhommel et al., 2017; Goult et al., 2007) (Figure S3). Expectedly, the construct D34 appears larger than D12 and D23, due to the long 49-residues linker between D3 and D4 (Table 1). Interestingly, D34 was the only fragment with a clear tendency to dimerize in solution, and it was therefore the only sample where inline size-exclusion chromatography SAXS (SEC-SAXS) was needed to measure the scattering for its monomeric form.

The three-domain constructs, D13 and D24, while similar in the maximum size d_{max} (about 125 Å) differ in their radius of gyration R_g, with the former being slightly larger (37.2 ± 0.1 vs 35.2 ± 0.1 Å, respectively). From the *P(r)* profile these differences can be attributed to the fact that D24 is more extended and rod-like while D13 has three inter-domain distances (two shoulders at larger *r*, Figure S4) consistent with a slightly curved shape.

Comparing the overall parameters of the two- and three-domain constructs in Table 1 allows two conclusions to be made. First, both the R_g and d_{max} increase only marginally when comparing D34 and D24. Assuming that the conformation of D34 alone is consistent with the one present in D24 it would suggest that D2 or its neighboring linker associates closely with the rest of the D24 fragment. Alternatively, it is possible that D34 alone is more extended than in D24, potentially due to missing interactions from D2. Second, the structural parameters of D12 and D23 are remarkably similar. However, when all three domains are present in D13 the R_g increases markedly and exceeds that of D24. As D1 is the only domain not present in D24, this suggests that D1 contributes to the larger R_g of D13 by being positioned at an open angle to D23.

The SAXS profile of the four domain construct D14 (which lacks the C-terminal tail region) (Figure 2A) indicates that the construct adopts a rather rigid conformation in solution. The *P(r)* displays two peaks revealing that D14 is an extended particle with two major inter-domain distances, characteristic of dumb-bell like proteins (Figure 2C). The normalized Kratky plot (Durand et al., 2010) has a defined shape with a maximum at sR_g ~4.5 further suggesting that the flexibility is limited (Figure 2E). The structural parameters of D14 are larger than those of the smaller constructs (Table 1). Together with the observations made on the truncation constructs, the data suggest that D14 has an extended curved shape with limited flexibility.

In conclusion, while the linker regions may in principle allow for considerable movements of the individual domains, the data from D14 and its three-domain constructs are indicative of a preferred arrangement of the PDZ domains in solution. It is furthermore clear that these preferred conformations differ between the constructs. These results are in solid discord with the prevalent ‘beads-on-a-string’ model and to further substantiate these findings, extended ab-initio and hybrid modeling of the SAXS data was conducted.

Low resolution SAXS models of D14 in solution

Ab-initio modeling was performed with GASBOR (Svergun et al., 2001) and MONSA (Svergun, 1999) as described in the methods section. GASBOR represents each amino acid as a bead and refines the spatial positioning of these using only the scattering profile of D14 and the number of residues. MONSA creates a bead model by simultaneously fitting scattering curves of multiple constructs, and can therefore extract additional information about the

relative positioning of the domains. The MONSA modeling was simultaneously fitting the data from D12, D13, D24 and D14, and also using the R_g values and volumes of the individual domains. When the constructs D23 and D34 were included the fits exhibited systematic deviations, although the overall model was consistent with the one generated from the abovementioned four constructs. This observation indicates that the structures of the individual tandems D23 and D34 are somewhat modified in the context of the larger constructs (data not shown) further supporting the aforementioned hypothesis that D34 is more extended alone due to missing stabilizing interactions from D2.

The GASBOR model (Figure 3A) reveals PDZK1 as an extended, asymmetric, modular L-shaped protein. The MONSA model (Figure 3C) has an encouragingly similar shape to that of the GASBOR model, as indicated by the normalized spatial discrepancy (NSD) value of 1.4, which is a measure of quantitative similarity between sets of three-dimensional points (objects with similar shape would have NSD of approximately 1). The MONSA model furthermore reveals the relative positions of the domains within the D14 construct. The domain organization of the MONSA model is fully consistent with the qualitative predictions of the previous section; in particular, the D1 domain contributes to the bend of the conformation (Figure 3C). The fits from MONSA (χ^2 =1.52, 1.70, 1.20, 0.75) and GASBOR (χ^2 = 0.9) are acceptable to good as expected for systems with some limited degree of flexibility (Figure 3B, D).

Using the available crystal structures of the four individual PDZ domains a rigid body model was generated using the program BUNCH (Petoukhov and Svergun, 2005) as described in the methods section (Figure 3E). BUNCH fits

multiple SAXS datasets from multidomain proteins representing the domain with their atomic models and the linkers as chains of dummy residues. The BUNCH model was generated using the same set of constructs as the MONSA model, and, similarly, the inclusion of D23 and D34 worsened the fits while yielding an overall similar domain arrangement (data not shown). The BUNCH model corresponds well to the GASBOR and MONSA models (NSD of 3.4 and 2.4, respectively) further supporting the positioning of the domains as suggested by MONSA. The discrepancies of the fits from BUNCH are somewhat larger than those for MONSA (Figure 3F, χ^2 = 1.2, 1.12, 1.37, 1.15); again as to be expected from systems containing a certain degree of flexibility. Furthermore, the exact domain orientation of the PDZ domains in the BUNCH model should not be interpreted too literally given that the protein is moderately flexible. However, the observed agreement between the three different models supports the idea that PDZK1 is a relatively defined extended, asymmetric L-shape protein in solution with only moderate flexibility.

PDZK1 linker regions are surprisingly compact

The SAXS data indicate that the three-domain constructs along with D14 have a defined conformation in solution, despite the presence of long linker regions. This is further supported by considering the R_g and d_{max} of the different constructs versus the number of domains. Figure 4A, B illustrate that adding PDZ domains and their linkers non-linearly increase the R_g and d_{max}, most likely due to the presence of intra-particle interactions, which become apparent already in the three-domain constructs. The addition of D4 and a 49-residue linker results only in a modest increase in the R_g as seen by the difference

between D13 and D14 (Figure 4A), suggesting that PDZK1 does not fully sample the available conformational space. Similar surprising compactness has also been seen in another PDZ containing protein with a long linker (Goult et al., 2007). It is therefore conceivable that the linker regions of PDZK1 play an active and even central role in the arrangement of the four domains within the protein.

Sequence conservation of non-PDZ regions in PDZK1

Bioinformatics analysis in the form of sequence conservation, secondary structure and disorder prediction was performed on the full length sequence of PDZK1. The sequence conservation is sparse in the linker regions of PDZK1, but two stretches of residues in the C-terminal tail stand out with ~11 and ~14 consecutive and highly conserved residues (Figure S5). From secondary structure prediction most of the non-PDZ regions consist of disordered regions without any significant secondary structure (Figure S6), with the exception of the beginning of the linker between D3 and D4 (residues 348-399). This linker is predicted to be most likely buried (Figure S5) and potentially folds into an alpha helix (Figure S6). Some of the regions with predicted disorder, such as the first linker, part of the third linker as well as the C-terminal tail score high for being a protein binding site (Figure S5 and S6), suggesting that they may act as intra-molecular binding sites. Indeed, the C-terminal four residues presents a PDZ-binding motif and have been proposed to be involved in intra- as well as inter-molecular interactions (Lalonde and Bretscher, 2009; Yang et al., 2014).

There are therefore two non-PDZ regions which may play a prominent role in the structure and function of PDZK1, namely the linker between D3 and

D4 as well as the C-terminal tail. For the remaining two non-PDZ regions the bioinformatics results were not as conclusive. However, these regions are considerably shorter than the tail and the last linker, and it is possible that they help to define the intramolecular characteristics of PDZK1.

Structural importance of the linker between D3 and D4

To elucidate the role of the linker between D3 and D4 in the overall conformation of PDZK1, we replaced the predicted alpha-helical residues 351-359 with a flexible GS-linker (Figure 1) in D14, resulting in the construct MUT. The SAXS data collected on MUT showed that this protein is significantly more extended than the wild type (with a R_g 46.6 ± 0.6 vs 40.6 ± 0.2 Å, respectively), with a d_{max} increase of ~20 Å (Table 1, Figure 4). The R_H from DLS measurements confirms the difference seen in the SAXS R_g between MUT and D14 (47 ± 2 versus 39 ± 2 Å, respectively). In addition, the slower decay of the *P(r)* profile at higher *r* as compared to the D14 construct further supports the view that the mutant is more extended (Figure 2C) and possibly more flexible as indicated by the lack of features in the normalized Kratky plot (Figure 2E).

To further explore this increased flexibility caused by the unfolded linker in MUT, various PDZK1 constructs (FL, D14, MUT, D34) were exposed to chymotrypsin at high concentrations (molar ratio of 1:100) and the proteolysis was followed over time (Figure 5). FL and D14 were quite resistant to proteolytic degradation indicating that they adopt a global, well-defined tertiary structure (Figure 5A, B). This stands in contrast to the fast cleavage of MUT, which contains the mutated linker stretch supporting the structural role of the linker regions in PDZK1 folding (Figure 5C). The importance of the linker is further

seen in the rapid degradation of D34, which may be due to the missing stabilizing interactions from D1 and D2 (Figure 5D). This picture is further substantiated by the thermostability measurements of various PDZK1 constructs (Figure 6) using intrinsic tyrosine and tryptophan fluorescence. Individual PDZK1 domains (D1, D2, D3, and D4) display a rather broad unfolding transition in agreement with the small surface they expose after denaturation. The stability amongst the individual PDZ domains differ largely, with D4 being the most stable one (Figure 6A). In contrast to the individual domains, D14 and FL unfold in a highly cooperative manner with FL being slightly more stable than D14. This is further confirmed by circular dichroism unfolding transitions (data not shown). MUT, however is strongly destabilized compared to D14 and exhibits at least two unfolding transitions indicating that not all individual domains are coupled (Figure 6B).

The C-terminal tail of FL does not induce a more compact conformation compared to D14

So far, our study has focused on the structural and functional importance of the longest PDZK1 linker, between D3 and D4. The longest non-PDZ region, however, is that of the 59 residues long C-terminal tail which is present in the full-length construct FL (Figure 1). It has previously been suggested (Lalonde and Bretscher, 2009) that the C-terminal tail of PDZK1 promotes an overall ring-shaped conformation of the full-length protein through the binding of the C-terminal four residues (DTEM) of the tail to the first PDZ domain. To determine whether we observe this in solution, SAXS data were collected on FL. The SAXS profile of FL does not reveal a ring-shaped protein conformation (Figure

2B, Table 1), which is further confirmed by a classification algorithm (Franke et al., 2018) which classifies FL as an extended protein. The addition of the C-terminal tail results in a small increase of R_g and d_{max} (by 2.3 Å and 5 Å, respectively compared to D14) suggesting that the tail does not induce a more compact conformation under the studied conditions (Table 1, Figure 2B, C, Figure 4). The R_H of FL (45 ± 1 Å) is 6 Å larger than that of D14 (39 ± 2 Å), further corroborating the SAXS results. Furthermore, the C-terminal tail region adds to the intra-domain distance as can be seen in the larger second peak in the *P(r)* of FL as compared to D14. Overall, the modest increase in the structural parameters and the difference in the shape of the *P(r)* may suggest, but alone cannot prove, that the tail folds back on the PDZ domains. However, it is clear that the tail neither promotes a ring-like structure of FL nor leads to a more compact shape compared to the tailless construct.

GASBOR models of FL are overall similar to those of D14, with a somewhat bulkier appearance, possibly reflecting an increased flexibility (Figure S7). A total of ten rigid body models generated with BUNCH, using only the full-length construct dataset, produce a range of possible conformations of the tail in the space around D3-D4 (Figure S7). Based on these SAXS models, it could be concluded that the C-terminus does not extend beyond D3, but instead is proximal to D3-D4, possibly mediated by intra-molecular interactions.

Peptide binding to PDZK1

To this point the SAXS and biochemical data support the view of a defined arrangement of the PDZ domains in PDZK1. To further our understanding of PDZK1 as scaffolding protein, we next characterized its response to ligand

binding. Previous studies using *in vitro* binding assays and pull down assays showed the association of PDZK1 with the C-terminal region of the human peptide transporter PEPT2 (Noshiro et al., 2006) and sodium proton exchanger NHE3 (Zachos et al., 2009). We therefore investigated the binding of various PDZK1 constructs to fluorescently labeled C-terminal peptides of PEPT1 (PEPT1-CT-FITC), PEPT2 (PEPT2-CT-FITC) and NHE3 (NHE3-CT-FITC) using fluorescence anisotropy measurements. A summary of the obtained dissociation constants (K_D) is shown in Table S1. PEPT2-CT-FITC binds with significantly higher affinity to PDZK1 (FL) compared to the C-terminal peptide derived from PEPT1 ($K_{D,\ PEPT1}$ = 286 ± 36 µM, vs. $K_{D,\ PEPT2}$ =4.9 ± 0.2 µM). Among the individual domain constructs, domain four of PDZK1 has the highest affinity to the C-terminus of PEPT2 (Figure 7A), albeit the other PDZ domains also bind PEPT2-CT-FITC in the high micromolar affinity range. PEPT2-CT binds to D4 via its canonical PDZ binding pocket, as illustrated by a high-resolution crystal structure of the D4-PEPT2-CT complex (Figure S7, Table S3). The presence of additional PDZ domains in the PDZK1 constructs lead to an apparent increase in the affinity to PEPT2-CT-FITC (Figure 7A-C). Since the PDZK1 constructs are the titrants in this experimental setup while PEPT2-CT-FITC remains constant, this observed increase in apparent affinity is likely caused by the presence of multiple binding sites and affinity changes of one or more PDZ domains in the full length context. The latter is further supported by the fact that MUT shows a reduced binding affinity for the C-terminal peptides of PEPT2 and PEPT1, potentially caused by the distorted linker between D3 and D4 which influence the stability and arrangement of the domains towards each other.

The binding affinities of the PEPT1 and PEPT2 peptides can in turn be contrasted with those of NHE3-CT-FITC. In the latter case, it is clear that the main binding partner of the NHE3-CT-FITC is D1 and the presence of the other domains does not improve the affinity (Table S1, Figure S8), but might rather obstruct binding (more than fivefold lower binding affinity in FL). As such, the binding affinities of PDZK1 differ whether they are measured in the frame of the full-length protein or of individual domains, thereby highlighting the importance of approaching studies of PDZ domains in their full-length context.

To understand whether peptide binding induces an overall structural change we performed SAXS ligand titrations of the PDZK1 D14 construct with C-terminal peptides of PEPT2 (PEPT2-CT) and NH3 (NHE3-CT). The SAXS data consistently demonstrate that PEPT2-CT and NHE3-CT associate with D14 without inducing any overall structural change on the protein (Figure S2G, H). Furthermore, even when the peptides are present in excess (by 5-fold), the contribution to the scattering from the peptides alone is very small (Figure S2G, H). Hence, both peptides bind D14 with no detectable structural or volumetric change, with the only discernable difference being an increase in overall electron density (Figure S2G, H). It is possible that the moderate flexibility of PDZK1 may be masking the structural change that could result from ligand binding.

Following up the notion gained from both the bioinformatics and the SAXS study for a structural and/or functional role of the C-terminal tail, binding affinities were also determined for a peptide consisting of the last 8 residues from PDZK1 (PDZK1-CT-FITC, Table S1, Figure S8). Only low-affinity binding in the high μM range was detectable for the individual domains, and although

the affinity increased for the four domain construct D14 it still remained more than 35-fold lower than that of PEPT2-CT-FITC. Fluorescence anisotropy measurements of FL resulted in a twofold lower affinity compared to D14 indicating that the C-terminus of PDZK1 may fold back on D14 as suggested by the SAXS studies.

Conclusion

The aim of this study was to understand how the different PDZ domains of PDZK1 are arranged in the full-length protein. Despite extensive efforts to determine the three-dimensional structure of this multi-domain protein by X-ray crystallography, PDZK1 has resisted crystallization. Instead, a SAXS study on deletion and full-length constructs, featuring global ab-initio and rigid body modeling, yielded a consistent model of the domain organization. Our data show that D14 has a preferred extended, L-shaped conformation with a small degree of flexibility. This picture challenges the view of PDZK1 being a flexible disordered beads-on-a-string protein as suggested by bioinformatics analysis (Supplementary material, Figures S3 and S4). Instead our results are in accordance with the expected role of PDZK1 as a scaffolding protein confining its binding partners in space. Furthermore, the defined conformation of the three and four-domain constructs as shown by SAXS, together with biochemical stability studies, confirm that there are pertinent intra-molecular interactions, which stabilize the molecule. We propose that the linker regions are not simple disordered spacers, but contribute actively to the defined conformation of PDZK1 through the formation of structural elements and

binding sites, enabling the aforementioned stabilizing interactions within the protein.

So far, the most extensively studied non-PDZ region of PDZK1 has been the C-terminal tail; also called the C-terminal regulatory region. Indeed, multiple studies have identified this region, and the phosphorylation of specific residues of this region, as of crucial functional importance for the regulation of receptors (Nakamura et al., 2005; Turner et al., 2011). Our bioinformatics analysis clearly supports this fact and has further highlighted that the C-terminal part contains two potentially quite interesting regions, one of which corresponds to the last eight C-terminal residues and has been partly characterized here. We show that the C-terminal tail could associate intra-molecularly, but it is not clear whether it has a preferred binding site within PDZK1. In addition, the C-terminal tail does not induce PDZK1 to adopt a ring shape structure as previously suggested (Lalonde and Bretscher, 2009), but this does not exclude the possibility that it may adopt a sub-structure itself. It is clear that the C-terminal tail is potentially a very interesting avenue for further exploration, where earlier investigations have pinpointed its crucial role for heterodimerization with another scaffolding protein such as NHERF2 (Yang et al., 2014).

Through characterizing the different affinities of PDZK1 and its constructs to the C-terminal peptides of NHE3 and PEPT2 we further reason that there are different ways how PDZK1 selects and interacts with its binding partners. While in the case of NHE3, the first domain of PDZK1 is specific and sufficient for binding, it is different for PEPT2. Here all four individual domains bound the C-terminal peptide of PEPT2 albeit with varying affinities. In the full length context, this leads to an apparent increased binding affinity, but it is not

clear whether only one or multiple binding sites can be occupied at the same time or the affinity for individual PDZ domains changes in the full length context. Considering that PDZK1 has been suggested to couple PEPT2 and NHE3 producing a concerted function equivalent of a Na^+ dependent oligopeptide transporter, these different modes may help to sequester these two proteins without making them compete (Noshiro et al., 2006). Indeed, Na^+/H^+ transporters such as NHE3 are frequently found in macromolecular complexes, as the generated gradient may be needed to power the partner transporter.

The four PDZ domains of PDZK1 are undoubtedly elemental for understanding the function of PDZK1. However, as outlined here, the complexity of PDZK1 function may also be sourced from its non-PDZ regions, which provides protein binding, phosphorylation and post-translational modification sites (Supplementary material S5). These linker regions also contribute to the remarkable structural integrity of PDZK1 as shown here and potentially modulate the affinities to its partner proteins. We may therefore begin to fully understand the function and dynamics of PDZK1 and those of similar scaffolding proteins by also considering the role of these, often overlooked, intermediate regions.

Additional information

Acknowledgements

We thank the Sample Preparation and Characterization facility of EMBL Hamburg for support with nanoDSF measurement and crystallization. Melissa Gräwert and Deimante Rasenyte for help with the DLS measurements; Vivian Pogenberg for help in setting up the fluorescence anisotropy measurements and discussions on data analysis; Samuel Pazicky and Maria Garcia Alai for help with CD measurements; Cy Jeffries for helpful suggestions for the SAXS data analysis; Haydyn Mertens for the help with collecting the SEC-SAXS data at the P12 BioSAXS beamline, PETRAIII Storage Ring, DESY, Hamburg. We thank Ylva Ivarsson (Uppsala University) and Toby Gibson (EMBL Hamburg) for helpful discussions on PDZ domains. This work was funded by the Horizon2020 program of the European Union (iNEXT grant, project No. 653706) and the Swedish Research Council (grant number 621-2013-5905).

Author contributions

Conceptualization, C.L; Methodology, N.R.H, D.I.S and C.L; Investigation, N.R.H, J.P, P.S and A.F; Writing – Original Draft, N.R.H, D.I.S. and C.L.; Writing – Review & Editing, All authors; Funding Acquisition, C.L and D.I.S; Resources, D.I.S and C.L; Supervision, D.I.S and C.L.

Declaration of Interests

The authors declare that they have no conflict of interest with the contents of this article.

References

Adams, P.D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 213–221.

Aslanidis, C., and de Jong, P.J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. *18*, 6069–6074.

Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., Casadio, R., and Ben-Tal, N. (2004). ConSeq: The identification of functionally and structurally important residues in protein sequences. Bioinformatics *20*, 1322–1324.

Bezprozvanny, I., and Maximov, a (2001). PDZ domains: More than just a glue. Proc. Natl. Acad. Sci. U. S. A. *98*, 787–789.

Birrane, G., Mulvaney, E.P., Pal, R., Kinsella, B.T., and Kocher, O. (2013). Molecular Analysis of the Prostacyclin Receptor's Interaction with the PDZ1 Domain of Its Adaptor Protein PDZK1. PLoS One *8*.

Blanchet, C.E., Spilotros, A., Schwemmer, F., Graewert, M.A., Kikhney, A., Jeffries, C.M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., et al. (2015). Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). J. Appl. Crystallogr. *48*, 431–443.

Broadbent, D., Ahmadzai, M.M., Kammala, A.K., Yang, C., Occhiuto, C., Das, R., and Subramanian, H. (2017). Roles of NHERF Family of PDZ-Binding Proteins in Regulating GPCR Functions (Elsevier Inc.).

Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. *41*.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 12–21.

Cheng, H., Li, J., Fazlieva, R., Dai, Z., Bu, Z., and Roder, H. (2009). Autoinhibitory Interactions between the PDZ2 and C-terminal Domains in the

Scaffolding Protein NHERF1. Structure *17*, 660–669.
Chi, C.N., Bach, A., Strømgaard, K., Gianni, S., and Jemth, P. (2012). Ligand binding by PDZ domains. BioFactors *38*, 338–348.
Delhommel, F., Cordier, F., Bardiaux, B., Bouvier, G., Colcombet-Cazenave, B., Brier, S., Raynal, B., Nouaille, S., Bahloul, A., Chamot-Rooke, J., et al. (2017). Structural Characterization of Whirlin Reveals an Unexpected and Dynamic Supramodule Conformation of Its PDZ Tandem. Structure *25*, 1645–1656.e5.
Dosztányi, Z., Mészáros, B., and Simon, I. (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics *25*, 2745–2746.
Dunn, H.A., and Ferguson, S.S.G. (2015). PDZ Protein Regulation of G Protein-Coupled Receptor Trafficking and Signaling Pathways. Mol. Pharmacol. *88*, 624–639.
Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010). NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. J. Struct. Biol. *169*, 45–53.
Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 486–501.
Fanning, A.S., and Anderson, J.M. (1996). Protein-protein interactions: PDZ domain networks. Curr. Biol. *6*, 1385–1388.
Feng, W., and Zhang, M. (2009). Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. Nat. Rev. Neurosci. *10*, 87–99.
Franke, D., Jeffries, C.M., and Svergun, D.I. (2015). Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. Nat. Methods *12*, 419–422.
Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., et al. (2017). ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J. Appl. Crystallogr. *50*, 1212–1225.
Franke, D., Jeffries, C.M., and Svergun, D.I. (2018). Machine Learning

Methods for X-Ray Scattering Data Analysis from Biomacromolecular Solutions. Biophys. J. *114*, 2485–2492.
Gallardo, R., Ivarsson, Y., Schymkowitz, J., Rousseau, F., and Zimmermann, P. (2010). Structural diversity of PDZ-lipid interactions. ChemBioChem *11*, 456–467.
Good, M.C., Zalatan, J.G., and Lim, W.A. (2011). Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. Science (80-.). *332*, 680–686.
Goult, B.T., Rapley, J.D., Dart, C., Kitmitto, A., Grossmann, J.G., Leyland, M.L., and Lian, L.Y. (2007). Small-angle X-ray scattering and NMR studies of the conformation of the PDZ region of SAP97 and its interactions with Kir2.1. Biochemistry *46*, 14117–14128.
Harris, B.Z., and Lim, W. a (2001). Mechanism and role of PDZ domains in signaling complex assembly. J. Cell Sci. *114*, 3219–3231.
Ivarsson, Y. (2012). Plasticity of PDZ domains in ligand recognition and signaling. FEBS Lett. *586*, 2638–2647.
James, C., and Roberts, S. (2016). Viral Interactions with PDZ Domain-Containing Proteins—An Oncogenic Trait? Pathogens *5*, 8.
Jones, D.T., and Cozzetto, D. (2015). DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. Bioinformatics *31*, 857–863.
Kabsch, W. (2010). XDS. Acta Crystallogr D Biol Crystallogr *66*, 125–132.
Karplus, P.A., and Diederichs, K. (2012). Linking crystallographic model and data quality. Science (80-.). *336*, 1030–1033.
Karthikeyan, S., Leung, T., and Ladias, J.A.A. (2001). Structural Basis of the Na+/H+ Exchanger Regulatory Factor PDZ1 Interaction with the Carboxyl-terminal Region of the Cystic Fibrosis Transmembrane Conductance Regulator. J. Biol. Chem. *276*, 19683–19686.
Kato, Y., Watanabe, C., and Tsuji, A. (2006). Regulation of drug transporters by PDZ adaptor proteins and nuclear receptors. Eur. J. Pharm. Sci. *27*, 487–500.
Kocher, O., Birrane, G., Tsukamoto, K., Fenske, S., Yesilaltay, A., Pal, R., Daniels, K., Ladias, J.A.A., and Krieger, M. (2010). In vitro and in vivo analysis of the binding of the C terminus of the HDL receptor scavenger receptor class B, type I (SR-BI), to the PDZ1 domain of its adaptor protein

PDZK1. J. Biol. Chem. *285*, 34999–35010.
Kocher, O., Birrane, G., Yesilaltay, A., Shechter, S., Pal, R., Daniels, K., and Krieger, M. (2011). Identification of the PDZ3 domain of the adaptor protein PDZK1 as a second, physiologically functional, binding site for the C-terminus of the HDL receptor SR-BI. J. Biol. Chem.
Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: A Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. *36*, 1277–1282.
Lalonde, D.P., and Bretscher, A. (2009). The Scaffold Protein PDZK1 Undergoes a Head-to-Tail Intramolecular Association That Negatively Regulates Its Interaction with EBP50 †. 2261–2271.
Lee, H.-J., and Zheng, J.J. (2010). PDZ domains and their binding partners: structure, specificity, and modification. Cell Commun. Signal. *8*, 8.
Li, J., Poulikakos, P.I., Dai, Z., Testa, J.R., Callaway, D.J.E., and Bu, Z. (2007). Protein kinase C phosphorylation disrupts Na+/H+ exchanger regulatory factor 1 autoinhibition and promotes cystic fibrosis transmembrane conductance regulator macromolecular assembly. J. Biol. Chem. *282*, 27086–27099.
Manjunath, G.P., Ramanujam, P.L., and Galande, S. (2017). Structure function relations in PDZ-domain-containing proteins: Implications for protein networks in cellular signalling. J. Biosci.
McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.
Mueller-Dieckmann, C., Bowler, M.W., Carpentier, P., Flot, D., McCarthy, A.A., Nanao, M.H., Nurizzo, D., Pernot, P., Popov, A., Round, A., et al. (2015). The status of the macromolecular crystallography beamlines at the European Synchrotron Radiation Facility. Eur. Phys. J. Plus *130*.
Nakamura, T., Shibata, N., Nishimoto-Shibata, T., Feng, D., Ikemoto, M., Motojima, K., Iso-O, N., Tsukamoto, K., Tsujimoto, M., and Arai, H. (2005). Regulation of SR-BI protein levels by phosphorylation of its associated protein, PDZK1. Proc. Natl. Acad. Sci. U. S. A. *102*, 13404–13409.
Noshiro, R., Anzai, N., Sakata, T., Miyazaki, H., Terada, T., Shin, H.J., He, X., Miura, D., Inui, K., Kanai, Y., et al. (2006). The PDZ domain protein PDZK1

interacts with human peptide transporter PEPT2 and enhances its transport activity. Kidney Int. *70*, 275–282.
Orthaber, D., Bergmann, A., and Glatter, O. (2000). SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. J. Appl. Crystallogr. *33*, 218–225.
Von Ossowski, I., Oksanen, E., Von Ossowski, L., Cai, C., Sundberg, M., Goldman, A., and Keinänen, K. (2006). Crystal structure of the second PDZ domain of SAP97 in complex with a GluR-A C-terminal peptide. FEBS J. *273*, 5219–5229.
Park, J., Kwak, J.-O., Riederer, B., Seidler, U., Cole, S.P.C., Lee, H.J., and Lee, M.G. (2014). Na+/H+ Exchanger Regulatory Factor 3 Is Critical for Multidrug Resistance Protein 4-Mediated Drug Efflux in the Kidney. J. Am. Soc. Nephrol. *25*, 726–736.
Pawson, T. (2007). Dynamic control of signaling by modular adaptor proteins. Curr. Opin. Cell Biol. *19*, 112–116.
Petoukhov, M. V., and Svergun, D.I. (2005). Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data. Biophys. J. *89*, 1237–1250.
Sobhy, H. (2016). A Review of Functional Motifs Utilized by Viruses. Proteomes *4*, 3.
Sugiura, T., Kato, Y., Kubo, Y., and Tsuji, A. (2006). Mutation in an adaptor protein PDZK1 affects transport activity of organic cation transporter OCTNs and oligopeptide transporter PEPT2. Drug Metab Pharmacokinet *21*, 375–383.
Sugiura, T., Kato, Y., Wakayama, T., Silver, D.L., Kubo, Y., Iseki, S., and Tsuji, A. (2008). PDZK1 regulates two intestinal solute carriers (Slc15a1 and Slc22a5) in mice. Drug Metab. Dispos. *36*, 1181–1188.
Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. *76*, 2879–2886.
Svergun, D.I., Petoukhov, M. V, and Koch, M.H.J. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophys. J. *80*, 2946–2953.
Teyra, J., Sidhu, S.S., and Kim, P.M. (2012). Elucidation of the binding

preferences of peptide recognition modules: SH3 and PDZ domains. FEBS Lett. *586*, 2631–2637.
Trave, G. (2011). PDZ-peptide complexes: As exciting as ever. Structure *19*, 1350–1351.
Turner, E.C., Mulvaney, E.P., Reid, H.M., and Kinsella, B.T. (2011). Interaction of the human prostacyclin receptor with the PDZ adapter protein PDZK1: role in endothelial cell migration and angiogenesis. Mol. Biol. Cell *22*, 2664–2679.
Walther, C., Caetano, F.A., Dunn, H.A., and Ferguson, S.S.G. (2015). PDZK1/NHERF3 Differentially Regulates Corticotropin-releasing Factor Receptor 1 and Serotonin 2A Receptor Signaling and Endocytosis. Cell. Signal. *27*, 519–531.
Wang, S., Yue, H., Derin, R.B., Guggino, W.B., and Li, M. (2000). Accessory protein facilitated CFTR-CFTR interaction, a molecular mechanism to potentiate the chloride channel activity [In Process Citation]. Cell *103*, 169–179.
Whitten, A.E., Cai, S., and Trewhella, J. (2008). MULCh: Modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies. J. Appl. Crystallogr. *41*, 222–226.
Woestenenk, E.A., Hammarström, M., Van Den Berg, S., Härd, T., and Berglund, H. (2004). His tag effect on solubility of human proteins produced in Escherichia coli: A comparison between four expression vectors. J. Struct. Funct. Genomics *5*, 217–229.
Yang, J., Singh, V., Chen, T.E., Sarker, R., Xiong, L., Cha, B., Jin, S., Li, X., Ming Tse, C., Zachos, N.C., et al. (2014). NHERF2/NHERF3 protein heterodimerization and macrocomplex formation are required for the inhibition of NHE3 activity by carbachol. J. Biol. Chem. *289*, 20039–20053.
Ye, F., and Zhang, M. (2013). Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. Biochem. J. *455*, 1–14.
Zachos, N.C., Li, X., Kovbasnjuk, O., Hogema, B., Sarker, R., Lee, L.J., Li, M., de Jonge, H., and Donowitz, M. (2009). NHERF3 (PDZK1) contributes to basal and calcium inhibition of NHE3 activity in Caco-2BBe cells. J. Biol. Chem. *284*, 23708–23718.

Figure Legends

Figure 1. **Schematic representation of the different PDZK1 constructs.**
Full-length PDZK1 (FL) is composed of four PDZ domains (blue ovals) connected by three linkers (brown) with a length of 16 (117-132), 19 (220-238) and 49 residues (325-373), respectively, as well as 6 N-terminal residues (1-6) and a C-terminal tail of 59 residues (461-519). Mutated residues in the linker region of the MUT construct are highlighted in light blue.

Figure 2. **SAXS profiles of tetra-domain constructs.**
MUT (purple) and FL (orange) compared to D14 (light-blue). (A, B) I(s) versus s as log-linear plots on absolute scale using scattering from water. (C, D) *P(r)* versus r profiles from the data normalized to unity at their maximum. (E, F) Dimensionless Kratky plots for the data as shown in A, B.

Figure 3. **Representative ab-initio and rigid body models of D14.**
(A) GASBOR ab-initio model of D14. (B) Fit of GASBOR model log(I(s)) versus s for D14. (C) MONSA ab-initio model showing the average domain dispositions within D14. D1 - orange; D2 - red; D3 - cyan; D4 - blue. (D) Model fits from MONSA to the constructs D12 (χ^2=0.72), D24 (χ^2=1.7), D13 (χ^2=1.5) and D14 (χ^2=1.26). (E) BUNCH model of D14 assuming flexible residues for 117-132, 220-238, and 325-374. (F) PDZ domains coloring is consistent with the MONSA model, with the linkers shown in grey. Fits: D12 (χ^2=1.12), D24 (χ^2=1.19), D13 (χ^2=1.37) and D14 (χ^2=1.15).

Figure 4. ***R_g*** **does not increase linearly with the number of domains in PDZK1.**
(A) Scatter plot of number of domains in each construct vs. R_g derived from Guinier analysis. A total of 12 different constructs were analyzed. (B) Scatter plot of number of domains in each construct vs. the d_{max} from the P(r) analysis. The data-points corresponding to the largest R_g and d_{max} results from the analysis of the MUT construct (triangle).

Figure 5. **Limited proteolysis of selected PDZK1 domain constructs with chymotrypsin**.

Limited proteolysis kinetics of selected PDZK1 domain constructs: (A) FL, (B) D14, (C) MUT, and (D) D34. The constructs were incubated with chymotrypsin at a molar ratio of 100:1 for different time points (5s, 5min, 10min, 40min, 2h, overnight; lanes from 2 to 7 respectively) and analyzed by SDS page. Respective untreated PDZK1 constructs (lane 2) and a molecular weight marker (lane 1), are shown as references.

Figure 6. **Thermostability data for PDZK1 constructs**.

Thermal stability unfolding curves using differential scanning fluorimetry of single domains (A) and multi-domain constructs (B). The table in panel (C) shows T_m values derived by a non-linear regression analysis using the Boltzmann equation, the protein concentration used during the experiment and a score for reversibility after cooling. The fluorescence ratio (F350nm/ F330nm) is represented as a function of temperature. The dashed lines indicate the apparent melting temperatures T_m.

Figure 7. **Binding of PDZK1 domain constructs to PEPT2-CT-FITC using fluorescence polarization.**

Binding isotherms were derived from the change of fluorescence anisotropy of PEPT2-CT-FITC as a function of protein concentration. Experimental data were fit by a 1:1 binding model shown as black line. (A) Binding of PDZK1 single domain constructs to the C-terminal peptide of PEPT2 (PEPT2-CT-FITC). Data points are black for D1, green for D2, blue for D3 and red for D4. (B) Binding of PDZK1 double-domain constructs to PEPT2-CT-FITC. Data points are colored black for D12, blue for D23 and red for D34. (C) Binding of PDZK1 triple-domain constructs to PEPT2-CT-FITC. Experimental data are colored blue for D13 and red for D24. (D) Binding of PDZK1 tetra-domain constructs to the C-terminal peptide of PEPT2. Data are shown in red for FL, black for D14 and green for MUT. All data points were recorded as triplicates with standard deviation illustrated by the error bars. The K_D values were calculated from three replicate experiments and are listed in Table S1.

Tables and Table Legends

	Constructs							
	D12	**D23**	**D34**	**D13**	**D24**	**D14**	**FL**	**MUT**
Structural Parameters								
I(0) (cm-1) [from *P(r)*]	0.023	0.021	N/A *	0.027	0.028	0.041	0.049	0.043
R_g (Å) [from *P(r)*]	30.8±0.2	30.8±0.2	34.4± 0.3	37.2±0.1	35.2±0.1	40.6±0.2	42.3±0.6	46.6±0.6
I(0) (cm^{-1}) [from Guinier]	0.023	0.021	N/A *	0.029	0.028	0.04	0.050	0.043
R_g (Å) [from Guinier]	29.5 ± 0.2	30.8 ± 0.3	33.4 ± 0.5	35.7 ± 0.2	34.8 ± 0.1	39.4 ± 0.1	42.6 ± 0.9	44.3 ± 0.6
d_{max} (Å)	113.40	110.00	121.00	126.00	125.00	150.00	155.00	168.00
Dry volume calculated from seq. ($Å^3$)	32325	29206	32884	46517	47521	64832	73310	64095
Guinier region	72-149	52-165	10-103	42-158	11-142	16-147	16-79	18-96
Molecular-mass determination								
Partial specific volume (cm^3g^{-1})	0.7392	0.7392	0.7387	0.7399	0.7407	0.7408	0.7383	0.7398
Contrast	2.936	2.928	2.959	2.932	2.921	2.926	2.957	2.941
Molecular mass from (from *I(0)*) (kDA)	29	28	N/A	37	38	50	66	55
Molecular mass from Standard P. (kDA)	32	28	N/A	38	39	60	60	54
Calculated monomeric M_r from sequence (kDA)	26.4	23.8	26.8	37.9	38.7	52.7	59.8	52.2

Table 1. SAXS Parameters Table. The forward scattering of construct D34 cannot be determined, as this construct was collected with SEC-SAXS.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Christian Löw (christian.loew@embl-hamburg.de).

METHOD DETAILS

Cloning and Construct design

The cDNA clone (IRAUp969D0730D) containing the gene sequence of human PDZK1 (UniProtKB - Q5T2W1) was purchased from Source Bioscience. The designed PDZK1 constructs were amplified and cloned into the pNIC28-Bsa4 expression vector (N-terminal His-tag) (Woestenenk et al., 2004) using ligation independent cloning (Aslanidis and de Jong, 1990). The N-terminal His-tag was followed by a 16 amino acid long linker with a TEV cleavage site. The correct insertion of the gene sequence was verified by DNA sequencing. Details for individual PDZK1 constructs are shown in Table S1. Domain borders of constructs were designed based on known X-ray and NMR structures of PDZ domains.

Expression and purification of PDZK1 constructs

All protein constructs were expressed and purified according to the following protocol. Recombinant proteins were expressed in *E. coli* BL21(DE3). Cultures of 1 L LB medium in 2.5 L baffled conical flasks were inoculated from a LB overnight culture and grown at 37 °C to an OD_{600nm} of ~0.8. Then the temperature of the cultures were reduced to 18 °C and protein expression was induced by the addition of 0.2 mM isopropyl β-D-thiogalactopyranoside (IPTG). After 16 h of induction, the cells were harvested at 10,000 × g for 10 min and the cell pellets were stored frozen at −80 °C. Frozen cell pellets were thawed on ice and resuspended in lysis buffer (20 mM NaP pH 7.5, 150 mM NaCl, 5%

Glycerol, 0.5 mM TCEP, 15 mM imidazole, 1 mg/mL lysozyme, 5 U/ml DNase, one EDTA-free Complete inhibitor tablet per 100 ml). For 1 g of cell pellet, 5 ml of lysis buffer was used. Resuspended cells were incubated under stirring at 4 °C for 30 min. Cells were disrupted with an Emulsiflex microfluidizer (Avestin C3) at 15,000 p.s.i. chamber pressure in three runs and centrifuged at 18,000 × g for 30 min at 4 °C. In the first step of purification IMAC was used. The supernatant was incubated with 2 ml (settled) of Ni-NTA Agarose resin (Invitrogen) under rotation for 1 hour and then transferred to a 10-mm-(i.d.) open gravity flow column and washed with 15 bed volumes of IMAC wash buffer (20 mM NaP pH 7.5, 150 mM NaCl, 5% Glycerol, 0.5 mM TCEP, 15 mM imidazole). Target proteins were eluted by the addition of five bed volumes of IMAC elution buffer (20 mM NaP pH 7.5, 150 mM NaCl, 5% Glycerol, 0.5 mM TCEP, 250 mM imidazole). For construct D3 and D4 the His tag was cleaved with TEV protease. The protein was dialyzed overnight against 1 l of gel filtration buffer in a 2 kDa dialysis bag at 4 ºC. The TEV-cleaved sample was recovered by negative IMAC, concentrated by ultrafiltration and gel filtration was performed. For the remaining constructs, IMAC-eluted fractions were collected and concentrated by ultrafiltration using Corning® Spin-X® UF concentrators with a cut-off limit of 5 kDa for single domains, 10 kDa for multiple domains and 30 kDa for FT, D14 and MUT constructs prior gel filtration. The concentrated protein samples were loaded on a HiLoad Superdex™ 75 pg 16/60 GL column (GE Healthcare) equilibrated with GF buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP) using an ÄKTAPure chromatography system. The elution profile was monitored at 280 nm and collected fractions containing the target protein were concentrated to approximately 5–20 mg/ml protein, flash frozen in liquid nitrogen and stored at −80 °C until further use.

Analytical gel filtration

To assess the quality of the purified proteins, samples were analyzed by analytical gel filtration on a home-packed Superdex 75 pg (SD75) 5/150 home-packed column (GE Healthcare) using the 1260 Infinity Bio-inert high-performance liquid chromatography system (Agilent Technologies). The system was equipped with an auto-sampler and 20 µl of samples at 1 mg/ml were injected on the gel filtration column and analyzed in duplicates. Runs were

performed at 4 °C at a flow rate of 0.2 ml/min in GF buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP). Prior to injection of the samples, the gel filtration column was equilibrated with three column volumes of GF buffer and calibrated with known protein standards.

Peptide preparation

Peptide stock solutions were prepared by weighing lyophilized powder in an analytical balance. The powder was dissolved in GF buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP) to an estimated concentration of 15 mM. The concentration of the FITC-labelled peptides was determined using a molar extinction coefficient of 70.000 M^{-1} cm^{-1} at 495 nm, while the concentration of PEPT2-CT and NHE3-CT has been assessed by measuring the absorption at 215 nm. Typically, the spectroscopically determined concentrations were by 20-30% lower as expected from the weight. The following peptides were used in this study.

Peptide	**Sequence**
PEPT2-CT	N-acetyl-IKLETKKTKL-COOH
PEPT2-CT-FITC	N-FITC-IKLETKKTKL-COOH
PDZK1-CT -FITC	N- FITC-SSSNSEDTEM-COOH
NHE3-CT-FITC	N- FITC- RPPAALPESTHM -COOH
NHE3-CT	N- acetyl- RPPAALPESTHM -COOH
PEPT1-CT-FITC	N-FITC-MSGANSQKQM-COOH

Limited Proteolysis

Stability tests for the expressed PDZK1 constructs against proteolytic degradation were performed at neutral pH at room temperature on purified protein in the presence of chymotrypsin (chymotrypsin: target protein molar ratio was 1:100). The reaction was stopped at different time points by addition of SDS sample buffer and subsequent heating. The samples (1 µg of protein per well) were then analyzed on NuPAGE 4-12% Bis-Tris Protein Gels.

Fluorescence anisotropy

Fluorescence anisotropy measurements were performed in GF buffer (see buffer contents above) supplemented with BSA at a concentration of 1 mg/ml to prevent unspecific interaction of the peptide with the plate. The fluorescently labelled peptide was kept constant at a concentration of 2 nM in a total volume of 150 µl far below K_D. PDZK1 constructs were serially diluted by a factor of 0.66. Samples were prepared in triplicates in black flat-bottom polystyrene NBS 96 - well plates (Greiner, Product No. Art.-Nr.: 655209). Anisotropy measurements were conducted in an Infinite M1000 plate reader (TECAN) at 25 °C. Excitation wavelength was 470 nm and the emitted light was recorded at 530 nm. The anisotropy measurements were setup in such a way that PDZK1 constructs were titrated in excess compared to the fluorescently labelled peptides assuming that only one peptide is bound to any of the PDZK1 constructs at a time allowing a direct comparison of the affinities for the different constructs. Resulting data were analyzed with a 1:1 binding model using Prism (GraphPad software). The affinities for the PEPT2 peptide of the individual domains account well for the observed affinities in the MUT construct but differ for D14, indicating that the spatial arrangement in the full length context plays an important role in binding.t

Crystallization and data collection

Crystallization trials were carried out by vapor diffusion in 96-well sitting-drop plates at 19°C. The drops had a total volume of 300 nL with a 2:1, 1:1 and 1:2 ratio of protein to mother liquor. Prior to crystallization, D4 at a concentration of 15 mg/mL was mixed with the PEPT2-CT at a molar ratio of 1:5 (protein to ligand). Initial crystallization hits were obtained using the commercial screen Classic I (Qiagen). The crystallization condition was further optimized and contained 20% PEG 4000, 0.1 M trisodium citrate pH 6.45 and 20 %(v/v) isopropanol. Crystals appeared after 2 days and were harvested on day 4. The crystals were flash-frozen in liquid nitrogen with 20% glycerol as cryo-protectant. Data were collected at 100K at the ID30-B beamline at ESRF (Grenoble, France) (Mueller-Dieckmann et al., 2015).

Data processing, refinement and structural analysis

Diffraction data were indexed, integrated, and scaled with the XDS program suite (Kabsch, 2010). The high-resolution cut-offs were estimated according to CC1/2 (Karplus and Diederichs, 2012). Data statistics and parameters are summarized in Table 2. Crystals of D4 in complex with the PEPT2 peptide belonged to space group P6(5)22 and contained one molecule per asymmetric unit. The structure of PDZ4 domain of mouse PDZK1 (PDB ID: 4R2Z) was used as a model for molecular replacement using Phaser (Adams et al., 2010). One molecule was positioned with a translation function Z-score of 23.9. The resulting electron density map was then subjected to the AutoBuild program, part of the Phenix program suite (Adams et al., 2010). An initial model of 79 residues in one chain and 127 water molecules was obtained. The structure was completed with sessions of manual model building using Coot combined with model refinement using Phenix (Adams et al., 2010).

SAXS data-acquisition and data processing

The SAXS data were collected at the P12 BioSAXS Beamline at the PETRAIII Storage Ring, DESY, Hamburg, Germany (Blanchet et al., 2015) using a standard batch-mode set-up with automated sample delivery. Samples and the corresponding matched solvents were measured at 15 ^{o}C under continuous sample flow using a total exposure time of 1 s (recorded as 30 $\times$ 50 ms frames), and concentration ranges of typically 0.5 mg/ml to 10 mg/ml, with 4-5 concentrations each. For the peptide titration studies, the concentration of D14 was kept constant at 100 µM, and the peptide concentrations of NHE3-CT and PEPT2-CT were 50 µM, 100 µM, 500 µM and 2000 µM respectively. The primary 2D-data underwent standard automated processing (radial averaging, subtraction, etc.) combined with additional manual and statistical evaluation (e.g., for radiation damage) to produce the final 1D-SAXS profiles I(s) vs s, where $s=4\pi sin\theta/\lambda 2$ and 2θ is the scattering angle and the X-ray wavelength λ =1.24 nm. On-line SEC-SAXS data were collected on the D34 construct, as this construct readily formed dimers in solution. Data were collected at 10 °C at an initial concentration of 11 mg/ml, with one frame per second for an hour at a flow-rate of 0.4 ml/min.

The molecular mass of all constructs was assessed from the forward (zero-angle) scattering intensity ($I(0)$) using the method of (Orthaber et al., 2000) by placing the data on an absolute scale using scattering from water. Values for contrast and partial specific volumes were calculated using the MULCh suite of analysis tools (Whitten et al., 2008). Guinier analysis were performed using PRIMUS (Franke et al., 2017) to determine the radius of gyration (R_g) and forward scattering $I(0)$. PRIMUS was also used to calculate the probability distribution of distances between atom pairs (P(r) profiles) within each protein construct from which the d_{max}, R_g, and $I(0)$ values were determined.

SAXS data modeling

Ab-initio modeling was performed using GASBOR (Svergun et al., 2001) that represents the structure as a collection of dummy residues. Multiple runs yielded an average most representative model. Finally, a second ab-initio modeling procedure, MONSA, was used that refines the spatial positioning of the individual domains against the datasets of the constructs. Further, multiple MONSA models were made to ensure that the domain dispositions were consistent between runs.

The crystal or NMR structures of the individual PDZ domains [Protein Data Bank (PDB) ID PDZ Domain 1: 4F8K, Domain 2: 2EEI, Domain 3: 3R68, Domain 4: 2VSP] were used with the program BUNCH (Petoukhov and Svergun, 2005) to refine the positions of the relevant domains against the scattering data of all the constructs (see Figure 1). In these refinements, BUNCH generates dummy atoms to account for those parts of the sequence of unknown structure, in this case the loop regions of PDZK1, and refines their shape/mass distribution in space.

BUNCH refinements were run multiple times in order to obtain an ensemble of structures representing the domain organization/average mass distributions within each construct. To improve confidence in the final average positions of the modules within the constructs containing more than 2 domains, BUNCH calculations were performed that simultaneously fit multiple data sets. The datasets corresponding to D23 and D34 were excluded in the modeling of

the D14 construct, because of systematic deviations in the fit pointing to possible alterations in the structures of these constructs in the free states.

Bioinformatic analysis

Protein disorder was predicted using two platforms: ANCHOR (Dosztányi et al., 2009) and DISOPRED3 (Jones and Cozzetto, 2015). All results were generated with the sequence of FL and default settings. Secondary structure prediction was performed through PSIPRED (Buchan et al., 2013) and sequence conservation and solvent accessibility through ConSurf (Berezin et al., 2004).

DLS measurements

Dynamic Light Scattering experiments were collected on the following constructs: FL, D14, MUT, D13 and D24 with a Wyatt DynaPro NanoStar DLS with a laser wavelength of 824.7 nm. The concentration of the constructs were 5 mg/ml, with the exception of D13 which had a concentration of 3.25 mg/ml. Ten DLS acquisitions, 10 seconds each, were made in triplicates at 15 °C . The scattering data were processed with the Dynamics DynaPro Control software v7.6.0.48.

Thermal stability measurements

Thermal stability of the purified proteins was measured by differential scanning fluorimetry (DSF) using the Prometheus NT.48 instrument (NanoTemper Technologies, GmbH). For this, 10 µL of each sample at a concentration of 0.5-2 mg/mL was loaded in a capillary. Measurements were done in triplicates. The temperature was increased by a heating rate of 1 °C/min from 15 °C to 80 °C and the fluorescence at emission wavelengths of 330 nm and 350 nm was measured. To check the reversibility of the unfolding reaction the sample was cooled at a cooling rate of 1 °C/min and the fluorescence recorded. Most constructs precipitated after unfolding or at higher temperatures, so only small domain constructs were partially reversible. Data were processed and analyzed using Prism (GraphPad software). Transition regions of unfolding and refolding profiles were fitted using the Boltzmann equation to determine T_M.

QUANTIFICATION AND STATISTICAL ANALYSIS

SAXS peptide binding studies were analyzed with OLIGOMER to determine the volume fractions of the peptide and the protein in the solution. The contribution of the peptide to the scattering was assessed through singular value decomposition using SVDPLOT (Konarev et al., 2003), by determining the number of components needed to adequately describe the system. To assess whether the datasets were statistically identical, a goodness of fit test independent of errors, correlation map, was used (Franke et al., 2015).

DATA AND SOFTWARE AVAILABILITY

PDB accession

Coordinates and structure factors of D4 with the C-terminal peptide of PEPT2 were deposited in the protein data bank (PDB), with the following accession code: 6EZI.

SASBDB accession

Scattering data of the D14 construct with its models have been deposited to the small angle scattering biological data bank (SASBDB), with the following accession code: SASDD85.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Materials		
Vector for protein expression pNIC28-Bsa4	EMBL-HAmburg	N/A
cDNA clone (IRAUp969D0730D) containing the gene sequence of human PDZK1	Source Bioscience	N/A
DH5α *E. coli* Cloning strain	EMBL Hamburg	N/A
BL21(DE3) *E. coli* Expression strain	EMBL Hamburg	N/A
PEPT2-CT	GL Biochem (Shanghai)	N/A
PEPT2-CT-FITC	GL Biochem (Shanghai)	N/A
PDZK1-CT -FITC	GL Biochem (Shanghai)	N/A
NHE3-CT-FITC	GL Biochem (Shanghai)	N/A
NHE3-CT	GL Biochem (Shanghai)	N/A
PEPT1-CT-FITC	GL Biochem (Shanghai)	N/A
Luria Bertani Broth (LB)	Melford	Cat #GL1703
Ni-NTA Agarose beads	Invitrogen	Cat #R901-15
crystallization reagents	Qiagen	N/A
TCEP (Tris(2-carboxyethyl)phosphine hydrochloride)	Soltec Ventures	Cat #51805-45-9
DNase I (Deoxyribonuclease I)	AppliChem	Cat #A3778
cOmplete™, EDTA-free Protease Inhibitor Cocktail	ROCHE	Cat #5056489001
chymotrypsin	ROCHE	Cat #11418467001
isopropyl β-D-thiogalactopyranoside (IPTG)	Roth	Cat #2316.5
di-Sodium hydrogen phosphate heptahydrate	Roth	Cat #X987.2
Sodium dihydrogen phosphate monohydrate, >98%, p.a., ACS	Roth	Cat #K300.2
Sodium chloride, >99,5%, p.a., ACS, ISO	Roth	Cat #3957.2
Glycerol, Rotipuran >99,5%, p.a.	Roth	Cat #3783.1
Lysozyme	Roth	Cat #8259.2
Imidazol, Pufferan >99%, p.a.	Roth	Cat #X998.4
HEPES, Pufferan >99,5%, p.a.	Roth	Cat #9105.3
Roti®-Mark 10-150 Protein-Marker	Roth	Cat #T850
NuPAGE® LDS Sample Buffer (4X)	Invitrogen	Cat #NP008
NuPAGE® Novex® 4-12% gels	LifeTechnologies	Cat #NP0323BOX
InstantBlueTM	Expedeon	Cat #ISB1L
Software and Algorithms		
XDS / XSCALE / XDSCONV	Kabsch, 2010	

PHENIX	Adams et al., 2010	
PHASER	McCoy et al., 2007	
COOT	Emsley et al., 2010	
MolProbity	Chen et al., 2010	
ATSAS	Franke et. al. 2017	
MULCh	Whitten et. al 2008	
ConSeq	Berezin *et al.*, 2004	
ANCHOR	Dosztányi, Mészáros and Simon, 2009	
PSIPRED	Buchan *et al.*, 2013	
Protein Data Bank (PDB)		www.pdb.org
PyMol		Schrödinger LLC; http://www.pymol.org
GraphPad Prism		https://www.graphpad.com
Deposited data		
Crystal structure of D4 in complex with the C-terminal peptide of PEPT2	This paper	PDB: 6EZI
Scattering data and models of the D14 construct	This paper	SASBDB: SASDD85
Crystal structure of D4 used as a model for molecular replacement	pdb.org	PDB: 4R2Z
Crystal structure of D1 used as a model for SAXS modeling	Birrane et al., 2013	PDB: 4F8K
NMR structure of D2 used as a model for SAXS modeling	pdb.org	PDB: 2EEI
Crystal structure of D3 used as a model for SAXS modeling	Kocher et al., 2011	PDB: 3R68
Crystal structure of D4 used as a model for SAXS modeling	pdb.org	PDB: 2VSP

Figure 1

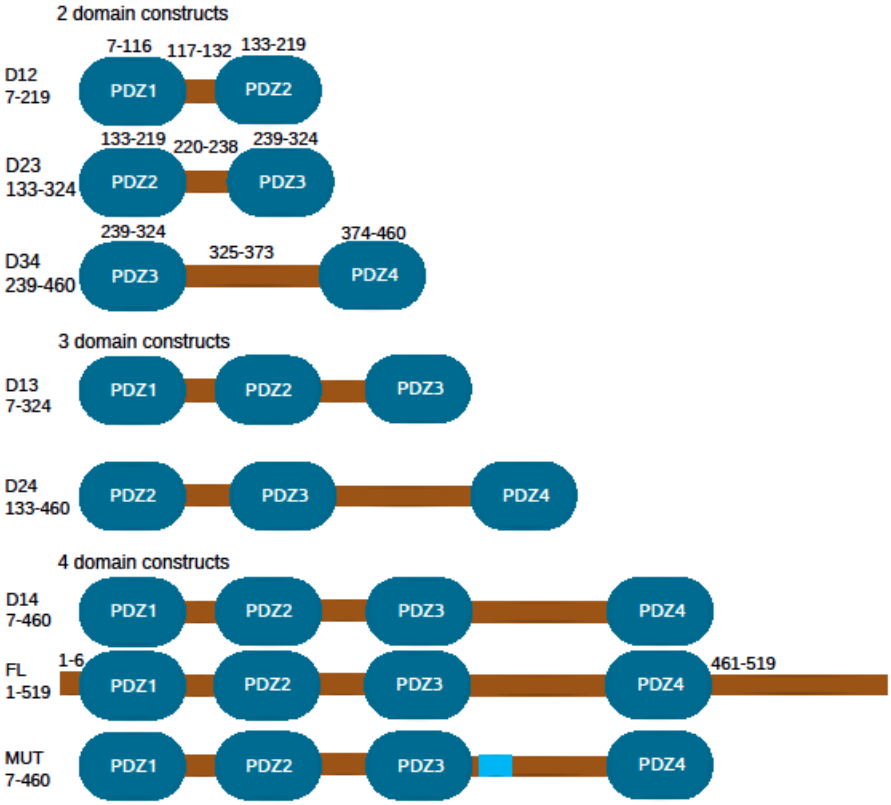

2 domain constructs
7-116
117-132
133-219
D12
7-219
PDZ1
PDZ2
133-219
220-238
239-324
D23
133-324
PDZ2
PDZ3
239-324
374-460
D34
239-460
325-373
PDZ3
PDZ4
3 domain constructs
D13
7-324
PDZ1
PDZ2
PDZ3
D24
133-460
PDZ2
PDZ3
PDZ4
4 domain constructs
D14
7-460
PDZ1
PDZ2
PDZ3
PDZ4
FL
1-519
1-6
461-519
PDZ1
PDZ2
PDZ3
PDZ4
MUT
7-460
PDZ1
PDZ2
PDZ3
PDZ4

Figure 2

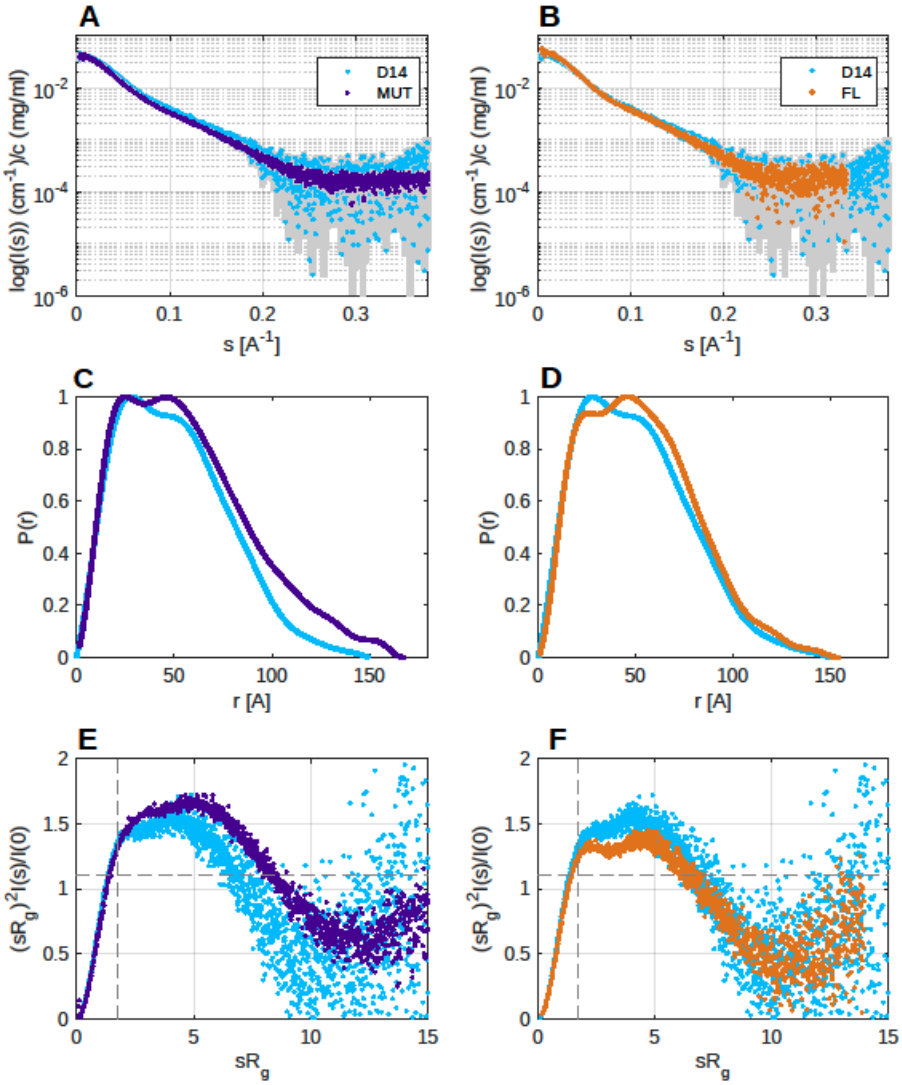


Figure 3

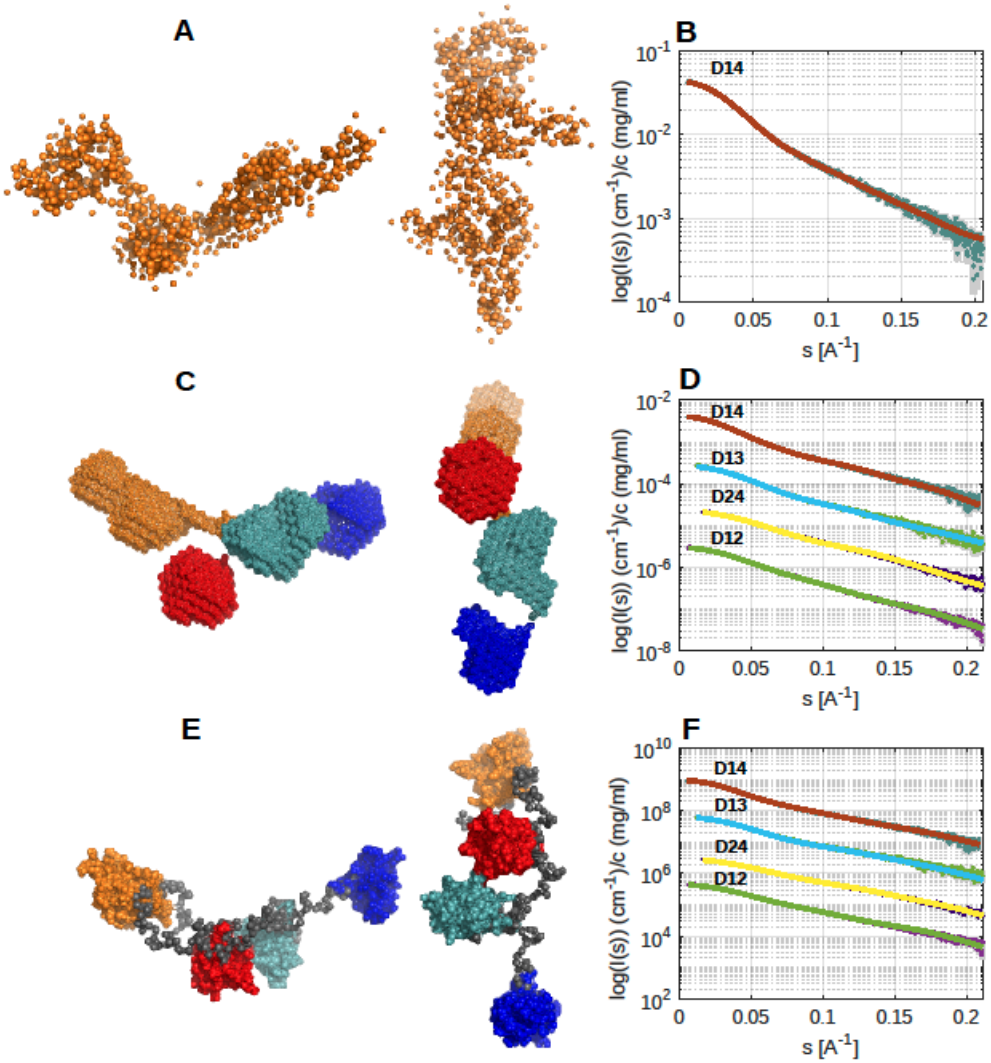


Figure 4

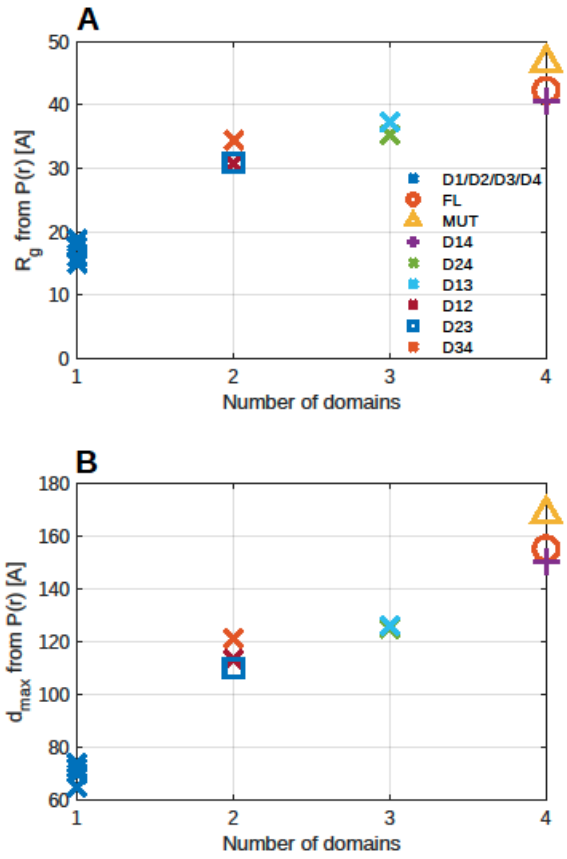


Figure 5

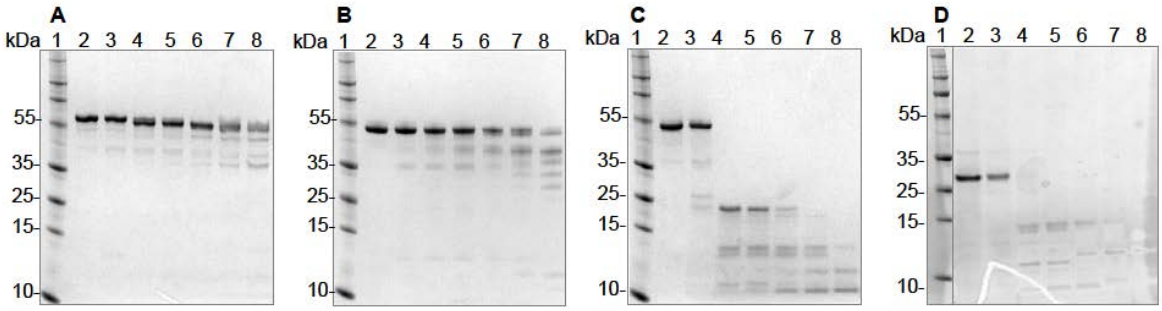
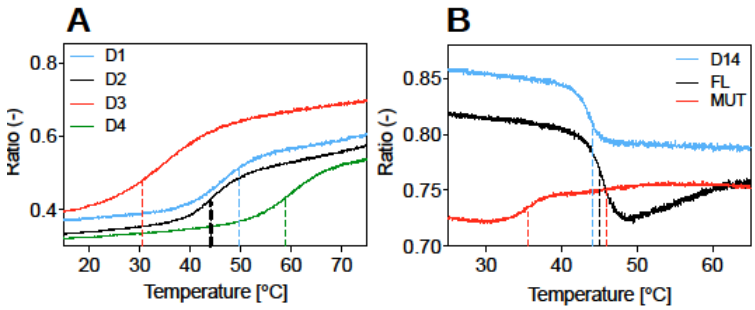


Figure 6



PDZK1 constructs	T_m [°C]	Protein concentration [μM]	Reversibility
D1	49.7±0.1	66	-
D2	44.0±0.1	80	+
D3	30.5±0.3	105	+
D4	58.8±0.1	105	++
D14	44.1±0.1	19	-
MUT	35.5±0.2/46.0±0.6*	19	-
FL	45.1±0.1	17	-

- not reversible
+ partial reversible
++ reversible
* T_{m1}/T_{m2}

Figure 7

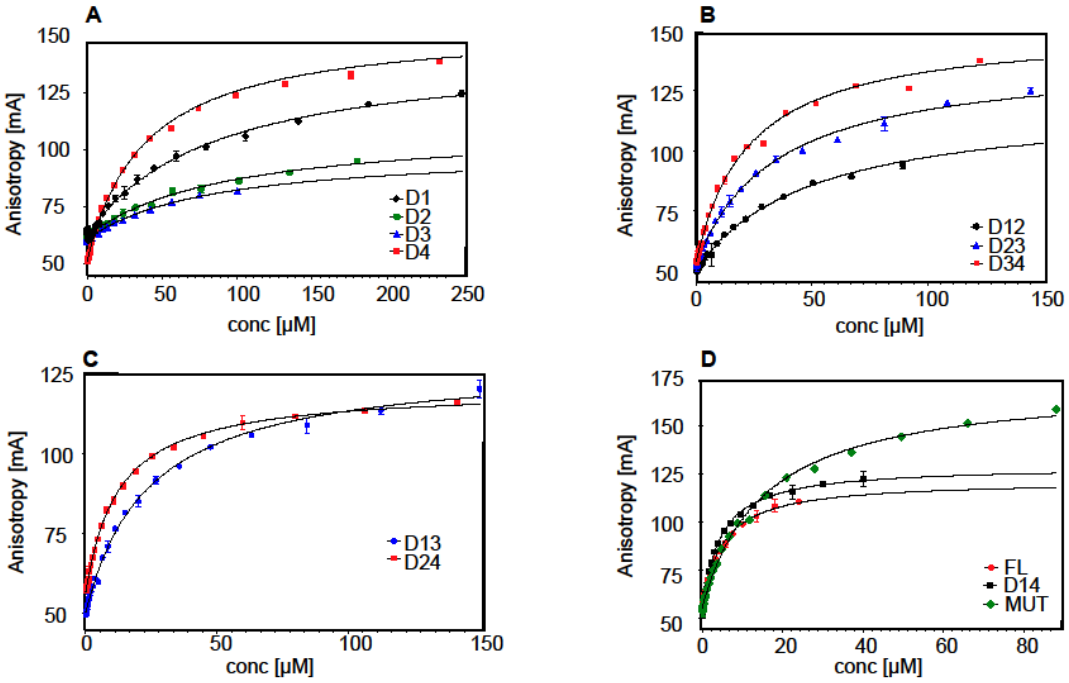


Figure 8